International Journal of Interactive Multimedia and Artificial Intelligence

December 2019, Vol. V, Number 7
ISSN: 1989-1660

unir LA UNIVERSIDAD EN INTERNET

"International Journal of Interactive Multimedia and Artificial Intelligence was selected for inclusion in the Science Citation Index Expanded beginning with vol. 4(3) March 2017. The journal will be listed in the Journal Citation Reports with a Journal Impact Factor and other metrics when released in June 2020."

Marian Hollingsworth

INTERNATIONAL JOURNAL OF INTERACTIVE MULTIMEDIA AND ARTIFICIAL INTELLIGENCE
ISSN: 1989-1660 –VOL. 5, NUMBER 7

Editor's Note

The International Journal of Interactive Multimedia and Artificial Intelligence - IJIMAI (ISSN 1989 - 1660) provides an interdisciplinary forum in which scientists and professionals can share their research results and report new advances on Artificial Intelligence (AI) tools or tools that use AI with interactive multimedia techniques.

After its recent tenth anniversary, the journal has achieved an important milestone. From 2015 to 2018 IJIMAI was indexed at Web of Science through Emerging Science Citation Index. This meant a great increase in visibility and number of received papers. This year, Clarivate Analytics has accepted the inclusion of IJIMAI in the Journal Citation Reports. Specifically, IJIMAI is being indexed and abstracted in Science Citation Index Expanded, Journal Citation Reports/Science Edition and Current Contents®/Engineering Computing and Technology. The Web of Science Categories in which IJIMAI is included are "Computer Science, Artificial Intelligence" and "Computer Science, Interdisciplinary Applications". This way, IJIMAI is indexed in Science Citation Index Expanded beginning with vol. 4(3) March 2017 so that the journal will be listed in the 2019 Journal Citation Reports with a Journal Impact Factor when released in June 2020. Given this great achievement, IJIMAI Editorial Board has to thank authors for all the papers sent and all the papers published, as well as reviewers for their support to obtain high-quality in papers, and specially our readers because without them this milestone would not have been possible.

The present regular issue includes research works based on different AI methods such as convolutional neural networks, genetic algorithms, lightning attachment procedure optimization, or those of multi-agent systems. These methods are applied into various fields as video surveillance, gesture recognition, sentiment analysis, territory planning, search engines, epidemiological surveillance or robotics.

The first article of this issue, written by Khiat and Hamdadou [1], targets a problem of spatial localization in territory planning through the proposal of a multicriteria group decision support system based on a multi-agent system. This system reproduces the behavior of decision makers considering the multiplicity and diversity of criteria as well as of own decision-makers. Moreover authors deal with a bounded temporal dimension so that the proposed system finds a solution before fixed deadline expires.

The second article, from Khari et al. [2], focuses on static gesture recognition based techniques. Specifically the aim is to recognized signs of American Sign Language (ASL). They propose a fine-tuned VGG19 Model, a convolutional neural network. The advantage of this method is that it does not rely on features extraction and helps in reducing the computational power required. Besides, it achieves a high recognition rate when tested on an ASL dataset, outperforming other methods.

Also related to computer vision, in the third article, Joshi et al. [3] present an approach for smart video surveillance to track an already classified unidentified vehicle to handle its occlusion. It is difficult to retain features during occlusion especially in case of complete occlusion. The authors propose a computationally efficient approach that works through two periods, a tracking period when there is no occlusion and a detection period when occlusion occurs. The tests on six scenarios prove that the algorithm presents good robustness against high noise and low illumination circumstances.

In the area of computer security, Rodríguez et al. [4] propose a symmetric-key cryptographic algorithm for text, which applies genetic algorithms philosophy, entropy and modular arithmetic. When comparing the proposed algorithm against RSA and DES, good performance is achieved in several factors, proving that genetic algorithms are a good option when facing problems in computer security.

Retaking the area of computer vision, this time applied to robotics, Sudin et al. [5] propose a novel localization method consisting of a corner extraction algorithm and a distance estimation algorithm for efficiently identifying salient goalposts for Robocup soccer humanoid competitions. The research arises from the fact that recent rules of Robocup discard the use of the middle pole to deliberately provide less prior information for the humanoid vision system. The technique provides highly accurate corner detection and distance estimation compared to other techniques.

Next research applies to web search engines, which process huge amounts of documents and queries so early termination algorithms are convenient to avoid processing all documents. Mansouria et al. [6] modify the WAND algorithm used in search engines, to early terminate with faster and more precise results. Also, they define new fine metrics to improve the evaluation of the retrieved information.

Kumar and Harish [7] present an article about sentiment analysis. Automatic detection of sarcasm or irony from content in microblogging reviews is a challenging task and the authors propose feature fusion to provide knowledge to the system by alternative sets of features obtained using linguistic and content based text features. The features are fused and classified using Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), Decision Tree (DT) and ensemble classifiers. To enhance the performance of the classifiers, they propose a weighted majority voting schema to create an ensemble from the decision of each classifier. The proposed approach is able to capture ironic utterances present in the reviews out-performing existing methods on benchmark dataset.

The next article deals with convergence of media, specifically radio and online spaces. Laor [8] studies the success of radio programs beyond their native FM environment, focusing on their attempts at achieving popularity on social networks. Success on social networks is measured by user involvement and interaction with posted content and comments. The results show that radio program activity on social networks promotes higher levels of interaction with listeners beyond broadcast schedule. Besides, integration of various media forms such as videos or images increases the likelihood of a post becoming popular.

Schrepp and Thomaschewski [9] describe the development and first validation studies of a modular framework for the creation of user experience (UX) questionnaires. This framework allows to select the UX aspects that are relevant for a certain product from a list of existing UX scales, as sometimes none of the existing questionnaires contain all the scales needed to answer a given research question.

In the field of spatial data mining, Zemri and Hamdadou [10] present the SOLAM (Spatial On Line Analytical Mining) system, an extension of Spatial On Line Analytical Processing (SOLAP) with Spatial Data Mining (SDM) techniques. They integrate the EPISOLAP system, targeted to epidemiological surveillance, with a spatial generalization method allowing the predictive evaluation of health risk. The spatial generalization allows exploring the data at different semantic and spatial scales while reducing the unnecessary dimensions.

Next three articles propose solutions to electrical engineering problems, focusing on power systems. The first one, by Mohamed et al. [11], presents the genetic moth swarm algorithm, which is an hybrid approach based on genetic algorithms and moth swarm algorithms, for

DOI: 10.9781/ijimai.2019.10.006

determining the optimal location and sizing of renewable distributed generation sources on radial distribution networks. The aim is to minimize the electrical power loss under security constraints. The experiments done prove the effectiveness of the presented approach compared with other methods under several power system constraints and conditions.

The second one of Ibrahim et al. [12] proposes the use of artificial neural networks to improve the performance of static synchronous series compensators (SSSC) integrated into combined wind farms (CWF). Their results illustrate that the performance of CWF can be improved using SSSC adjusted by a neural network, when they compare with CWF with ordinary SSSC and CWF with SSSC tuned by a multiobjective genetic algorithm.

The last article on power systems, by Kamel and Youssef [13], proposes an approach to determine the optimal allocation of different shunt compensation devices in power systems. The approach combines the use of lightning attachment procedure optimization and loss sensitivity indices. The developed algorithm is validated using standard IEEE 14-bus and IEEE 30-bus test systems. It achieves the objective functions with better performance compared to other optimization methods, such as teaching learning-based optimization, genetic algorithm and particle swarm optimization.

Next article, written by Padilla-Zea et al. [14], targets the long-term unemployment problem. It presents a gamified educational platform to empower social economy entrepreneurship skills in long-term unemployed people. The learning path is shaped as a story that guides the work throughout the training process. Two pilots were developed in Italy and Spain, involving five facilitators and around 60 learners. The analysis shows that this game can address various learner profiles that will benefit based on their features. This multi-faceted use of the game allows targeting multiple learners and competences.

This regular issue ends with a study by Rodríguez et al. [15] that aims to carry out a comparison of image recognition methods for the purpose of evaluating exercises performed in an immersive environment for motor skills training. The compared methods are convolutional neural network, K-nearest neighbor, support vector machine and decision tree. The assessment of the techniques is carried out using images captured from an immersive environment, obtaining that the convolutional neural network model has the best performance.

Dr. Rubén González Crespo

Dr. Elena Verdú

References

[1] S. Khiat and H. Djamila, "A Temporal Distributed Group Decision Support System Based on Multi-Criteria Analysis," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 7, pp. 7-21, 2019, doi: 10.9781/ijimai.2019.03.002.

[2] M. Khari, A. K. Garg, R. Gonzalez-Crespo, and E. Verdú, "Gesture Recognition of RGB and RGB-D Static Images Using Convolutional Neural Networks," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 7, pp. 22-27, 2019, doi: 10.9781/ijimai.2019.09.002.

[3] R.C. Joshi, A. G. Singh, M. Joshi, and S. Mathur, "A Low Cost and Computationally Efficient Approach for Occlusion Handling in Video Surveillance Systems," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 7, pp. 28-38, 2019, doi: 10.9781/ijimai.2019.01.001.

[4] J. Rodríguez, B. Corredor and C. Suárez, "Genetic Operators Applied to Symmetric Cryptography," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 7, pp. 39-49, 2019, doi: 10.9781/ijimai.2019.07.006.

[5] M. N. Sudin, S. S. N. H. Abdullah and M. F. Nasudin, "Humanoid Localization on Robocup Field using Corner Intersection and Geometric Distance Estimation," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 7, pp. 50-56, 2019, doi: 10.9781/ijimai.2019.04.001.

[6] Z. I. Mansouria, Z. Lougmiri and S. Mohamed, "MWAND: A New Early Termination Algorithm for Fast and Efficient Query Evaluation," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 7, pp. 57-69, 2019, doi: 10.9781/ijimai.2019.04.002.

[7] K. H. M. Kumar and B. S. Harish, "Automatic Irony Detection using Feature Fusion and Ensemble Classifier," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 7, pp. 70-79, 2019, doi: 10.9781/ijimai.2019.07.002.

[8] T. Laor, "'Hello, is There Anybody Who Reads Me?' Radio Programs and Popular Facebook Posts," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 7, pp. 80-87, 2019, doi: 10.9781/ijimai.2019.10.003.

[9] M. Schrepp and J. Thomaschewski, "Design and Validation of a Framework for the Creation of User Experience Questionnaires," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 7, pp. 88-95, 2019, doi: 10.9781/ijimai.2019.06.006.

[10] F. A. Zemri and D. Hamdadou, "SOLAM: A Novel Approach of Spatial Aggregation in SOLAP Systems," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 7, pp. 96-104, 2019, doi: 10.9781/ijimai.2019.10.002.

[11] E. A. Mohamed, A. A. Mohamed, and Y. Mitani, "Genetic-Moth Swarm Algorithm for Optimal Placement and Capacity of Renewable DG Sources in Distribution Systems," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 7, pp. 105-117, 2019, doi: 10.9781/ijimai.2019.10.005.

[12] Y. Ibrahim, S. Kamel, A. Rashad, Loai Nasrat, F. Jurado, "Performance Enhancement of Wind Farms Using Tuned SSSC Based on Artificial Neural Network," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 7, pp. 118-124, 2019, doi: 10.9781/ijimai.2019.05.001.

[13] S. Kamel and H. Youssef, "Voltage Stability Enhancement Based on Optimal Allocation of Shunt Compensation Devices Using Lightning Attachment Procedure Optimization," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 7, pp. 125-134, 2019, doi: 10.9781/ijimai.2019.09.002.

[14] N. Padilla-Zea, S. Aceto, and D. Burgos, "Social Seducement: Empowering Social Economy Entrepreneurship. The Training Approach," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 7, pp. 135-150, 2019, doi: 10.9781/ijimai.2019.09.001.

[15] A. O. R. Rodríguez, M. A. Riaño, P. A. Gaona-García, C. E. Montenegro-Marín, Í. Sarría, "Image Classification Methods Applied in Immersive Environments for Fine Motor Skills Training in Early Education", International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 7, pp. 151-158, 2019, doi: 10.9781/ijimai.2019.10.004.

TABLE OF CONTENTS

OPEN ACCESS JOURNAL

ISSN: 1989-1660

The International Journal of Interactive Multimedia and Artificial Intelligence is covered in Clarivate Analytics services and products. Specifically, this publication is indexed and abstracted in: *Science Citation Index Expanded, Journal Citation Reports/ Science Edition, Current Contents®/Engineering Computing and Technology*.

A Temporal Distributed Group Decision Support System Based on Multi-Criteria Analysis

Sofiane Khiat*, Djamila Hamdadou

Computer Science Laboratory of Oran (LIO) - Université Oran1, Oran (Algeria)

Received 8 August 2018 | Accepted 15 February 2019 | Published 8 March 2019

Abstract

Decision support consists of proposing tasks and projects by taking into account temporal constraints and the use of resources with the aim of finding a compromise solution between several alternatives. Indeed, on the one hand, centralized resolution systems and methods are generally inappropriate to the real case because of the local unavailability of decision makers. On the other hand, the data of the decisional problem are generally poorly expressed in a negotiation environment. Other techniques and approaches treat the same decision-making problem and impose a distributed vision for coherent decisions. For this purpose, Multi-Agent Systems (MAS) allow modeling a distributed resolution of the group decision support problem. In this article, we propose a new model of a multi-criteria group decision support system based on a multi-agent system modeling a spatial problem. We consider that each decision maker is assimilated to an agent that has a decision-making autonomy, in which he interacts with other agents in the debate through a negotiation process in order to reach an acceptable compromise. In this study, we propose coordination mechanisms among agents to highlight the simulated negotiation. Therefore, the proposed system finds a solution before fixed deadlines' time expire. We experiment the suggested negotiation model to solve the decisional problem of spatial localization in territory planning.

Keywords

Group Decision Support System, Spatio Temporal Modeling, Multi Agent Systems, Multicriteria Decision Analysis, Negotiation, Territory Planning.

DOI: 10.9781/ijimai.2019.03.002

I. Introduction

THE field of decision support has been the subject of research conducted by several scientists of various fields. Therefore, it has constituted an attractive pole for different studies and applications. Among these applications, we find several cases which are related to problems with spatial reference: urban and regional planning, transport, management of water resources, environmental management, evaluation of the territory and location of industrial activities, etc. The problems which are related to the area of spatial decision support focus on the selection of geographical sites [14] [4] [37]. This latter is based on the choice between several sets of physical criteria [12] [13].

Moreover, decision support systems aim to help decision makers in their tasks by providing them with all the relevant elements for decision-making and spatial planning. In fact, territorial decision support involves several conflicting criteria, whose importance are not the same. The territorial decision support involves several decision-makers and institutions, which generally have divergent preferences and objectives where various points of view must be taken into account for a final decision [23]. The decision appears as a compromise between several interests and divergent points of view that imply the use of a negotiation strategy between the various involved actors who must take decisions as quickly as possible by taking into account a functioning constrained time. Consequently, this permits to obtain an answer to an interrogation before a given deadline in order to allow decision makers to act as quickly and appropriately as possible.

* Corresponding author.

E-mail address: khiat.sofiane@gmail.com

Territory planning (TP) is based on a prospective and strategic vision that takes into account the potential, physical, social, economic, and environmental constraints of the concerned territory [14]. This type of problem involves several decision makers (persons and institutions) with different interests who have generally divergent preferences and objectives and whose different points of view must be taken into account for public decision.

The realization of spatial localization in TP relies on mathematical methods and computer tools such as geographical information systems (GIS). The application of GIS is oriented to several areas: urban development [39], environmental management [10], the territory evaluation [12], industrial diagnosis [41] etc.

More particularly, spatial decision support systems are interesting, especially in the development of a model of group decision support which is dedicated to the problems of space localization in TP: the problem which entails the search for a surface on a geographical map satisfying a set of criteria and finding a compromise between several interests that appeal to the expertise of several people, particularly those which are concerned by the decision.

However, the group of decision-makers will be modeled by a set of entities called computer agents. These agents represent each decision-maker in a multi-agent system (MAS) [11] [27].

The problem addressed in the current study concerns the proposition of a system that models the different decision makers who have their own information, constraints, decision strategies, preferences, and objectives generally not shared or communicated. Hence, the need for a negotiation process integrated into a group decision support system (GDSS) allows finding a common agreement for this group, in the face of a conflict. Several reasons can be mentioned besides the group and

the decision-making for a common goal, which we quote:

1. *The multi-criteria aspect ("several conditions per alternative "):* allows the identification and the measurement of the alternatives or solutions on which the decision will be made. It is therefore to build a family of criteria that can represent as closely as possible the costs and benefits of the actions
2. *The geographical distribution* of decision-making actors is justified by organizational realities. Decision-making can thus bring together distributed actors on one or more sites.
3. *The bounded temporal dimension ("a definite period of time")* is essential because it induces a beginning and an end to the activity, so it is ensured by an evaluation strategy and the means of communication.

The research work conducted in this study is summarized, in this paper, as follows: In section 2, we present a literature review on group decision support systems, their characteristics, and their topological evolution. In section 3, we propose solutions related to the problem of distributed negotiation with a specific deadline. Section 4 is dedicated to describe our proposal for a distributed group decision support system. In Section 5, we describe how the proposed system works. Section 6 is dedicated to the description of the design and implementation of the proposed approaches.

Finally, we will conclude this article by summarizing the various contributions we have proposed to the problem of distribution and negotiation in group decision support systems. We will end by reciting some tracks of research that seem relevant to the problem dealt with in this study.

II. Related Work

The presented work is integrated in the context of Decision Support Systems (DSS). The later is considered in two main dimensions: the individual and the collective dimension.

The individual dimension is to provide a decision support to an expert decision maker in a field and to propose the resolution of a particular problem. Solving the problem follows a pre-established decision support process that is based on breaking down the problem into tasks and subtasks to have a satisfactory solution. Several decision support systems with a single decision maker have been proposed by using multi-criteria analysis methods:

In [1], the authors presented a method of multicriteria decision support to evaluate the decision of internalization / outsourcing as a part of a sustainable development strategy and they evaluate the strategic importance of the activities. The proposed method makes it possible to calculate an overall performance index by using the method AHP (Analytical Hierarchy Process) with indicators. In [2], the authors presented a case study on the implementation of a multi-criteria approach to the performance and risk exposure of a bank. The proposed methodology is based on the PROMETHEE II method implemented in an integrated decision-making system.

In the same optics as our research, other studies have been conducted by exploiting geographic information systems (GIS): In [3], a methodology for evaluating built-up urban space was proposed (authors developed a decision support system for housing valuation). This system integrates a problem editor, a database management module, a set of multi-criteria decision support methods and an adequate human-computer interface that can be integrated with GIS tools. In [4], the objective was to provide a decision support in urban infrastructure which is planned to users. In addition, visualization of available alternatives on maps provide an added value to decision-making processes in urban infrastructure assessment issues. The development of this system was motivated by a real urban case study.

In [38] a fuzzy hierarchical analysis method (FAHP) combined with a geographical information system (GIS) has been proposed. The authors presented a process for ranking industrial sites in Algeria. The proposed process of decision-making is based on the AHP method. Also the GIS is used to prepare geographic data in screening phase and to visualize ranked zones on a map in the evaluation phase. In the same area of use of the AHP multicriteria method and techniques that deal with transport problems, in [40] an analysis of the factors of urban mobility in the situation of cities has been proposed. The objective is to take into consideration all the elements involved in mobility in urban environments, in which their behavior was studied.

The works cited above do not always reflect reality because decision-making does not concern a single decision-maker, which has led to the development of group decision support systems, where a set of decision-makers, sometimes geographically dispersed, with different values and with potentially conflicting issues are involved. As a result, we identify the second dimension, the collective one.

The collective dimension concerns the collaborative aspect because it consists of providing collective decision support, where each decision-maker is involved in each step of the decision-making process. As a consequence, several works have been proposed:

In [5], a group decision model based on ELECTRE GD has been proposed. It is a group decision method constructed on ELECTRE III. The proposed model generates a collective solution that helps decision makers with different interests to reach (through an iterative process) an agreement on how to classify their alternatives. In [6], a methodology for remote group decision support (GDM) in case of emergency is proposed. In this model, some decision-makers are identified to formulate a group decision-making framework and a multi-criteria decision-making process is carried out, in which different results are obtained from diverse decision makers to verify the effectiveness of emergency management.

Other works have been invested in coordination between decision-makers, for a global decision-making, which was considered as a common interest. Cao and all [7] propose a theoretical vision of coordination in the use of the multi-criteria tools for the decision support system (DSS) intended for the groups. The authors proposed an extension by formulating parallel and sequential coordination methods for the distribution of multi-criteria tools. These methods can be used by DSS users to coordinate and structure the distribution of multi-criteria tools for groups. The study proposed in [8] had as main objective the implementation of parallel and sequential coordination methods in a web-based multicriteria group decision support system. The authors presented two methods of coordination that influenced the collaborative group decision process and the final consensual solution in the context of distributed group decision support multicriteria analysis. In [9] the authors set up a web-based multicriteria decision support system, which solved multi-criteria arrangement problems in a collaborative group of decision makers in sequential or parallel coordination mode and in a distributed and asynchronous environment.

In a context of simulating the behavior of decision-makers, several researchers have proposed group decision support systems (GDSS) with architectures based on multi-agent modeling (MAS). The decision-makers are modeled in such systems by intelligent agents. Below, we are going to identify some works on this aspect.

The author in [10] made a simulator based on a multi-agent system whose objective was to provide the negotiators with an instrument to test the consequences of a regulation in order to reach an acceptable compromise. In [11], a three-layer system structure had been proposed. This structure allowed for the implementation of a distributed intelligent decision-making system for a marketing decision. The authors developed the marketing system supported by a distributed decision support.

The modelling of decision-makers by intelligent agents in a group decision support system is a very interesting field in current research because it has become more important especially when the data are of spatial type (geographic area). Several researchers have proposed GDSS models to address spatial location issues by considering a set of decision-makers. However, few are the works that considered the multi-criteria aspect and the multi actors aspect at the same time. The main existing works in the literature were carried out in our research team. Consequently, we are going to quote the most significant works carried out:

In [12], the authors proposed a decision quality optimization study in the context of spatial data management using a decision support model. The suggested model allowed experts to carry out diagnoses and proposed adapted alternatives by modelling negotiation and multi-stakeholder participation using multi-agent systems. In [13], the authors' objective was to integrate multi-criteria analysis methods (MCAM) into a decision support system based on SOLAP technology (Spatial On-Line Analytical Processing), which was modeled and implemented as part of decision support systems dedicated to epidemiological surveillance. In [14], the authors proposed a strategy for the design and development of a spatial group decision support system and multicriterion. A multi-agent modeling (MAS) with a negotiation protocol based on mediation is proposed to conduct the spatial localization process. The latter was implemented in the territory planning. In the context of group decision support systems (GDSS) that model negotiation, a coordinator is involved in order to help decision-makers to negotiate [15]. The main role of a coordinator is to find a satisfactory compromise for all decision-makers. The Negotiation in multi-agent systems (MAS) gives growth to two different approaches [16], [17]:

1. *Analytical and normative approach based on game theory.* Zlotkin and Rosenschein [18] applied game theory tools in multi-agent systems (MAS).
2. *An approach based on the behavior of actors in the negotiation process.* Most of the works in this direction lead to the development of negotiation protocols. The best known among them is the Contract Net Protocol (CNP) [19] developed for coordination.

The first approach can only be used if the mathematical models support negotiation. It has an analytical or numerical solution (Monte Carlo method for example) [20]. The second is the only one possible to explore complex processes, in which relationships between agents are not reducible to mathematical models [10].

We have made improvements to integrate the previous approaches into a distributed decision support system proposed in this article. It is obvious that a large number of solutions and the distribution of choices, makes a large part of the decision support system. The metaheuristic based on time constraints has proved to be an efficient decision support tool for assignment problem solving. For example, in [42] the authors presented an optimization algorithm that solved a Rich Vehicle Routing Problem (RVRP) and arose from a research project carried out for an important Spanish distribution company.

Lastly, in a coordination strategy, it is obvious that a negotiation by proposal of solutions cannot be buckled to infinity. As a result, the strategy that we proposed takes into account time parameter and proposes a policy of the time management.

Claude Duvallet [21] was mainly interested in the study of the real time aspect in multi-agent systems. He proposed ANYMAS model (ANYtime MultiAgents System) for the design of a real-time multi-agent system (RTMAS) based on the use of anytime algorithms.

In this study, we suggest a model of a decision support system based on agents. The decision makers who engage in decision making can be geographically remote (distributed), where each decision maker is modeled in this system by an intelligent agent. All agents follow a collective decision support process guided by an elected coordinator agent. Such a system is propped by negotiation simulation mechanisms.

III. Our Contribution

We place our contribution in the context of critical decision-making situations, where collective decision-making activities are generally characterized by synchronous cooperation sessions within *distributed* environments which are evolving and often unpredictable related to problems of *multiplicity of decision-makers and their preferences.*

- *Distributed (location problem):* Decision-makers act simultaneously and from distributed access points on shared objects by following implicit or explicit coordination rules and by using a set of tools that allow them to progress in a coordinated manner.
- *Multi criteria and multi decision makers:* The members who have different interests, skills, and experiences express their preferences in the form of a choice between several possible solutions to several criteria, which can relatively be of different nature: economic, social, environmental, technical ect. Decision making requires a synergy of efforts from several members, so that each one of them can use their know-how.

The main contribution proposed, in this article, particularly in the area of collaborative decision support, is to organize the performance of interdependent tasks over time by taking into account the temporal constraints and the use of data with the objective of optimizing one or more criteria. The methods of traditional resolution known as centralized, are generally poorly adapted to the real case because the data is geographically dispersed. For this purpose, the multi-agent systems (MAS) constitute a paradigm quite appropriate and powerful which allow the modeling and the distributed resolution of the spatial location problem in a Territory Planning (TP). In this work, we are interested in a problem which consists of searching for a surface on a geographical map that satisfies a set of criteria. In addition, we propose to model and solve this problem in a distributed way by using a multi-agent approach. We consider that each decision maker is assimilated to an agent who has a decisional autonomy, and who can also cooperates with the other agents in order to reach a mutually acceptable global solution. Negotiation is a powerful mechanism for finding mutually acceptable compromises. Literary, the proposed approaches are based on a multilateral protocol that have a coordinator agent and a group of negotiators participating agents, who try to find a compromise that best satisfies the various decision-makers. Indeed, our objective is to propose mechanisms of cooperation between agents by electing one agent among the group of agents in order to ensure consistency in decisions that are locally taken. This new approach is encouraging because it looks like the way humans negotiate. During a negotiation, the proposed solutions allow the agents to interact with an offer or a proposal related to their points of view, and their preferences. The main objective of our contribution is to:

- Design, develop, and implement a Group Decision Support System (GDSS) to represent the multiplicity and diversity of actors by proposing a negotiation protocol.

Other more specific objectives are targeted:

- Deploying the proposed GDSS in a distributed architecture.
- Representing the preferences of each decision maker by exploiting the main advantages offered by multi criteria analysis methods.
- Guarantying the temporary aspect basing on a negotiation strategy.

IV. The Proposed Decision Support System

The problem addressed consists in seeking a view as precise as possible of a given situation and obtaining a maximum of relevant information that satisfy a set of criteria and preferences which makes it possible to reach a consensus.

For that, we implement a distributed group decision support system operating in three main phases: (1) pre-negotiation; (2) negotiation; and (3) post-negotiation. The proposed system is illustrated in Fig. 1.

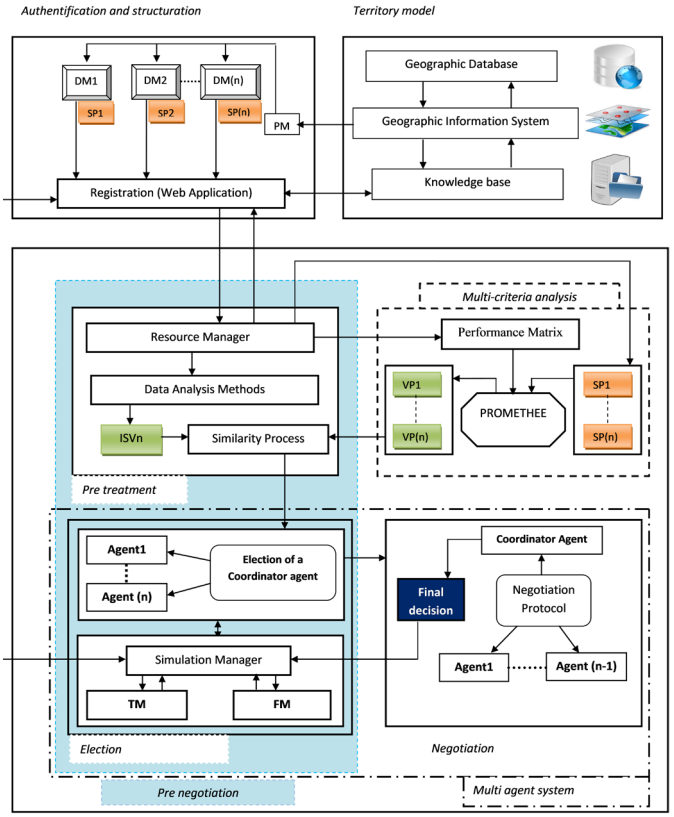


Fig. 1. Proposed Group Decision System.

The "Pre-negotiation" phase allows exploring and opening the negotiation space. The last phase "Post-negotiation" closes this space. The second phase "negotiation" includes an important stage known as a proposal of solution (decision support). The later contains four main stages that constitute the building blocks of any decision-making process, whose Simon's [22] model shows the different steps for individual decision making.

Table I summarizes some concepts used in the proposed decisional approach:

A. Authentication and Structuring

For a distributed group decision problem, each decision maker in this group is modeled by an agent of a MAS, who allows the interaction between the different agents by a negotiation process. Accordingly, each decision-maker is located in geographically dispersed locations and he must involve his preferences through a web application. This phase also allows to display the final solution in output (the solution of the problem). This solution is stored in the knowledge base taking into account multiple possible uses of solutions to evaluate scenarios for similar future situations.

TABLE I. Description of Acronyms

Acronyms	Description	Role
PM	Performance Matrix	A set of data describing the decisional problem, each performance represents the evaluation of each alternative against each of the criteria.
SPi	Subjective parameters of each decision maker i	Values that express the preferences of the each decision-maker (weight of a criterion, its preference and indifference).
ISVn	Initial Solution Vectors	A ranking of all the alternatives in a vector by a specific order.
VP	Vector of Preference	The vector that expresses the ranking of alternatives for each decision-maker after execution of the multi-criteria method.
TM	Time Manager	Time controller by event sequence.
FM	Fault Manager	Failure of a Technical and physical system controller.
n	Number of decision makers involved in solving the decision problem	Number of decision makers counter.
DMi	Decision makers from 1 to n	System users.
PROMETHEE	Preference Ranking Organization METHod for Enrichment Evaluations	A multicriteria analysis method based on the elementary mechanism which is the comparison of the alternatives according to each criterion(two by two).

B. Territory Model

This model makes possible to adjust the analysis of the object and the space, to explain why we find this or that phenomenon here and not elsewhere, to interrogate a set of modules to provide a set of possible solutions to a given situation or problem, and to manage modifications and recordings in the knowledge base. It has its own structure and functioning, it inscribes itself in the space and the time (spatio-temporal scales). The territorial model comprises three components which are ultimately in interaction.

1. The Geographic Information System Module GIS

It is an information system designed to collect, store, process, analyze, manage, and present all types of spatial and geographic data. The essential function of GIS is to enable knowledge management of the territory, it is able to [23]:

- Manage the geographic database.
- Archive information in a knowledge base.
- Manipulate and query geographic databases.
- Provide a spatial representation of the studied systems.
- Visualize the data.

When decision-makers are able to identify alternatives and criteria through using the analytical capabilities of GIS in which a value (score) is assigned to each criterion the set of alternatives and their scores relative to the different criteria constitutes the matrix evaluation (Matrix Performance).

2. The Geographic Database Module

It is a set of spatial and non-spatial data structured and organized to be searchable and analyzable interactively or automatically. A geographic

database usually relates to a defined area. It is managed by GIS software. Furthermore, It integrates the data itself as well as their metadata.

3. The Knowledge Base Module

This knowledge base consists of the decision-making session's directory. It is implemented to save and store the data related to the final decision, the trace of decision-making sessions, intermediate results, system elements, and shared documents. Its main purpose is to store all the solved cases in the past (problems and their solutions) and to provide these results to decision-makers. The flowchart in Fig. 2 shows this interaction.

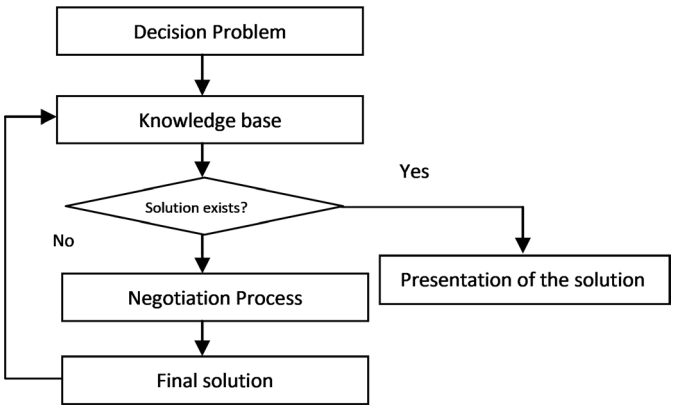


Fig. 2. Flowchart of a knowledge base decision support system.

C. Negotiation Model

The proposed system illustrated in Fig. 1 is composed of several modules which deals with a given group decision problem where the negotiation module is composed, mainly, of four sub-modules described below.

1. The Pre-Negotiation Module

We identify two major elements that mark this phase: *pretreatment* (pre-processing) and *election*. These two elements are ubiquitous in the rest of the decision-making process:

Pretreatment: In order to better adapt our proposed system to control mechanisms, we have equipped this module by a data manager which is able to detect any hardware and software infrastructure included in the system to better keep the environment stable, especially it is important to maintain an initial state between the system and the users. This makes it possible to define the deadlines in a negotiation process and control the interactions of decision-makers. The challenge of this notion is to allow the best choice of constraint that can express the best decision. The second notion consists of ranking the alternatives from the best to the worst without the notion of preference being present. This allows to assign to each available alternative a rank (a ranking vector). This approach is ensured by the data analysis method to solve this problem. Accordingly, we have exploited the different steps of this approach:

- Clustering ("solution categorization").
- Data analysis method ("choice of the best criterion").

The clustering consists of grouping objects in order to build predefined categories or classes. This type is a part of the classification problems. This is done by using a set of examples named as a set of solutions or alternatives (learning set) . They are made up of objects, whose category membership is not known (unsupervised aspect). Methods that solve this type of problem can be used to build / discover categories. Moreover, this type of method is used when the decision maker is not able to specify the categories [24].

However, the data analysis method makes it possible to choose the best criterion. In this paper, we are interested in determining a new measure for proposing solutions to decision-makers adapted to problems of multi-criteria decision support. This technique makes possible to classify the solutions using the unsupervised classification (use of k-means) in order to apply a learning algorithm. It is a question of finding a partitioning of the individuals which better represents the classes of each individual. This partition is then presented in the form of a decision tree.

Election: after obtaining the ranking of all the alternatives in a vector that will be called vector of initial solutions, this is compared to each preference vector (VP) of the decision maker to have the best similarity, in order to choose a preference vector of a decision-maker (among all decision-makers) that contains a ranking of solutions similar to the vector of initial solutions. This makes it possible to choose the participant who will be responsible for the good progress of the negotiation named the coordinator (in MAS). Additionally, this step is ensured by the similarity process. A time manager is set up to ensure the negotiation within the deadlines, to detect other breakdowns, and the event of exceeding time deadlines. The fault manager will automatically trigger if an agent disappears or leaves the negotiation. Therefore, in this case the negotiation must be restarted.

The two steps mentioned above are placed in a distributed decision support system, in which a simulation manager is set up to choose the best actor within a well-defined time interval in order to solve a specific decision problem.

2. The Multicriteria Analysis Module

Multicriteria decision analysis consists of constructing models that deal with decision problems taking into account several criteria. Each criterion addresses a set of homogeneous consequences. This is an important factor to evaluate a given scenario or to appreciate an occasion of alternative [25] and that the decision-maker must consider which are important and which are less important. The Multi Criterion Analysis (MCA) allows to deal with the multiplicity, the divergence, and the nature (quantitative or qualitative) of the criteria in order to reach acceptable compromises [41].

MCA is based on a coherent of criteria's family constructed and started from a set of consequences or evaluation (performance) of each alternative of A = {a1, a2 ...., an} on a family of criteria F= {g...., gn} which is provided by gj (ai). These evaluations can be summarized in Table II. The application of this definition is called a Table of Performance (Performance Matrix).

TABLE II. Performance Matrix

	g1	g2	……	g(n)
a1	g1(a1)	g2(a1)	……	gn(a1)
a2	g1(a2)	g2(a2)	……	gn(a2)
……	……	……	……	……
a(n)	g1(an)	g2(an)	……	gn(an)

In the context of our study, we chose the multi-criteria analysis method PROMETHEE II [26] which consists of ranking the alternatives according to an order of preference. This problem seeks to obtain a complete preorder on the set A of each participant, who must introduce his preferences (in the form of subjective parameters), depending on the application and the situation treated. Table III summarizes the various subjective parameters used in the multi-criteria method. They can be classified into two categories: "intercriteria parameters" and "intracriteria parameters" [23].

TABLE III. SUBJECTIVE PARAMETERS

Parameters	Symbol	Meaning
weight	Pj	Qualifies the relative importance of a given criterion Cj with respect to the other criteria.
Preference threshold	PTj	The threshold at which the difference between the two alternatives is perceptible and makes one preferable to the other.
Veto threshold	Vtj	Allows fixing an additional notion. If this threshold is exceeded on a criterion, then the alternative cannot be taken into consideration. Thus, it defines an intolerable situation for one of the participants.
Indifference threshold	Itj	This is the smallest significant difference. Below that threshold, it is impossible to separate the two actions.

The first characteristic of the proposed group decision support system is decentralization. Such a system is lacking a central coordinator entity for the organization of the multi-agent system which involves a negotiation between the participants.

3. The Multi-Agent System Module MAS

In the first place, it is necessary that the information is available to decision-makers as soon as possible. However, Multi-Agent Systems (MAS) are particularly appropriate when dealing with Group Decision Support Systems (GDSS). Indeed, the agents make possible to reproduce the global functioning of a GDSS from the entities which compose it (GDSS) and interactions. The MAS is a tool which makes it possible to express an application and a behavior of the decision makers by autonomous agents, who play roles and render services in an organization.

The different distributed decision-makers who have their own objectives are modeled by agents that have their own objectives and preferences. This implies that the decision process is distributed among the different entities which are basically involved in this group decision via a web application. The MAS allows the representation of interactions between various entities that can cooperate, negotiate, and communicate. In addition, an agent that evolves in an environment must be able to receive information from this environment, and to act in the same environment by following a decided behavior according to the agent's reasoning. The agent is characterized by his architecture and his behavior, so that he can accomplish what is expected of him. Depending on the architectures and capacities, the agents are classified into several types that qualify them as cognitive, reactive, or hybrid [27].

In the context of our study, our system involves reactive agents. Each agent is controlled by a time manager for its overall operation and interactions with other agent in the platform.

For this purpose, we endow the MAS module with a negotiation protocol based on the election that involves one negotiator agent (coordinator agent) among the others, and a set of participating agents who represent different actors that are involved in a collective decision. We are interested in a particular application class: applications based on distributed group decision support systems by intelligent agents. In this class of applications, we will look for a final decision in relation to a given situation within a constrained time frame. Therefore, we will find a solution before fixed deadlines time expiries. Fig. 3 illustrates the interaction between the negotiator agent and a participating agent who is controlled by this environment.

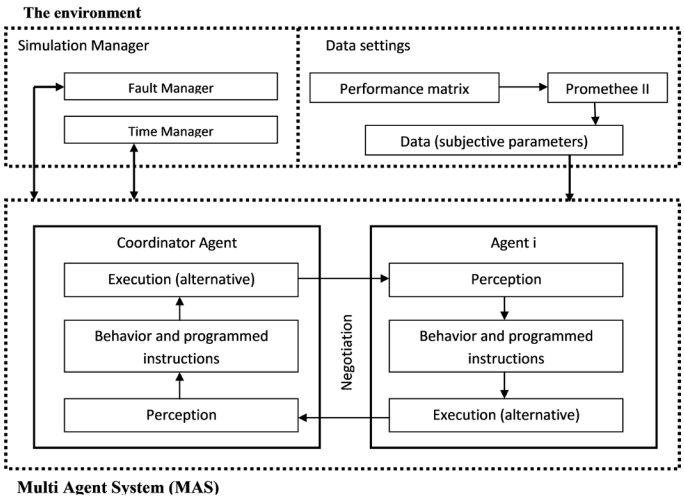


Fig. 3. Overall architecture of the proposed system.

4. The Negotiation Module

Distributed Group Decision Support System (GDSS) refers to any computer system based on a multi-agent model, where each participant can be an offeror or applicant of a service in the face of a conflict, as opposed to the Client / Server model, where each participant has a specific role (either he offers a service, which makes him a server "elected agent", or he is a service requester, which makes him a client).

Our contribution is to search for a surface on a geographical map that satisfies a set of criteria and finds a common agreement between the participants. Basically, this includes negotiation ensured by an elected coordinator, who offers solutions (alternative) from an initial vector of solutions.

In our research, we are interested in the negotiation in multi-agent systems (MAS). In such a system, the negotiation can resolve conflict situations between participants through the following characteristics:

- *The proposition of solutions* (alternatives) through a negotiation protocol that can be carried out directly from individual to individual [28].
- *The use of a coordinator* [29]: the process of negotiation based on sending messages between a coordinator and participants through a protocol is the most widely used, easily adapted one that actually models the way humans react to each other.
- *The election of a coordinator* among all participants to decentralize the negotiation protocol and distribute it.

V. THE PROCESS OF PROPOSED SYSTEM

In this section, we present a flowchart that illustrates the collaborative and distributed decision support process based on our observations and our analysis of the models proposed in the literature [30] , [7] and adapted to our design of cooperative decision support modules. The model we propose is able to support the decision-making process in the particular multi-decision-maker context which is distributed with time constraints.

The flowchart in Fig. 4, demonstrates the sequence and approach taken by the system to conduct the collective distributed decision support process.

The decision-making flowchart is divided into several parts. The first part is reserved for the notion of initiation of the decision-making process and the formulation of the problem.

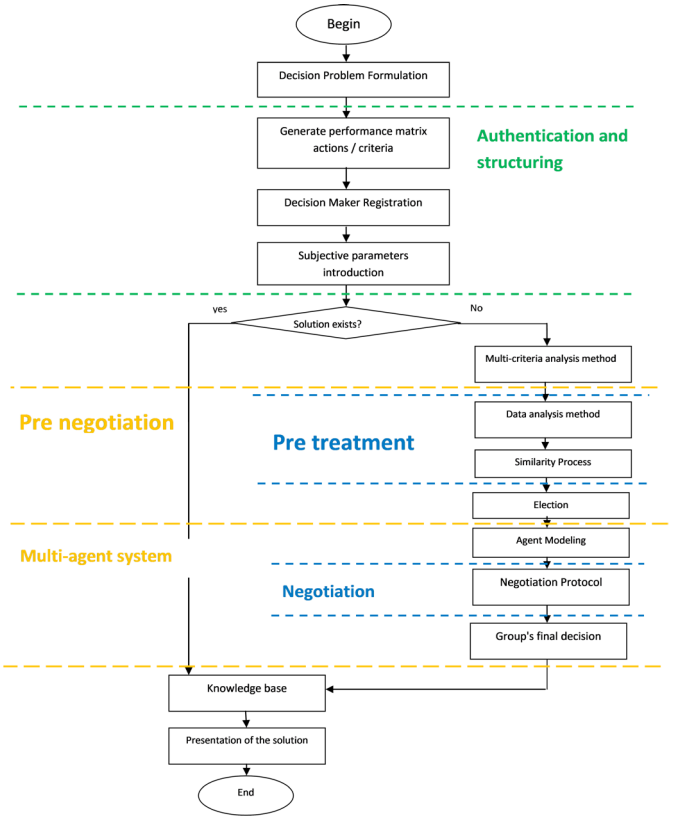


Fig. 4. Flowchart of the proposed approach.

At this level, the system is responsible for several tasks mentioned below:

1. The determination of data (performance matrix);
2. The invitation of decision makers;
3. The configuration of the decision-making environment (triggering and managing the temporal sequence) ;
4. The creation of a representation and a shared context (registration of decision makers).

The decision support model that we propose allows the decision maker to describe their preferences through a web application to solve a decision problem. The system tries to check if the solution already exists, if yes, the adapted proposal to this problem is exposed directly to the decision-makers, otherwise it is a new problem that requires the execution of the system processes to start another part of multi-criteria analysis; this notion is described in section V.A.

A decision support environment must therefore allow the simulation manager to trigger decision-making mechanisms; some of them are related to the time management and others are related to the management of breakdowns and tasks.

To summarize, the simulation manager proposes a list of decision-makers according to their registration order and profiles. The simulation manager is therefore responsible for choosing a negotiation coordinator and triggering other mechanism, such as the time manager, which synchronizes processes and determines the temporary deadlines of the responses (the decisions) in the negotiation phase. The measures of the time manager will be explained in the following sections.

A. PROMETHEE Method

It is a multi-criteria decision support analysis method, and it is the acronym for Preference Ranking Organisation METHod for Enrichment Evaluations. PROMETHEE was proposed for the first time by Jean Pierre Brans [26].

PROMETHEE belongs to the family of outclassing methods, in which two particular mathematical treatments are proposed: the first makes it possible to classify the alternatives in a partial pre-order that leads to incomparability (the PROMETHEE I method). The second allows classifying the potential alternatives according to a total pre-order (PROMETHEE II method).

These methods address any multi-criteria problem of the type:

TAB {g1 (a1), g2 (a2)..... gn (m) /a Є A} where A = {a1, a2 ... am} is a set of alternatives evaluated on a set of n criteria G = {g1, g2 ... gn}. Let F = {1, 2 ... n} be the set of criteria indices.

The data relating to such a problem can be represented in a table TAB (n × m) of dimension called performance matrix. Detailed in section IV.C.2.

The PROMETHEE method is exploited by each decision maker. Besides, it consists of establishing a process of numerical comparison of each alternative compared to all other alternatives. Consequently, it is possible to calculate the most important (merit) or the least important (demerit) of each alternative compared to all the others [26]. The result of this comparison allows the ordered ranking of alternatives (solutions) in a table called ordered ranking vector for each decision-maker. Therefore, it must consider a degree of two important parameters: the threshold of preference and indifference chosen quite easily by the decision-maker.

Both methods PROMETHEE I and II have the same initial reasoning, but their objectives are different in terms of classification of alternatives. PROMETHEE I makes it possible to identify relationships by a partial classification; whereas PROMETHEE II provides a classification of all the actions known as a total aggregation. We are interested in the PROMETHEE II method because of its advantages.

1. Why PROMETHEE II

The PROMETHEE II method is among the most used methods in the category of outranking methods, because it offers a number of advantages of which we quote [31]:

1. It integrates the recent developments in the modeling of preferences in a simple way.
2. It has a mathematical basis, so that it programs and improves its functionality easily.
3. It builds a valued outranking upgrade that reflects the preference intensity.
4. It provides the decision-maker with a complete and partial ranking of the alternatives to choose.

We chose the PROMETHEE II method because it deals with a large number of alternatives, whereas the other methods such as AHP treat a limited set of alternatives [38].

In the next section and in the decision-making processes, the system must provide decision-makers with several tools to help them to negotiate. This latter is realized by a protocol which makes it possible to propose a set of solutions (alternatives) with an order relationship between various proposals. This modeling step is chained by a data analysis method and a similarity process that we explain in section VI.2. It is generally carried out with time constraints.

B. Classification and Similarity Process

We summarize our concept and our approach in several phases:

- *Phase1*: the integration of the decision maker's preferences via the web application in the PROMETHEE multicriteria method. This phase is independent from the others because it is autonomous as each decision maker has its own storage vector.
- *Phase 2*: The clustering phase, in which a k-means clustering algorithm is applied. The k-means is executed on the data set without taking into consideration the subjective nature of the problem.

- *Phase 3*: The classification of criteria based on preference relationships generated by phase 2 i.e. an application of the decision tree that allows us to have order of a ranking by a tree relationship with the notion of the greatest possible gain.
- *Phase 4*: The verification of the quality of the classification's results by an agent's election (mediating agent), this latter controls the negotiation process through several simulations.

The first phase is realized by the PROMETHEE method. In the following phase, we explain the impact of the clustering phase on the election process.

1. Clustering and Election Process

We process the clustering method (phase2) according to the following constraints:

Either a clustering problem of the performance matrix defined by:

A = {a1, a2 ... am} a set of objects (alternatives).

G = {g1, g2 ... gn} a set of criteria associated with set A.

The problem is to find a clustering method that represents the objects in set A. To do this, we build:

1. A performance matrix with quantitative parameters.
2. C = {1, 2….n} a group category set defined on all objects. This set is categorized by classes denoted by numbers (c1, c2, c3…….cn) and obtained by applying a clustering method.

In order to find a clustering method, we consider the problem of clustering as a problem of choice between several criteria which better represent the set of solutions (objects or alternatives). As a consequence, we chose the K-Means Method that aims to group a set of solutions together in the form of labeled classes [32]. Fig. 5 shows the clustering of the performance matrix (PM) processed by k-Means.

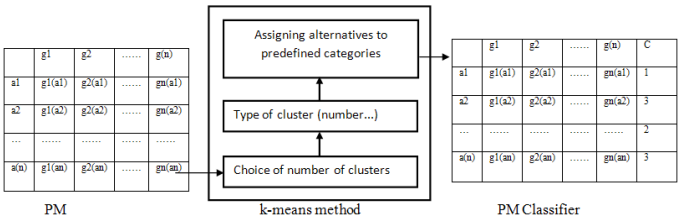


Fig. 5. The application of a clustering method.

Why K-Means

Different clustering methods and algorithms have been proposed and developed in literature. Studies have shown that no method is better than others on all clustering problems [33]. In this context, the K-Means algorithm is one of the best known and most used algorithms. It is based on an iterative process that is easy to program and adapt. However, the main advantage of this technique is its linear complexity in relation to the number of objects to that will be treated.

The objective of the next steps (phases 3 and 4) is to propose a new formulation of the coordination problem in GDSS using negotiation concepts. In this section we firstly define the application of decision trees on all alternatives and criteria. Then, we detail the two main developed contributions in this article to solve this type of problem, namely as the similarity process and the election of the coordinator agent. Finally, we describe the negotiation protocol.

2. Decision Tree

Is the classification phase of the criteria (phase 3). It is a question of finding partitions that better represent individuals (solutions) and model all the attributes (criteria) in the form of a decision tree. We take a set of classified data as an input (performance matrix processed by the K-Means Method), and we provide on output a tree that looks like an orientation diagram.

ID3 is one of the reference algorithms and the most used in the supervised classification type. The ID3 is founded on information theory research [35], this algorithm is based on examples already classified in a set of classes (result of the K-Means Method) that produces a decision tree [34].

Why ID3

We chose the ID3 algorithm in the design of our system because the popularity of this algorithm rests on its simplicity, especially from the point of view of the ease of implementation [36].

3. Similarity Process and Election in a Multi-Agent Environment

The objective of the process is to use the knowledge gained in the result of the decision tree to improve the quality of the final result. Indeed, the use of a reference vector for a comparison between several rankings increases the robustness of the solution. This happens by avoiding the problems related to the outsourcing of coordination and the processing related to the classifications of the proposals to reach the final solution. The similarity process consists of:

- Selecting the root of the decision tree (top node) that represents the most important attribute.
- Assigning a sort order of more to less important of the set of examples (alternatives) with respect to the selected attribute.
- Defining a reference vector called the initial solution vector that contains examples (alternatives) generated by the ranking order assigned to the solutions.
- Comparing a decision-maker's ranking vector with the initial solution vector, a similarity index becomes more important when the two vectors are similar in terms of ranking solutions.
- Choosing a ranking vector that corresponds to a decision-maker by the greatest similarity assessment.

In this step, the process carries out a series of evaluations of each ranking (triggered the fault manager in case the decision maker leaves the negotiation). This process is able to manage the failure by choosing the second ranking vector. This latter corresponds to the best second similarity index. It is an operational process that involves the selection of a ranking for an election of a decision-maker who has the most appropriate storage. Fig. 6 illustrates the selection of the ranking most similar to the initial solution vector.

Our objective is not to reproduce the operation of the similarity process automatically to choose a ranking vector if the process finds two or more appropriate ranking vectors, but to propose a framework for the realization of coordination between the decision-makers by an election of a ranking that models a sequence of proposals in a negotiation process.

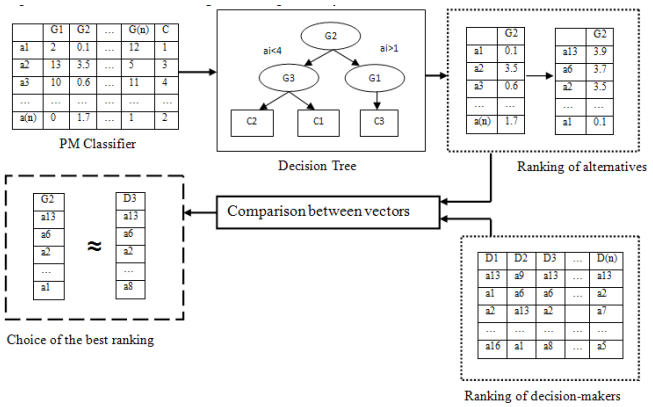


Fig. 6. Similarity Process.

Fundamentally, our approach is to develop coordination parameters that consider the choice of a ranking vector as a reference of choice of a decision-maker. This latter ensures negotiation, not in a direct way, but by a multi agent model that provides a detailed view of decision support process.

Having a multi-agent model integrated into the group decision support system provides a management of a negotiation protocol (described and detailed in section C.3). The MAS allows to model each decision-maker by an agent and to select a coordinator agent, who has the best ranking treated by the similarity process. Besides, the selected agent (coordinator agent) mentioned above is able to:

- Ensure a smooth running of the negotiation
- Use the functional capacities of a coordinator in a suitable way in a group decision support process.
- Control the behavior of the group through time management and provide solutions to reach a global compromise.

A Multi-Agent System (MAS) is an essential tool to support a cooperative work in a decision support system. In fact, the MAS brought by a GDSS exists in the interaction between the agents through negotiation. Furthermore, the decision makers find the satisfactory solution by making acceptance checks on the machine in order to see the results. The following section describes this notion by explaining the sequence of the negotiation process.

4. The Negotiation Protocol

In a multi-agent system, negotiation is characterized by a coordination mechanism that allows directing the actions (behaviors) of agents while leaving them freedom of choice by using an appropriate protocol. Our logic of unfolding the negotiation protocol is inspired and based on a protocol proposed by Hamdadou [23], in which it is divided into two notions, the *protocol description* and *negotiation process*:

Protocol Description: this stage of the group decision support process is the stage in which solutions (alternative) are sorted. We call this a production phase (compromise) and problem solving. This process is the fact of selection of the proposals by the coordinator (elected agent) before submitting the proposed solutions to all the agents (decision-makers).

The generation of an acceptable compromise to all agents must be based on a dynamic of exchange (a confirmation sequence) between the agents (MAS) and decision makers distributed, to not exceed the temporal deadlines.

Negotiation Process: in our work, a simpler approach is proposed. This approach consists of proposing solutions (alternatives) by the coordinator of his own ranking vector. Other agents are considered autonomous, as well as they always try to satisfy their desire in order to achieve the overall objective.

Negotiation is considered as a succession of sending messages governed by different steps as described below:

- In the first step, the coordinator has the proposals of the ranking vector that corresponds to the decision-maker chosen in the similarity process. He can decide how agents can organize themselves to respond to proposals. Thereafter, the coordinator also has the power to trigger the fault manager in the event that an agent's response is no longer valid (exceeding a deadline).
- In the second step, the negotiation process begins with the coordinator agent, who sends the best proposal at the respective time (the first solution located at the first location in the ranking vector and so on).
- In the third step the coordinator evaluates the answers provided by the other agents; he counts the number of participating agents who have accepted his proposal. If this number is greater than or equal to a threshold, then the negotiation is successful. Else the negotiation continues with several iterations of exchange of proposals and counter-proposals.
- The last step consists of a compromise at a coordinator level, As soon as all the answers of the participating agents are received; it communicates the best proposal to all the participants who are present in the negotiation process.

These last two steps are repeated until only one proposal remains, or the available time to negotiate has elapsed.

The coordinator has three features:

1. The first is to propose to each agent a solution (alternative), and to evaluate the time intervals for accepting responses by triggering the time manager.
2. The second is to measure the responses sent by the participating agents by a threshold. Also, the coordinator has the power to trigger the fault manager in order to eliminate failed agents. If the coordinator falls down, the fault manager tries to reactivate the similarity process.
3. The last feature is to decide which proposal satisfies all agents based on their rankings, and the time limit set for the finalization of the negotiation managed by the time manager. We detail the impact of time in a multi agent system in the next section (Time Management).

Thus, we can notice that the participating agent manages two types of conversations driven by the Negotiation Protocol. This conversation is to allow each agent to accept or reject the proposal sent by the coordinator agent. Furthermore, to be able to answer the proposal, the agent consults his ranking vector. If this proposal is in the first row (the first half) of its ranking vector, then he accepts it by sending an acceptance message to the coordinator agent, if not, he sends a refuse message.

The negotiation process ends when the final solution is validated by the coordinator agent, in this case this coalition for a compromise is represented as a recommendation registered in the knowledge base and decision makers can visualize this result.

C. Time Management

Claude Duvallet [21] proposed an agent-based decision support system whose behavior is anytime. The acquisition of behavior takes place at two levels:

1. The local level: this is the agent level. The latter is endowed with anytime behavior.
2. The global level: it is the multi-agent system level, which must also have the anytime behavior.

The modular architecture of our system has been chosen as an agent-based architecture. It is then extended with additional modules to realize a distributed architecture with a negotiation protocol by adding a time management inspired by anytime behavior. The coordination agent has a threshold function, which counts the number of proposals that must not exceed a certain number (NbAT). Besides, when the threshold (number of messages exchanged) is reached by one agent, the other agents must synchronize in order to launch the reelection of another coordinating agent; i.e. the coordinator has proposed several solutions without reaching an acceptable solution to all agents. Moreover, a group of agents that triggered re-election are the candidates for a new election. For this purpose, choosing a new coordinator needs the selection of an agent that has the second high similarity threshold.

For a given agent, when its alert threshold (AT) function indicates that its communications have exceeded this threshold *(1)*, he will then have two possibilities (See Fig. 7):

1. He sends an alert message to the coordinator *(1)*, which counts the number of alerts (NbAT). If this number is greater than half plus one, the coordinator triggers the fault manager *(3)* to restart the re-election *(4) (5)*.

2. He notifies the simulation manager *(2)*, so that the system sends a message to the concerned decision maker to validate his choice immediately *(6)*. If not, the agent that sends the alert is considered as a failing agent (he leaves the negotiation).

When an agent has to solve a problem, he invokes the time manager to initiate the alert threshold (for each agent). He can obtain an answer (end of negotiation) whose quality is not optimal (ranking of the final solution is not among the first ranked in its vector). In the case of having or receiving an alert, the concerned decision maker (the agent who represents the decision maker) must make his choice in relation to the solutions proposed by the coordinator to obtain the final result. Fig. 7 illustrates the role of each management mechanism in the proposed system.

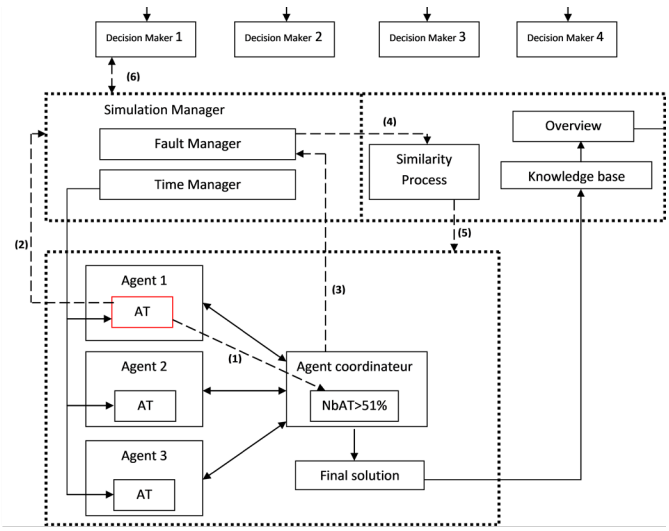


Fig. 7. Operation of the proposed system.

For each function of a mechanism, we define an associated formula. Table IV presents a summary of the different formulas used by the agents.

TABLE IV. Table of Measuring The Functioning of the Mechanisms

Appellation	Formula	Acronyms
Alert Threshold (triggering case)	$AT = \sum_{i=1}^{n} Pro(i)$ $\textbf{\textit{if}}\left(AT > \left(\frac{NbrAl}{2} \right) \right) \textbf{\textit{then}}$ *alert*	*AT* :Alert Threshold *Pro* : Proposal *NbrAl*: Number of Alternatives
Re-election alert	$NbAT = \sum_{i=1}^{n} AT$ $\textbf{\textit{if}}\left(AT > \left(\frac{n}{2} + 1 \right) \right) \textbf{\textit{then}}$ *re election*	*NbAT* : Number of Alert thresholds triggered by agents. *n* : number of agents
Overall time estimated by the time manager	$NT = \sum_{i=1}^{n} TPro(i)$	*NT* :Negotiation Time *Tpro* : Time of a proposal *n* : total number of solutions

Observing the behavior of the system will allow decision-makers to anticipate decision-making when situations of failure arise in a negotiation. This allowed us to become aware of different attractions and proposals in order to interpret the results. The estimated time allowed us to control the reactivation of functions and eliminate an iteration loop of consensus.

In order to support the representation of this approach of coordination, a UML modeling was used. It is the most useful tool for MAS modeling. Additionally, for a good representation of the different interactions between the coordinator (the elected agent) and the agents of the system, we propose a sequence diagram representation. For this diagram, the corresponding language is defined by primitives of the negotiation protocol, proposed from an agent to N agent:

Coordinating agent: sends messages to all agents:

- REQUEST (): the coordinator sends a message to the participants to indicate the beginning of the negotiation process, i.e. requesting or initiating negotiation.
- PROPOSE (): the coordinator proposes an agreement (contract) to all participating agents concerning a given solution.
- CONFIRM (): the coordinator sends a message to all agents to inform them that the negotiation is successful and the best alternative has been found.

Participating agent: The messages sent by the participants target the coordinator only. The other participants are not informed of these messages:

- INFORM (): after establishing a ranking of alternatives from the best to the worst (ranking of alternatives), each participating agent sends a life message to the coordinator (ready to negotiate).
- AGREE (): each participant indicates to the coordinator by this message that he accepts or agrees to this contract (solution proposed by the coordinator is to accept).
- REFUSE (): the participant indicates to the coordinator that his proposals are rejected. The agreement cannot be concluded in its current form and should be amended.

Fig. 8 illustrates the sequencing of these interactions via a UML diagram and Fig. 9 show a use case of our system.

VI. Case Study

The application of our distributed decision support architecture requires an infrastructure communication and simulation platforms for giving the geographically dispersed aspect of decision-makers who may interact with the system in the decision-making process. In a distributed context, where decision-makers are dispersed, the architecture of a GDSS consists of modeling a geographically dispersed structure or organization of a distributed decision support system. This allows introducing the preferences of a decision-maker via a web application, whose objective is to reproduce the behavior of a group modeled by a MAS in order to come to a collective decision about a problem. The tools used to implement our approach are:

1. *Multi Agent System:* The multi-agent platform "JADE" allows the realization of intelligent agents and to program their behaviors. It has a java interface that illustrates the decision-making process.
2. *Java Application:* uses a program (Application) in which the simulation can be run to perform the negotiation process. This application was developed with the "NetBeans" environment based on the "Java" language.
3. *User Web Interface (Decision Makers):* allows the interaction and introduction of the preferences (subjective parameters) of the decision-makers via a web page. This interface was developed by the "DREAMWEAVER" tool based on the "HTML" language.

In the following, we describe step by step the operation and progress of our approach through a case study.

The spatial group decision support system (SGDSS) proposed in this article focuses on the decision-making problem of choosing the most appropriate zone that best meets the needs of all decision-makers.

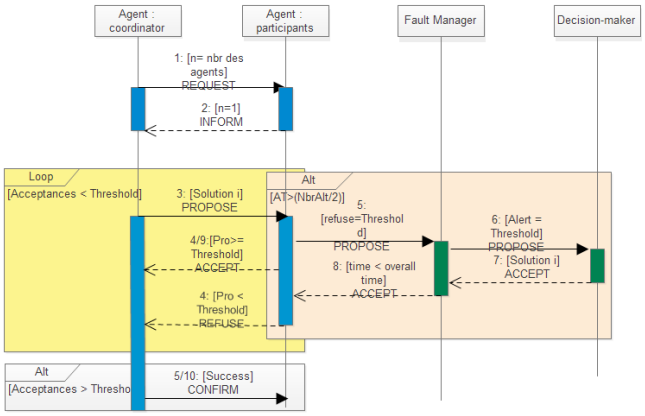


Fig. 8. Sequence diagram.

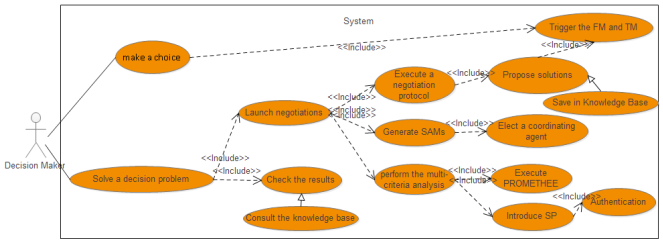


Fig. 9. Use case diagram.

As a consequence, our work is interested in the choice of a solution which satisfies most of the preferences of the decision-makers. For this reason our system is divided into two parts: Users part and Server part. The first consists of the web pages, where each decision maker has a registration window that represents his preferences, visualizations of results, and interactions. The second concerns the simulation of negotiation and processing for choosing the solution. This part is composed of several tasks, where each task is an independent treatment that will be used for negotiation thereafter.

A. The Addressed Decision Problem

For this case study, we relied on the work of Joerin [37] and taken up by Hamdadou in [23], 650 virgin ilots (alternatives) were proposed. The study area is located in the Canton of Vaud about 15 km North of Lausanne. The area of this zone is 52.500 km2. Its geographical limits in the Swiss coordinating system are 532 750-532 500 (m) and 158 000-164 000 (m). In this study, we also identify the diversity of factors (environmental, social, economic, etc.). It seems wise not to aggregate all of them in a single criterion. The land relevance for habitat characteristics for this application are seven criteria, namely: damage, noise, impacts, geotechnical, natural risks, equipment, accessibility, and climate. Table V describes in detail the different criteria considered in this study.

TABLE V. Description of Identified Criteria

Criteria	Type	Factors
Harm	Natural	Pollution of the air and odors
Noise	Social	Highways and railways
Impacts	Social	Groundwater and sectoral
Geotechnical and natural risks	Natural	Constraints, landslides, floods, earthquakes, fires
Equipment	Economic	Distance to gas, electricity, water, roads
Accessibility	Social	Average travel time between home and work
climate	Natural	Sun, fog, temperature

The problem addressed in this case study is related to the choice of a solution (alternative) that represents a land for housing. Indeed, our work is interested on the most appropriate choice of a zone for the construction of a dwelling.

The alternatives (actions) correspond to the objects of the negotiation.

The definition and evaluation of the criteria are identified according to the different actions that generate the matrix performance. This is illustrated in Fig. 10.

N°	ID_ZONE	HARM	NOISE	IMPACTS	GEOTECH	EQUIP	ACCESS	CLIMATE
1	202	1,00	0,88	0	1	816	8	0,92
2	209	1,00	0,45	0	1	1249	9	0,91
3	210	1,00	0,69	0	1	1165	9	0,89
4	211	1,00	0,48	0	1	1518	9	0,92
5	213	1,00	0,92	0	1	1356	9	0,89
6	215	1,00	1,00	0	1	1434	8	0,75
7	216	1,00	0,97	0	1	1490	10	0,83
8	218	1,00	1,00	0	1	1556	8	0,70
9	219	1,00	1,00	0	1	1638	12	0,68
10	220	1,00	1,00	0	1	1629	8	0,68
11	221	1,00	0,95	0	1	1641	10	0,84
12	223	1,00	1,00	0	1	1697	8	0,68
13	224	1,00	0,98	0	1	1758	10	0,70
14	225	1,00	1,00	0	1	1801	8	0,67
15	226	1,00	0,91	0	1	1809	10	0,84
16	228	1,00	1,00	0	1	1840	8	0,67
17	229	1,00	0,97	0	1	1870	10	0,68
18	230	1,00	0,09	0	1	1848	12	0,55
:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:
644	9516	1,00	1,00	6	6	1568	6	0,69
645	9517	1,00	1,00	6	6	1569	8	0,67
646	9519	0,57	0,82	6	6	1589	10	0,47
647	9525	1,00	0,98	6	6	1766	12	0,07
648	9534	1,00	0,26	6	6	1912	11	0,39
649	9548	1,00	0,14	6	6	2240	12	0,66
650	9550	0,00	0,03	6	6	2012	10	0,54

Fig. 10. The performance matrix.

B. Identification of Decision-Makers

In this study, the different decision makers involved in the group decision are:

- Decision Maker 1: Environmental associations.
- Decision Maker 2: Politician.
- Decision Maker 3: Economist.
- Decision Maker 4: Public.

Each decision-maker is represented by an agent; the creation of agents is performed using the MAS JADE platform (JAVA). We attribute to each participating agent a weight in order to express his importance in the negotiation process.

C. Simulation of the Negotiation Process

1. At the User Level

Registration of the Decision Makers: they can subscribe to the web page, a password and username is assigned to each member of the group. Each decision-maker has a window that allows him to:

- Introduce the subjective parameters.
- View input data: Performance Matrix.
- Receive confirmation messages.
- Evaluate the results.

Fig. 11 and Fig. 12 show, respectively, the different functions assigned to the decision-maker in his web page.

Fig. 11. Home page of a decision maker.

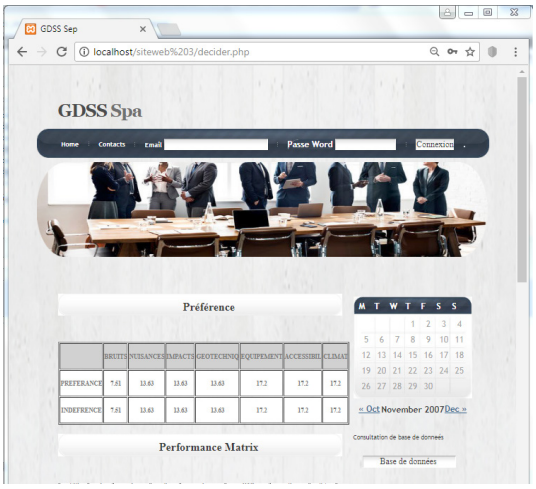
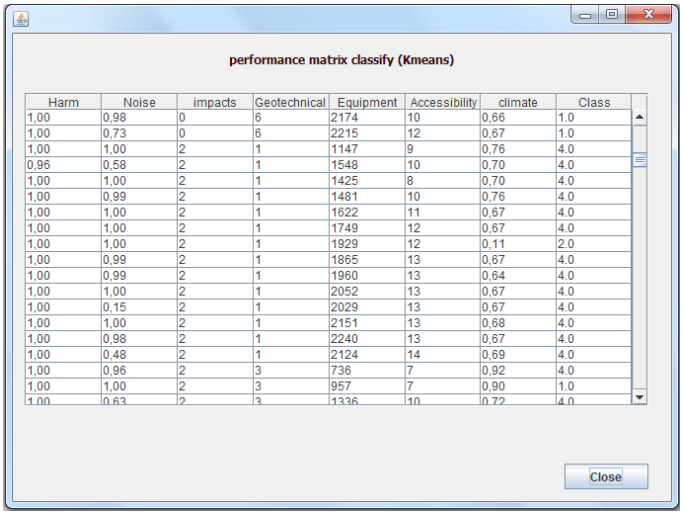


Fig. 12. Inputting subjective parameters.

2. At the Server Level

In this section, we will illustrate our proposal of the group decision support model, so beginning with the organization and creation of the necessary data for the negotiation simulation. Besides, this organization is to classify the matrix performance through using the K Means method, as shown in Fig. 13.



Harm	Noise	impacts	Geotechnical	Equipment	Accessibility	climate	Class
1,00	0,98	0	6	2174	10	0,66	1.0
1,00	0,73	0	6	2215	12	0,67	1.0
1,00	1,00	2	1	1147	9	0,76	4.0
0,96	0,58	2	1	1548	10	0,70	4.0
1,00	1,00	2	1	1425	8	0,70	4.0
1,00	0,99	2	1	1481	10	0,76	4.0
1,00	1,00	2	1	1622	11	0,67	4.0
1,00	1,00	2	1	1749	12	0,67	4.0
1,00	1,00	2	1	1929	12	0,11	2.0
1,00	0,99	2	1	1865	13	0,67	4.0
1,00	0,99	2	1	1960	13	0,64	4.0
1,00	1,00	2	1	2052	13	0,67	4.0
1,00	0,15	2	1	2029	13	0,67	4.0
1,00	1,00	2	1	2151	13	0,68	4.0
1,00	0,98	2	1	2240	13	0,67	4.0
1,00	0,48	2	1	2124	14	0,69	4.0
1,00	0,96	2	3	736	7	0,92	4.0
1,00	1,00	2	3	957	7	0,90	1.0
1,00	0,63	2	3	1336	10	0,72	4.0

Fig. 13. Clustering of alternatives by the Kmeans method.

The generation of the decision tree by a hierarchy of criteria is calculated from the clustered matrix performance. Fig. 14 shows the first level of the tree (root) that will be useful later.

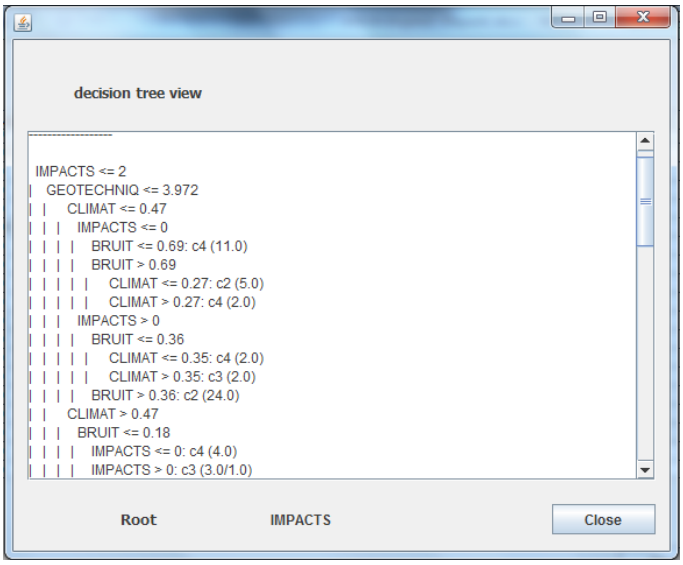


Fig. 14. Creating the decision tree using the ID3 algorithm.

A set of agents is set up to reproduce the behavior of decision-makers. This latter consists of a coordinator who is elected by the system (responsible for the negotiation), and a set of participating agents.

Generating preference vectors by using PROMETHEE II (multi criteria analysis method): in Fig. 15 and Fig. 16, we show a series of experiments' illustrations that demonstrate how our system attributes subjective values and parameters to each agent involved in the group decision to begin negotiation.

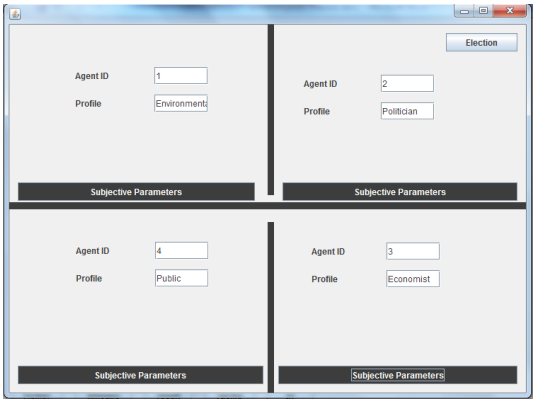


Fig. 15. Identifications of the parameters of each agent.

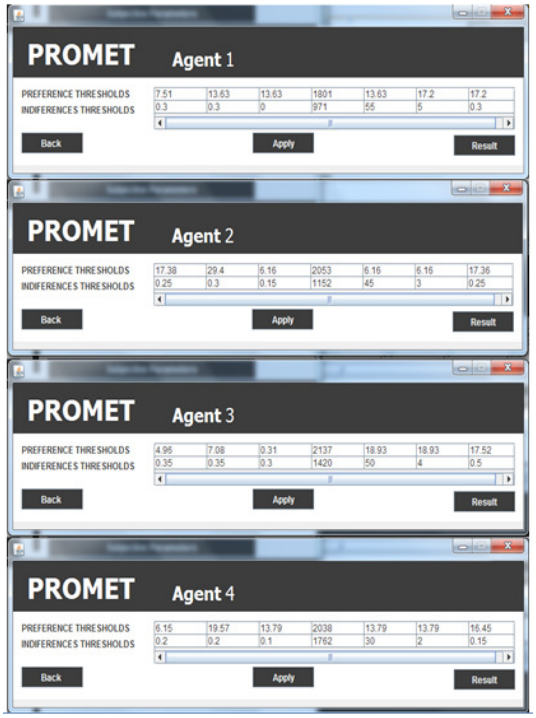


Fig. 16. Subjective parameters expressed by each decision-maker.

Loading input data: Performance matrix, subjective parameters allow calculating the ranking of alternatives (preference or ranking vector) of each decision-maker by using the PROMETHEE II method. An example of the result of this method is shown in Fig. 17.

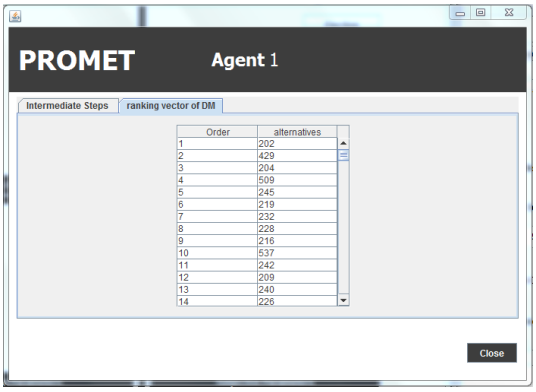


Fig. 17. Ranking calculated by the PROMETHEE II method.

The similarity process: runs to select a coordinator after generating each decision maker ranking. Fig. 18 illustrates the ranking of alternatives (calculated from the root), and the result of the similarity process.

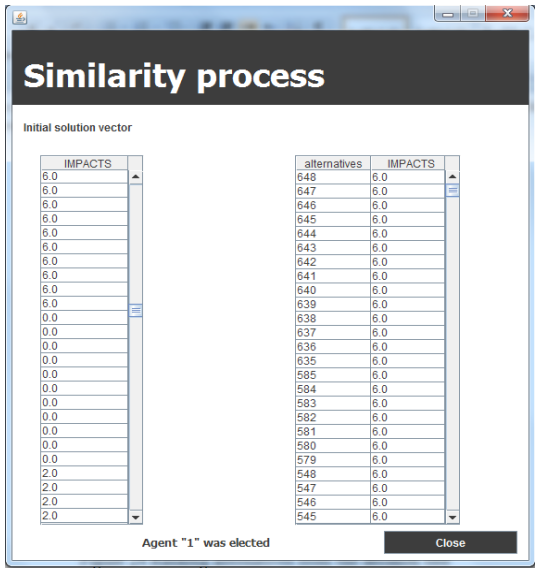


Fig. 18. Ranking alternatives from the decision tree and choosing the coordinator.

Election of the coordinator: it can be seen that agent '1' has been chosen as the coordinator shown at the bottom of the interface in Fig. 18.

Negotiation protocol: In our study, an acceptance threshold is set at (70%). If the majority accepts a proposal, it means the solution is chosen. For this purpose, the system signals the end of the negotiation and that the solution of the problem has been found. The different messages exchanged between the coordinator and the participants during the negotiation process are shown in Fig. 19 and provided by the functionality of the SMA module (the sniffer agent).

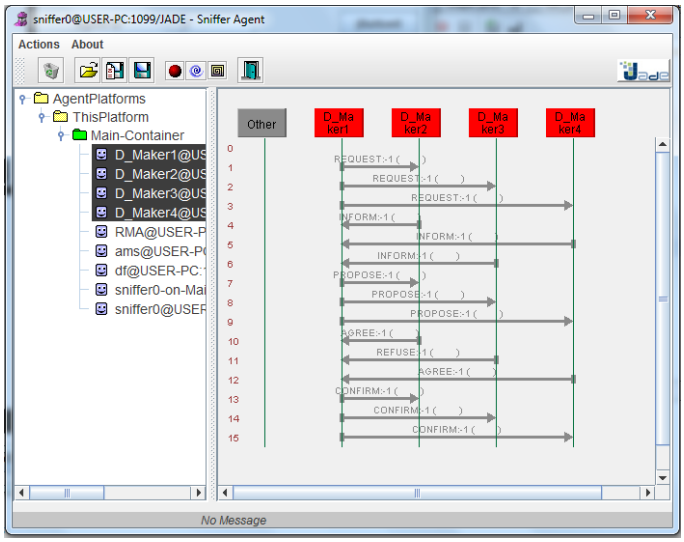


Fig. 19. The messages exchanged during the negotiation process via the sniffer.

The Group's final decision: As soon as the ultimate alternative is found, the participating agents reach a consensus. The alternative chosen is 202 as shown in Fig. 20 with a high acceptance rate. In addition, the server (system) sends the negotiation's results to the decision makers via the web interface shown in Fig. 21.

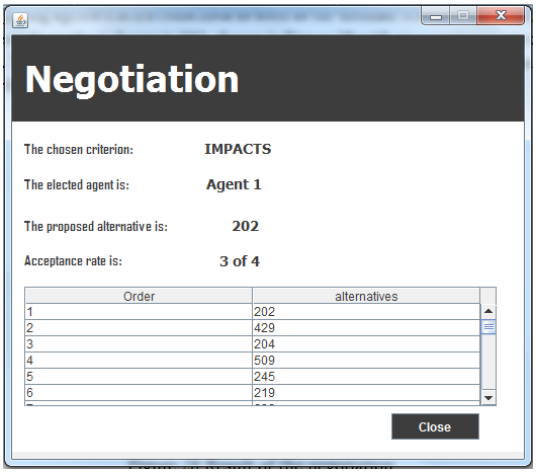


Fig. 20. Result of the negotiation.

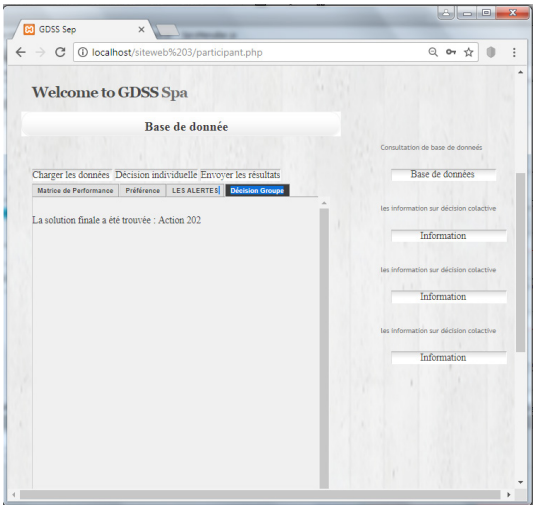


Fig. 21. Final result sent to decision-makers.

The values of the subjective parameters expressed by each decision maker are involved in the group decision and the matrix performance. Eventually, the results are stored in a database as shown in Fig. 22.

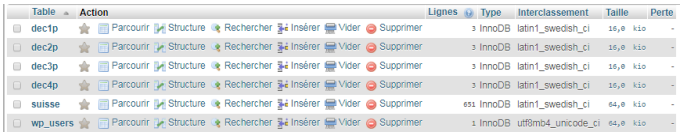


Fig. 22. Example of the database.

In a situation that there is not an ultimate acceptable solution according to the majority of the concerned decision-makers, the system triggers the re-election of a new coordinator. We can have a case of failure if:

- The alternatives have been exhausted.
- Re-election number is equal to the number of decision-makers (each decision-maker has become a coordinator).
- The estimated time for the negotiation is invalid (exceeding the deadline).

In this case we propose other strategies to solve this problem in order to have a deterministic protocol, such as the monotonous concession and game theory that can make the negotiation protocol more efficient.

VII. Conclusion and Future Works

In this article, we have uncovered a new research track in the field of multi-criteria decision support. We proposed a decisional group model based on a multi-agent system modeling a spatial problem.

We have developed a new approach that integrates:

- Data analysis based on a clustering method, which allows the management and manipulation of resources, in order to optimize the quality of negotiation in a spatial context.
- Multi-criteria analysis based on a multi-agent model, which reproduces the behavior of decision-makers in order to respond to the multiplicity and diversity of both the criteria and the decision-makers.

Several techniques and strategies are elaborated in this article, in order to enrich our approach which allows:

- The representation of the multiplicity of actors (decision-makers), their diversity, their behaviors and their interactions.
- The interaction of the different decision-makers facing a decision-making problem via a web application.
- The Integration of time management mechanisms during the negotiation process.
- We end this conclusion by noting the different research perspectives, which we intend to resolve in the future:
- Integration of GIS (Geographic Information System) and their functionality.
- Working on spatial data in real time.
- Development and design of a GDSS with distributed agents that operate in real time.
- Integration of other negotiation strategies between different agents.

References

[1] Boukherroub, T., Fondrevelle J., and GUINET, A. Multi-criteria decision making for the supply chain design: A review with emphasis on sustainable supply chains. 4th International Conference on Information Systems, Logistics and Supply Chain, ILS 2012.

[2] Doumpos, M. and Zopounidis C. “A multicriteria decision support system for bank rating.” Decision Support Systems 50(1): 55-63, 2010.

[3] Natividade-Jesus, E., J. Coutinho-Rodrigues, E., and Antunes, C. “A multicriteria decision support system for housing evaluation.” Decision Support Systems 43(3): 779-790, 2007.

[4] Coutinho-Rodrigues, J., Simão, A, and Antunes, C. H. “A GIS-based multicriteria spatial decision support system for planning urban infrastructures.” Decision Support Systems 51(3): 720-726, 2011.

[5] López, J. C. L., Carrillo, P. A. Á., and Valenzuela, O. A. “A Multicriteria Group Decision Model for Ranking Technology Packages in Agriculture.” Soft Computing for Sustainability Science 358: 137, 2017.

[6] Yu, L. and Lai K. K. “A distance-based group decision-making methodology for multi-person multi-criteria emergency decision support.” Decision Support Systems 51(2): 307-315, 2011.

[7] Cao, P. P., Burstein, F., and San Pedro, J. “Extending Coordination Theory to the Field of Distributed Group Multiple Criteria Decision-Making.” Journal of Decision Systems 13(3): 287-305, 2004.

[8] Álvarez Carrillo, P. A., Leyva López, J. C., and Sánchez Castañeda, M. D. L. D. “An Empirical Study of the Consequences of Coordination Modes on Supporting Multicriteria Group Decision Aid Methodologies.” Journal of Decision Systems 24(4): 383-405, 2015.

[9] López, J. C. L., Carrillo, P. A. Á., and Chavira, D. A. G., and Noriega, J. J. S. “A web-based group decision support system for multicriteria ranking problems.” Operational Research: 1-36, 2016.

[10] Marjorie LE BARS. “Un Simulateur Multi-Agent pour l’Aide à la Décision d’un Collectif: Application à la Gestion d’une Ressource Limitée Agro-environnementale” thèse doctorat, UNIVERSITE PARIS IX-DAUPHINE, 2003.

[11] Ai, W.-G., Sun, J., and Li, H. A distributed marketing decision support system based on multi-intelligent-agent. Machine Learning and Cybernetics. Proceedings of 2004 International Conference on, IEEE, 2004.

[12] Hamdadou, D., Bouamrane, K. A Multicriterion SDSS for the Space Process Control: Towards a Hybrid Approach. In: Gelbukh A., Kuri Morales Á.F. (eds) MICAI 2007: Advances in Artificial Intelligence. MICAI 2007. Lecture Notes in Computer Science, vol 4827. Springer, Berlin, Heidelberg, 2007.

[13] Zemri, F. A. and Hamdadou D. Integration of Multi Criteria Analysis Methods to a Spatio Temporal Decision Support System for Epidemiological Monitoring. ICAASE, 2014.

[14] Hamdadou, D., Bouamrane, K. A Spatial Group Decision Support System: Coupling Negotiation and Multicriteria Approaches. Intelligent Decision Technologies, vol. 10, no. 2, pp. 129-147, 2016.

[15] Bui, T. X. and Shakun M. F. "Negotiation processes, evolutionary systems design, and negotiator." Group decision and negotiation 5(4-6): 339-353, 1996.

[16] Osborne, M. J. and Rubinstein A. A course in game theory, MIT press, 1994.

[17] Lepperhoff, N. "SAM-Simulation of computer-mediated negotiations." Journal of Artificial Societies and Social Simulation 5(4), 2002.

[18] Rosenschein, J. S. and G. Zlotkin. Rules of encounter: designing conventions for automated negotiation among computers, MIT press, 1994.

[19] Davis, R. and Smith R. G. "Negotiation as a metaphor for distributed problem solving." Artificial intelligence 20(1): 63-109, 1983.

[20] Axtell R. "Why Agents? On the Varied Motivations for Agent Computing in the Social Sciences" Santa Fe University. Broking Institution: Center on Social and Economics Dynamics, 22 Pp, 2000.

[21] Duvallet. C, « Des systèmes d'aide à la décision temps réel et distribués: modélisation par agents », PHD Thesis, University of Havre, 2001.

[22] Simos, J. Evaluer l'impact sur l'environnement: Une approche originale par l'analyse multicritère et la négociation. Presses polytechniques et universitaires romandes, 1990.

[23] Hamdadou D. Un Modèle d'Aide à la Décision en Aménagement du Territoire, une Approche Multicritère et une Approche de Négociation, PHD Thesis, University of Oran, 2008.

[24] ROUBA B. «Minimisation des désagréments dans les clusters agrégés », Thesis in computer science, Alger, University of Oran 1 Ahmed Benbella, 2015.

[25] Hickey, R. and Jankowski P. "GIS and environmental decisionmaking to aid smelter reclamation planning." Environment and Planning A 29(1): 5-19, 1997.

[26] Brans J.P,»Elaboration d'instruments d'aide à la décision: méthode PROMETHEE», Université Laval, Québec, p 5, 1982.

[27] NACHET B. Modèle multi-agents pour la conception de systèmes d'aide à la décision collective, Thesis, University of Oran, 2014.

[28] Lander S. E. and Lesser. V. R. "Understanding the Role of Negotiation in Distributed Search Among heterogeneous Agents" International Joint Conference on Artificial Intelligence (IJCAI-93), 1993.

[29] Sycara K. "The PERSUADER" The Encyclopedia of Artificial Intelligence, Shapiro J. (Eds.) New York, 1992.

[30] Lemiere, B., Michel, P., Abriak, N. E., Haouche, L., Laboudigue, A., Alary, C. and all. "The GeDSeT project: constitution of a decision support tool (DST) for the management and material recovery of waterways sediments in Belgium and Northern France." Déchets, Sciences et Techniques 57: 10 p, 2010.

[31] Taibi Boumedyen M. L'analyse Multicritère comme outil d'aide a la décision : Application de la méthode PROMETHEE. PHD thesis, University Abou-Bekr-Belkaid Tlemcen, 2009-2010.

[32] McQueen, J. "Some methods for classification and analysis of multivariate observations". Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, CA, USA, Vol.1, pp.281–297, 1967.

[33] Candillier, L. "Contextualisation, visualisation et évaluation en apprentissage non supervisé". PHD thesis, Charles de Gaulle-Lille 3 University, 2006.

[34] Quinlan, J. R. "Induction of decision trees." Machine Learning 1(1): 81-106, 1986.

[35] Shannon, C. E. "A mathematical theory of communication." ACM SIGMOBILE Mobile Computing and Communications Review 5(1): 3-55, 2001.

[36] Rakotomalala, R. «Arbres de décision.» Revue Modulad 33: 163-187, 2005.

[37] Joerin.F, Décider sur le territoire: proposition d'une approche par utilisation de SIG et de méthodes d'analyse multicritère, PHD Thesis, University of Lausanne, 1997.

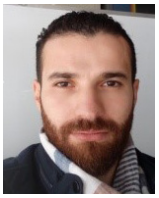
[38] Taibi, A. and Atmani B. "Combining Fuzzy AHP with GIS and Decision Rules for Industrial Site Selection." International Journal of Interactive Multimedia and Artificial Intelligence 4(6) : 60-69, 2017.

[39] Fraile, A., Larrodé, E., Magreñán, Á. A., and Sicilia, J. A. "Decision model for siting transport and logistic facilities in urban environments: A methodological approach." Journal of Computational and Applied Mathematics 291: 478-487, 2016.

[40] Fraile, A., Sicilia, J. A., González, R., and González, A. "Study of Factors Affecting the Choice Modal of Transportation in an Urban Environment Using Analytic Hierarchy Process". Workshop on Engineering Applications, Springer, 2016.

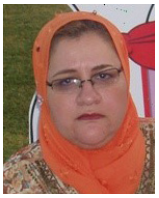
[41] Djamila, H. "A MultiCriteria Group Decision Support System for Industrial Diagnosis." INFOCOMP 10(3): 12-24, 2011.

[42] Sicilia, J. A., Quemada, C., Royo, B., and Escuín, D. "An optimization algorithm for solving the rich vehicle routing problem based on Variable Neighborhood Search and Tabu Search metaheuristics." Journal of Computational and Applied Mathematics 291: 468-477, 2016.



Sofiane Khiat

Sofiane khiat received his Master degree in Information System Technologies at University of Oran1 Ahmed BenBella, Algeria in 2013. Currently, he is a Ph.D. student in Computer Science Department at the same University. His research interests include Group Decision Support System, Artificial Intelligence, Multi-Agents System, Temporal Modeling and distributed systems.



Djamila Hamdadou

D. Hamdadou received her Engineering degree in Computer Science and her Master of Science degree from the Computer Science Institute in 1993 and 2000, respectively. She also obtained her doctorate in 2008. She received her PHD in 2012 from the Computer Science Department. Currently, responsible of the Research Team "Spatio Temporal Modeling and Artificial Vision: from the Sensor to the Decision" of the Laboratory of Computer Science of Oran LIO. Her research interests include Business Intelligence, Artificial Intelligence, and Spatio Temporal Modeling.

Gesture Recognition of RGB and RGB-D Static Images Using Convolutional Neural Networks

Manju Khari[1]*, Aditya Kumar Garg[1]*, Rubén González Crespo[2], Elena Verdú[2]

[1] Computer Science and Engineering, Ambedkar Institute of Advanced Communication Technologies and Research, Geeta Colony, Delhi (India)
[2] Universidad Internacional de La Rioja, Logroño (Spain)

Received 28 June 2019 | Accepted 10 September 2019 | Published 16 September 2019

Abstract

In this era, the interaction between Human and Computers has always been a fascinating field. With the rapid development in the field of Computer Vision, gesture based recognition systems have always been an interesting and diverse topic. Though recognizing human gestures in the form of sign language is a very complex and challenging task. Recently various traditional methods were used for performing sign language recognition but achieving high accuracy is still a challenging task. This paper proposes a RGB and RGB-D static gesture recognition method by using a fine-tuned VGG19 model. The fine-tuned VGG19 model uses a feature concatenate layer of RGB and RGB-D images for increasing the accuracy of the neural network. Finally, on an American Sign Language (ASL) Recognition dataset, the authors implemented the proposed model. The authors achieved 94.8% recognition rate and compared the model with other CNN and traditional algorithms on the same dataset.

Keywords

American Sign Language, Image Processing, CNN, Gesture Recognition.

DOI: 10.9781/ijimai.2019.09.002

I. Introduction

In Computer Accessibility, researchers keep on investigating new and assistive technologies for the people who suffer from disabilities. These technologies consist of devices and software that are intended to benefit the people who suffer from disabilities. Moreover these technologies may help users in performing tasks with much larger efficiency which in turn will help in increasing their quality of life. According to an estimate of World Health Organization (WHO) in 2019, there are about 466 million people who suffer from hearing loss [1]. In 2005, the number of deaf people was 278 million [2]. In around 15 years, there is a significant rise in the number of people having hearing problems.

Deaf people use numerous methods for communication purpose. One of these methods include Sign Languages that are made up of movements of the hands, torso, arms, head, facial expressions and eyes. Sign Language is a hand gesture language which is most common among the deaf people to be used as way of expressing their feelings, thoughts, and knowledge in the place of verbal communication. According to [3], there are more than 100 sign languages. Moreover, there is no a unified way of writing sign languages, being SignWriting a proposal on which some research work has been developed [4]. American Sign Language (ASL) is one of the most commonly used sign languages throughout the U.S. and Canada and also including Southeast Asia and West Africa. According to an estimate in [5], there are about 250,000–500,000 deaf people who rely on using American Sign Language (ASL). ASL fingerspelling consists of 36 signs and is also the sixth most used sign language in US.

* Corresponding author.
E-mail addresses: manjukhari@yahoo.co.in (M. Khari), adityagarg2607@gmail.com (A. Kumar Garg).

There are many works on gesture recognition for different purposes, some focusing on the whole body [6] while others focusing on a specific part as eyes [7] or hands [8]. Despite of the vast number of research works that have been published, there have been several limitations. Some of the limitations include: (1) many of the previous methods make use of add on devices, (2) most of the previous methods are based on getting more speed than getting a higher recognition rate, (3) most of the previous methods still make use of traditional learning algorithms that are based on feature extraction which requires high computation [9]. Gesture recognition can be classified into two different categories. The first one is electromagnetic gloves and sensors based detection and the other is Computer Vision based. The electromagnetic gloves and sensor based technique is very expensive and is not suitable for real life purposes [10]. On the other hand, the Computer Vision based technique can be further divided into Static Gesture Recognition and Dynamic Gesture Recognition. There are many challenges in the area of hand gesture recognition such as (1) feature extraction (2) variation in hand size, (3) hand partial occlusion.

In recent years, the field of deep learning is under rapid development. Particularly, Convolutional Neural Networks (CNN) are able to achieve far more effective results related to the field of Image Classification, Natural Language Processing, etc. With increasing popularity of CNN, many new CNN Models such as GoogLeNet [11], VGG16, VGG19 [12], and Inception V3 [13] have emerged and were able to achieve significant results in ImageNet Large-scale Visual Recognition Challenge (ILSVRC).

As these CNN models became more popular, a concept termed as Transfer Learning gained reputation. Transfer Learning is the transfer of parameters of a previously trained model to help the training of another model. The main advantage of Transfer Learning is to avoid the overfitting of models and also to reduce the time taken to train a

model. With the help of Transfer Learning, the model can be more easily converged. There are a huge number of research results that helped in confirming the performance of models trained with the help of Transfer Learning. For example, GoogLeNet Model which was pre-trained with ImageNet image library can be used to predict diabetic retinopathy [14].

The main contribution of this paper is to propose a Neural Network that will help in increasing the recognition rate on American Sign Language Dataset. The main focus of this paper is on Static Gesture Recognition based techniques. Presently, Static Gesture Recognition disadvantages include non-robust and inaccurate recognition under abrupt lightning changes and complex background. Inaccurate Gesture Segmentation in turn affects the accuracy of gesture classifications [15]. The proposed algorithm uses Transfer Learning, which not only helps in eliminating the need of feature extraction but also helps in reducing the computational power required to obtain a higher accuracy for Sign Recognition.

The rest of the paper is organized as follows. Section II describes the Related Work. Section III explains the proposed Work. Section IV describes the research methodology. Section V explains the results and comparative analysis. Section VI includes the conclusion.

II. Related Work

Pugeault and Bowden (2011) [16] used a Microsoft Kinect for the collection of Intensity and Depth Images of American Sign Language (ASL) 24 letters (except J and Z). On this dataset the authors used OpenNI + NITE framework for gesture detection and tracking. For extraction of features, the authors used a set of Gabor filters and then the classification was done using Random Decision Forest. The paper concluded that the average recognition rate is 73% when only Intensity Images were considered and 69% when only Depth Images were considered. The combined recognition rate is 75%.

Estrela et al. (2013) [17] proposed a framework on the basis of bag of features in combination with Partial Least squares (PLS).The authors split the experiment into two groups and computed results on the basis of ASL dataset. In the first group, the experiments compared Support vector Machine (SVM) classifiers and Partial Least Squares (PLS) classifiers. The accuracy achieved in the first experiment of the PLS and SVM classifiers is 66.27% and 62.85%, respectively. In the second experiment, BASE and SIFT feature descriptors are appraised. BASE feature extractor runs faster and consumed less memory while the SIFT feature extractor achieves a better accuracy. The accuracy with PLS is 71.51% and with SVM is 65.55%, respectively.

Chuan et al. (2014) [18] used a 3D motion sensor based system for implementing American Sign Language Recognition (ASL). The authors applied KNN and SVM for classification of 26 letters on the basis of features derived from the sensory data. According to the experiments, the result shows that the highest rate for average classification is 72.78% and 79.83% which was achieved by k-Nearest Neighbour (KNN) and Support Vector Machine (SVM) respectively. Rioux-Maldague and Giguère (2014) [19] presented a novel feature extraction technique which uses both the depth and intensity image that were captured from a Microsoft KnectTMsensor. Then, the authors used a Deep Belief Network on an American Sign Language dataset.

Ameen and Vadera (2017) [20] developed a CNN aimed for the classification of both the colour and depth Images. The CNN was applied to American Sign Language (ASL) database. The authors were able to achieve a precision equivalent to 82% and recall of 80%. Xie et al. (2018) [15] proposed a RGB-D Static Gesture Recognition based method using a fine-tuned Inception V3. In comparison to a traditional CNN, the authors adapted a two stage training strategy. The authors compared the proposed model with traditional methods and other CNN model. The highest accuracy authors were able to reach was 91.35%.

Dai et al. (2017) [21] proposed a SmartWatch-based American Sign Language (ASL) recognition system. The purpose of this system was to be more portable, comfortable and user friendly. The proposed system was designed such that each individual Sign having its own motion pattern can be transformed into accelerometer and gyroscope signals. In the next steps, these signals are analysed with the help of a Long-Short Term Memory Recurrent Neural Network (LSTM-RNN) trained with Connectionist Temporal Classification (CTC). Islam et al. [10] proposed a novel "K convex hull" method which is the combination of K curvature and convex hull algorithms. The "K convex hull" method developed is able to detect fingertip with high accuracy. In this paper, the system gathers ASL gesture images with black background and extracts mainly five features that are fingertip finder, pixel segmentation, elongatedness, eccentricity and rotation.

Tao et al. (2018) [22] proposed a Convolutional Neural Network (CNN) along with inference fusion and multiview augmentation. This method uses depth images captured by Microsoft Kinect. Chong and Lee (2018) [23] developed a prototype with the help of a Leap Motion controller (LMC). This study focussed on full ASL recognition i.e. all the 26 letters and 10 digits. The recognition rate for 26 letters with the help of Support Vector Machine (SVM) and Deep Neural Network (DNN) was 80.3% and 93.81% and for the Combination of 26 letters and 10 digits was 72.79% and 88.79%, respectively.

Lim et al. (2019) [24] proposed a two phase recognition system. The two main phases comprises of hand tracking and hand representation. In the first phase, the hand tracing is performed with the help of particle filter. And in the second phase, a compact hand representation is computed by averaging the segmented hand regions.

Hou et al. (2019) [25] proposed a smartwatch based Sign recognition system termed as SignSpeaker. SignSpeaker was developed by using a smartwatch and a smartphone. The average recognition rate achieved was 99.2% and 99.5%.

III. Proposed Work

A. VGG19 Model

The VGG Network was introduced by Simonyan and Zisserman in [12]. This network only uses 3x3 convolutional layers stacked on the top of each other in increasing depth. On the top of that a max pooling layer is introduced which handles the reducing of volume size. Max-pooling is performed over a 2 × 2 pixel window. After Max-Pooling, the Model consists of three Fully-Connected Layers (FC Layers): the initial two layers both consist of 4,096 Nodes while the third layer is used to perform 1000-way ILSVRC classification and therefore consists of 1000 channels (one for each class). Finally a soft-max layer is introduced.

All the hidden layer in the VGG19 Model are equipped with rectification (ReLU) [26]. The general architecture of a VGG19 Model is shown in Fig. 1.

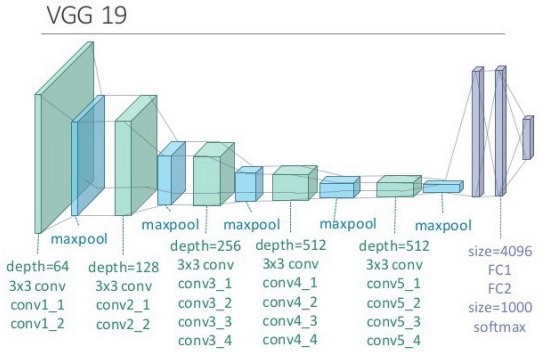


Fig. 1. VGG19 Model Architecture.

B. VGG19 Model Fine Tuning

The original model is pre-trained for the classification of 1000 classes. Since the number of classification classes are inconsistent for the experiment, the authors removed the topmost layer of the model and re-established a new fully connection layer of 24 classes for carrying out the experiment. For Fine-Tuning the model, i.e. for obtaining the appropriate top layer weights, all the other layer of the model are frozen and the model is trained for multiple rounds on the ASL Dataset. After training the model for a significant number of epochs, the model weights were saved. In the second stage, the authors did not trained the complete model due to the presence of relatively less data. Training the complete model will lead to overfitting. Hence, the authors adapt a strategy by freezing the first 16 layers of the model and training the rest of the model. For this stage, the authors used a low learning rate Stochastic Gradient Descent (SGD) and the model is trained.

C. Concatenation

The authors used the above strategy to develop two different VGG19 models namely VGG19-v1 and VGG19-v2. VGG19-v1 was trained by using only the RGB Images and similarly VGG19-v2 was trained only using the Depth Images.

On analysing the past researches on Static Gesture Recognition, there was a need to merge depth-information and colour information together for obtaining high accuracy and recognition rates. Thus, the authors combined the results of both VGG19-v1 and VGG19-v2, as shown in Fig. 2.

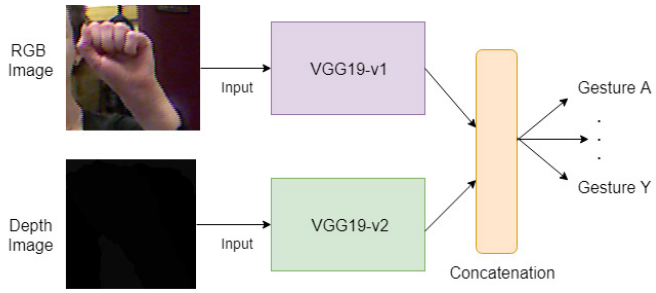


Fig. 2. Proposed Model.

IV. Methodology

A. Dataset

In this research, the ASL dataset published by Pugeault and Bowden in 2011 [16] is used. The dataset provides 24 (except y and z) English letter images in the form of gesture expressions. The ASL dataset is recorded by 5 different persons with the help of Kinect, with non-identical lightning conditions and background conditions. In the ASL dataset, there are approximately ~500 non identical Hand Gesture Images which correspond to each alphabet. Hence the dataset contains approximately 60,000 images for colour and depth. Fig. 3 shows some images from the dataset.

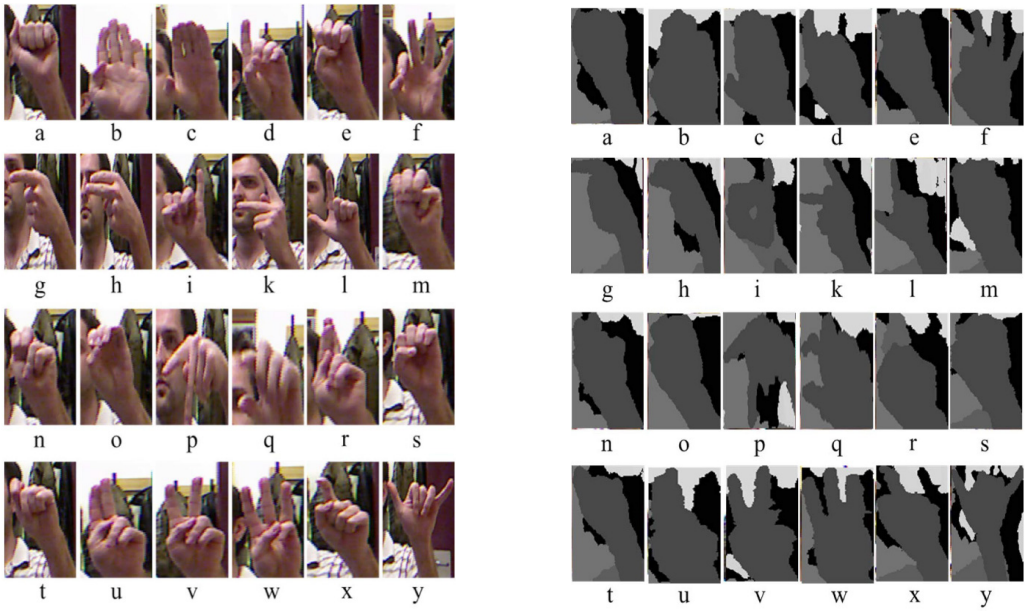
B. Data Augmentation

For Data Augmentation, the authors used a data enhancement tool known as ImageDataGenerator provided by Keras framework. In this tool, the authors set different parameters like roation_range, height_shift_range, Width_shift_range, ZCA_whitening etc. for implementing data augmentation. The transformation helped in increasing the amount of data and also in preventing overfitting.

C. Model Training

The Model proposed in Section III is trained using the ASL Dataset. Both the models i.e. VGG19-v1 and VGG-v2 are trained using the RGB and RGB-D images as proposed. Both the models are trained using the two stage strategy adopted by the authors and the final weights of the model were saved. For avoiding the overfitting of the models, the authors used Early Stopping method in the training process. Early Stopping is the method that monitors model for stopping the training process. With the help of Early Stopping, the authors monitor the accuracy of Validation Set i.e. Validation Accuracy. If the validation accuracy falls to a certain level, the Early Stopping process stops the training of the model after some consecutive epochs.

Now, as the training data consist of both Intensity and Depth images, the authors implemented feature concatenation. Both the Intensity and Depth images are provided as input to the fine-tuned VGG-19 model and then a concatenate function is set just before the topmost soft-max layer and then the soft-max layer is used for classification.



(a) (b)

Fig. 3. ASL Dataset (a) RGB Images (b) Depth Images.

V. Result and Discussion

A. Dataset Processing

1. Data Preprocessing

After carefully studying the dataset, the researchers found out that there are some unusual patterns in the dataset. In some folder there are unequal distribution of colour and depth images. To solve the issue, the researchers carefully removed the depth images that are different and do not correspond to any colour image. Since this research uses the VGG19 Model with an input image dimensions of 299 x 299 x 3, the researchers converted the single channel depth images to 3-channel images in which 1 channel preserves the original depth information and the remaining two channels are set to 0.

2. Augmentation

As the amount of data present in the original dataset is not very large, there was a need to perform Data Augmentation for the Dataset. Data Augmentation is defined as the way of creating new data which will have different orientations. Data Augmentation has two benefits which are as follows:

1. Data Augmentation helps in preventing overfitting.
2. Data Augmentation gives the ability to generate new data from limited data.

3. Final Dataset

After all the phases of cleaning and pre-processing through which the Dataset is passed, the Dataset was converted into two different subsets: Intensity and Depth. Both of these Subsets are divided into 3 categories namely Training Set, Validation Set and Test Set.

Training Set is the Set on which the model will be trained to recognize the ASL. Training Set contains approximately ~46000 Images. Validation Set is defined as the Set which will validate that whether the model trained using the Training Set is accurate or not. Test Set is the unseen Data that is used on the well trained model to predict the accuracy and performance of the model. Both the Validation Set and the Test Set contain approximately ~9000 Images. Fig. 4 shows the distribution of Images among the different sets.

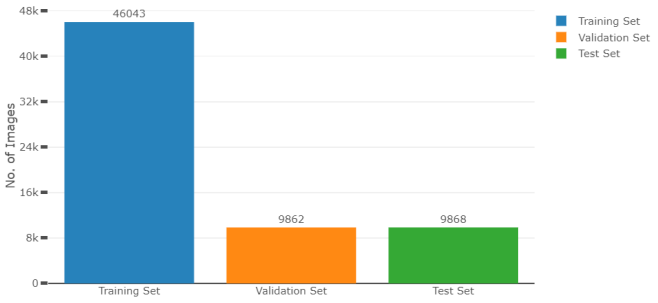


Fig. 4. Train, Test and Validation Split.

B. Tools

In this research, the authors performed comparative experiments. The proposed model is compared with deep learning algorithms and advanced machine learning algorithms. For ensuring the fairness of the experiment, all the models used the same dataset proposed in section III.C. The operating system which was used by the authors for the experiment was Ubuntu 18 and the GPU is Tesla K80. For CNN implementation, Keras framework is used with tensor flow as backend. Keras is an open-source library on neural network which is written in python.

C. Experimental Analysis

The experiments were performed on the ASL dataset. The Model was trained with the help of approximately ~46000 Intensity and ~46000 Depth Images and is tested on approximately ~9000 Intensity and ~9000 Depth Images. The average accuracy obtained by the authors is about 95.29% on the Training Set and about 94.80% on the Test Set. For evaluation of the performance of the proposed model, the authors proposed a Confusion Matrix on the basis of the Test Set Results for the combined Model. Fig. 5 shows the proposed Confusion Matrix.

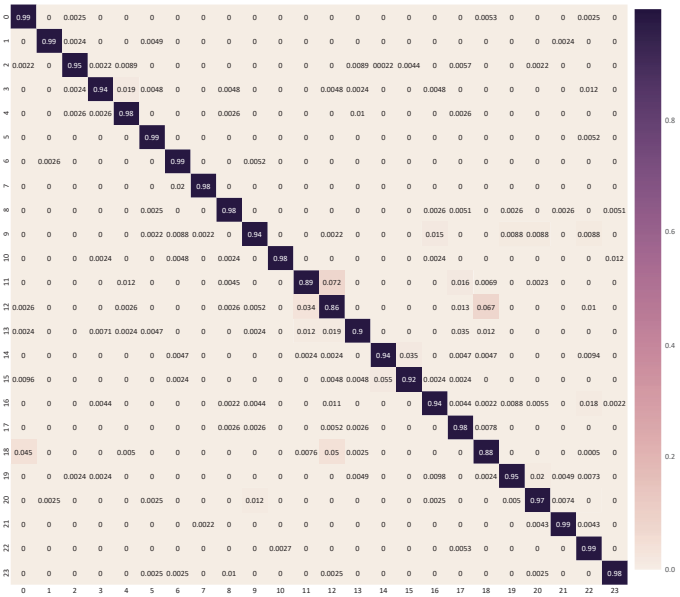


Fig. 5. Confusion Matrix.

D. Comparitive Analysis

The authors compared the results of the proposed model with results of the methods SIFT+PLS, H3DF+SVM and Gabor + RDF that were introduced in Section II. Table I and Fig. 6 represent the Accuracy Comparison. All these algorithms are not implemented by the author, the accuracy of these models is compared with the proposed model's accuracy. Apart from these traditional methods, the authors also directly tested other CNN algorithms such as VGG16, VGG19, CaffeNet and Inception V3 without performing any Fine Tuning. Table II and Fig. 7 represent the Comparison between their Accuracy. It can be seen from the results that the model proposed in this paper has the highest accuracy among all the models.

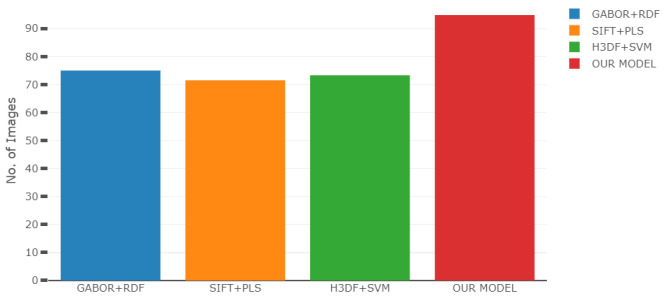


Fig. 6. Traditional Models vs. Proposed Model.

TABLE I. Comparison between Traditional Models and Proposed Model

Recognition Methods	Gabor+RDF	SIFT+PLS	H3DF+SVM	Our Model
Recognition Rate	75%	71.51%	73.3%	94.8%

TABLE II. Comparison between Other CNN Models and Proposed Model

Recognition Methods	CaffeNet	VGG16	VGG19	Inception V3	Our Model
Recognition Rate	73.75%	83.44%	87.37%	88.15%	94.8%

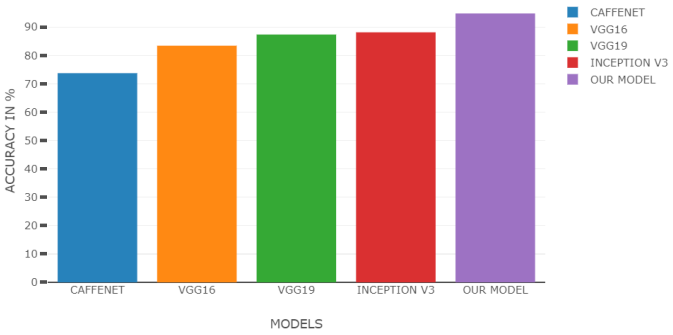


Fig. 7. CNN Models vs Proposed Model.

VI. Conclusion

CNN is currently a powerful artificial intelligence tool that can recognise patterns with high accuracy. In this paper, the authors proposed a fine-tuned VGG19 Model for implementing static gesture recognition. The VGG19 Model was fine-tuned using a two-stage process. In the first stage, all the layers of the model were frozen except the last layer and the model was trained on multiple rounds of ASL Dataset. In the Second stage, only the initial 16 layers of the model were frozen and rest of layers were trained on multiple rounds with low rate SGD due to presence of relatively less data. In comparison with other methods, many of which rely on Features Extraction, the proposed method can easily automate that task for classification. The proposed model is tested on ASL dataset and the recognition rate attained is 94.8%.

In addition, the authors did a comparative study with other models and on comparison it was determined that the proposed model outperforms certain traditional machine learning methods namely Gabor+RDF, SIFT+PLS and H3DF+SVM. Moreover, the model was compared with different CNN models such as VGG16, CaffeNet, VGG19 and Inception V3 without fine-tuning. The maximum recognition rate among these four models was 88.15% with is much lower than the recognition rate 94.8% of the proposed model. For future work, the authors will continue the research in the field of Computer Vision and for optimizing Neural Networks for complex gesture Recognition.

References

[1] W. H. O. (WHO), "Deafness and hearing loss," 2019. Available at: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss (last visited on 13 June 2019).

[2] L. Kin, T. Tian, R. Anuar, Z. Yahya, and A. Yahya, "Sign Language Recognition System using SEMG and Hidden Markov Model," *Conference on Recent Advances in Mathematical Methods, Intelligent Systems and Materials*, 2013, pp. 50–53.

[3] M. P. Lewis, G. F. Simons, and C. D. Fennig, *Ethnologue: Languages of the World*, 17 edn. Dallas: Sil International, 2013.

[4] E. Verdú, C. Pelayo G-Bustelo, M. A. Martínez and R. Gonzalez-Crespo, "A System to Generate SignWriting for Video Tracks Enhancing Accessibility of Deaf People," *International Journal of Interactive Multimedia and Artificial Intelligence,* vol. 4, no. 6, pp. 109-115, 2017. doi: 10.9781/ijimai.2017.09.002

[5] R.E. Mitchell, T.A. Young, B. Bachleda, M.A. Karchmer, "How many people use ASL in the United States? Why estimates need updating," *Sign Language Studies*, vol. 6, no. 3, pp. 306–335, 2006.

[6] A. Kumar, A. Kumar, S. K. Singh and R. Kala, "Human Activity Recognition in Real-Times Environments using Skeleton Joints," *International Journal of Interactive Multimedia and Artificial Intelligence,* vol. 3, no. 7, pp. 61-69, 2016. doi: 10.9781/ijimai.2016.379

[7] M. Raees and S. Ullah, "EVEN-VE: Eyes Visibility Based Egocentric Navigation for Virtual Environments," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 3, pp. 141-151, 2018. doi: 10.9781/ijimai.2018.08.002

[8] I. Rehman, S. Ullah and M. Raees, "Two Hand Gesture Based 3D Navigation in Virtual Environments," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 4, pp. 128-140, 2019. doi: 10.9781/ijimai.2018.07.001D.

[9] Aryanie and Y. Heryadi, "American sign language-based finger-spelling recognition using k-Nearest Neighbors classifier," *2015 3rd International Conference on Information and Communication Technology (ICoICT),* Nusa Dua, 2015, pp. 533-536. doi: 10.1109/ICoICT.2015.7231481

[10] M. M. Islam, S. Siddiqua and J. Afnan, "Real time Hand Gesture Recognition using different algorithms based on American Sign Language," *2017 IEEE International Conference on Imaging, Vision & Pattern Recognition (icIVPR),* Dhaka, 2017, pp. 1-6. doi: 10.1109/ICIVPR.2017.7890854

[11] C. Szegedy, W. Liu, Y. Jia, Y. *et al.*, "Going deeper with convolutions," In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, 2015, pp. 1-9. doi: 10.1109/CVPR.2015.7298594.

[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for largescale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[13] C. Szegedy, V. Vanhoucke, S. Ioffe *et al.*, "Rethinking the inception architecture for computer vision," *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, USA, June 2016, pp. 2818–2826.

[14] V. Gulshan, L. Peng, M. Coram, *et al.,* "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016. doi:10.1001/jama.2016.17216

[15] B. Xie, X. He, and Y. Li, "RGB-D static gesture recognition based on convolutional neural network," *The Journal of Engineering*, vol. 2018, no. 16, pp. 1515-1520, 2018, doi: 10.1049/joe.2018.8327

[16] N. Pugeault and R. Bowden, "Spelling it out: Real-time ASL fingerspelling recognition," *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, Barcelona, 2011, pp. 1114-1119. doi: 10.1109/ICCVW.2011.6130290

[17] B. Estrela, G. Cámara-Chávez, M.F. Campos, W.R. Schwartz and E.R. Nascimento, "Sign language recognition using partial least squares and RGB-D information," In *Proceedings of the IX Workshop de Visao Computacional, WVC*, 2013.

[18] C. Chuan, E. Regina and C. Guardino, "American Sign Language Recognition Using Leap Motion Sensor," *2014 13th International Conference on Machine Learning and Applications*, Detroit, MI, 2014, pp. 541-544. doi: 10.1109/ICMLA.2014.110

[19] L. Rioux-Maldague and P. Giguère, "Sign Language Fingerspelling Classification from Depth and Color Images Using a Deep Belief Network," *2014 Canadian Conference on Computer and Robot Vision*, Montreal, QC, 2014, pp. 92-97. doi: 10.1109/CRV.2014.20

[20] S. Ameen and S. Vadera, "A convolutional neural network to classify American Sign Language fingerspelling from depth and colour images," *Expert Systems*, vol. 34, no. 3, 2017, e12197.

[21] Q. Dai, J. Hou, P. Yang, X. Li, F. Wang, and X. Zhang, "The Sound of Silence: End-to-End Sign Language Recognition Using SmartWatch," in *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking,* 2017, pp. 462-464.

[22] W. Tao, M.C. Leu and Z. Yin, "American Sign Language alphabet recognition using Convolutional Neural Networks with multiview augmentation and inference fusion," *Engineering Applications of Artificial Intelligence*, vol. 76, pp. 202-213, 2018.

[23] T. W. Chong and B.G. Lee, "American sign language recognition using leap motion controller with machine learning approach," *Sensors*, vol. 18, no. 10, 3554, 2018.

[24] K. M. Lim, A. W. C. Tan, C. P. Lee and S. C. Tan, "Isolated sign language recognition using Convolutional Neural Network hand modelling and Hand Energy Image," *Multimedia Tools and Applications*, vol. 78, no. 14, pp. 19917-19944, 2019.

[25] J. Hou, X. Y. Li, P. Zhu, Z. Wang, Y. Wang, J. Qian, J. and P. Yang, "SignSpeaker: A Real-time, High-Precision SmartWatch-based Sign Language Translator," in *Proceedings of the 25th Annual International Conference on Mobile Computing and Networking (MobiCom '19),* Los Cabos, Mexico, 2019, article no. 24.

[26] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems*, vol. 1, Lake Tahoe, Nevada, 2012, pp. 1097-1105.

Manju Khari

Dr. Manju Khari is an Assistant Professor in Ambedkar Institute of Advanced Communication Technology and Research, Under Govt. Of NCT Delhi affiliated with Guru Gobind Singh Indraprastha University, Delhi, India. She is also the Professor- In-charge of the IT Services of the Institute and has experience of more than twelve years in Network Planning & Management. She holds a Ph.D. in Computer Science & Engineering from National Institute Of Technology Patna and She received her master's degree in Information Security from Ambedkar Institute of Advanced Communication Technology and Research, formally this institute is known as Ambedkar Institute Of Technology affiliated with Guru Gobind Singh Indraprastha University, Delhi, India. Her research interests are software testing, software quality, software metrics, information security, optimization, Artificial Intelligence and nature-inspired algorithms. She has 70 published papers in refereed National/International Journals & Conferences (viz. IEEE, ACM, Springer, Inderscience, and Elsevier), Besides this, she associated with many International research organizations as Editor of Springer , Wiley and Elsevier books and Guest editor of International Journal of Advanced Intelligence paradigms, reviewer for International Journal of Forensic Engineering, Inder Science, and editorial board member of International Journal of Software Engineering and Knowledge Engineering, World Scientific.

Aditya Kumar Garg

Aditya Kumar Garg is Associate Software Developer at Kuliza Technologies, Bangalore. He is currently working in backend Software development on latest technologies such as Spring Boot. He received his bachelor's degree from Ambedkar Institute of Advanced Communication Technologies and Research. His research interests include machine learning, Internet of Things, information security.

Rubén González Crespo

Dr. Rubén González Crespo has a PhD in Computer Science Engineering. Currently he is Vice Chancellor of Academic Affairs and Faculty from UNIR and Global Director of Engineering Schools from PROEDUCA Group. He is advisory board member for the Ministry of Education at Colombia and evaluator from the National Agency for Quality Evaluation and Accreditation of Spain (ANECA). He is member from different committees at ISO Organization. Finally he has published more than 200 paper in indexed journals and congresses.

Elena Verdú

Elena Verdú received her master's and Ph.D. degrees in telecommunications engineering from the University of Valladolid, Spain, in 1999 and 2010, respectively. She is currently an Associate Professor at Universidad Internacional de La Rioja (UNIR) and member of the Research Group "Data Driven Science" of UNIR. For more than 15 years, she has worked on research projects at both national and European levels. Her research has focused on e-learning technologies, intelligent tutoring systems, competitive learning systems, accessibility, data mining and expert systems.

A Low Cost and Computationally Efficient Approach for Occlusion Handling in Video Surveillance Systems

Rakesh Chandra Joshi, Adithya Gaurav Singh, Mayank Joshi, Sanjay Mathur *

Deptt. of E&CE, Govind Ballabh Pant University of Agriculture and Technology, Pantnagar (India)

Received 19 September 2018 | Accepted 27 December 2018 | Published 11 January 2019

Abstract

In the development of intelligent video surveillance systems for tracking a vehicle, occlusions are one of the major challenges. It becomes difficult to retain features during occlusion especially in case of complete occlusion. In this paper, a target vehicle tracking algorithm for Smart Video Surveillance (SVS) is proposed to track an unidentified target vehicle even in case of occlusions. This paper proposes a computationally efficient approach for handling occlusions named as Kalman Filter Assisted Occlusion Handling (KFAOH) technique. The algorithm works through two periods namely tracking period when no occlusion is seen and detection period when occlusion occurs, thus depicting its hybrid nature. Kanade-Lucas-Tomasi (KLT) feature tracker governs the operation of algorithm during the tracking period, whereas, a Cascaded Object Detector (COD) of weak classifiers, specially trained on a large database of cars governs the operation during detection period or occlusion with the assistance of Kalman Filter (KF). The algorithm's tracking efficiency has been tested on six different tracking scenarios with increasing complexity in real-time. Performance evaluation under different noise variances and illumination levels shows that the tracking algorithm has good robustness against high noise and low illumination. All tests have been conducted on the MATLAB platform. The validity and practicality of the algorithm are also verified by success plots and precision plots for the test cases.

Keywords

Cascade Object Detector, Kalman Filter, Machine Learning, Occlusion Handling, Video Signal Processing, Video Surveillance.

DOI: 10.9781/ijimai.2019.01.001

I. Introduction

In recent times quest for development of intelligent security systems has become the need of the hour for making residential and office premises safer. Due to increasing threat from different types of activities leading to breach of security, it has become impossible for conventional security systems to detect such activities and alert security system in advance.

Thus, an increasing reliance on surveillance systems has resulted in need for better target detection and tracking techniques. Methods such as Radio Frequency Identification (RFID) tracking are not useful in preventing above mentioned situations; hence there is a need of wide area surveillance. Target tracking via image processing for a video surveillance system provides an attractive solution, which can efficiently track a specific target, record its position throughout the video stream and also analyze its motion pattern.

Video tracking is the process of locating a moving object (or multiple objects) over time using a camera. It has a variety of uses, some of which are; human-computer interaction [1], security and surveillance [2] [3], video communication and compression [4], augmented reality [5], traffic control [6] [7], medical imaging [8] [9], video editing [10] [11], multimedia contexts [12] [13], complex object movements [14], video streaming [15], healthcare systems and smart indoor security systems. Video tracking is a time consuming process due to the amount of data that is captured and needs to be processed.

Further, the algorithm complexity increases if object recognition for tracking is also involved. The objective of video tracking is to maintain detectability of target object in consecutive video frames. To perform video tracking, an algorithm analyses sequential video frames and outputs the movement of targets between the frames. There are a variety of algorithms, each having its strengths and weaknesses. Considering the intended use, it is important to choose the algorithm best suited for the serving the purpose. Traditional tracking algorithms involve foreground extraction of the moving target from a static background and then tracks the coherent blobs of the target. Though, these algorithms are computationally efficient but track all the vehicles which exhibit motion in the stream. Similarly, other tracking algorithms like optical flow techniques as discussed in [16] and wavelet based vehicle tracking as illustrated in [17] also track all the vehicles which exhibit motion or all that are similar in appearance.

The lighting condition varies throughout the whole day (depending upon the weather condition for outdoor and lights for indoor). Low light conditions result into poor discrimination of objects from their background and sometimes lighting condition causes shadows or white-out effect. Most of the background subtraction techniques are sensitive to illumination change and it is difficult to handle the shade and shadow caused by the illumination change. Most algorithms which are able to handle these situations, need time on the order of several frames to estimate and train the background model [18]. Less work

* Corresponding author.

E-mail addresses: rakeshchandraindia@gmail.com (R. C. Joshi), agaurav51@gmail.com (A. G. Singh), mayankjoshi63@gmail.com (M. Joshi), mathur_sanjay@rediffmail.com (S. Mathur).

has been done in direction of tracking a specific vehicle throughout the video stream even in presence of other moving and similar in appearance vehicles, which is essentially the need for tracking an unidentified target.

The algorithm proposed in this paper has been developed keeping in mind the problem of tracking an unidentified vehicle entering a secured premise, such as a college campus. A basic complexity present in such situation is possible presence of other vehicles. The requirement thus becomes to track only the vehicle classified as unidentified and not the other vehicles irrespective of the fact that they are in motion or not. It is also required to track the target vehicle even through any possible occlusion and detect it post occlusion. The above mentioned problem can be divided into three different image processing procedures: Classification of vehicle if it is unidentified or not, target vehicle detection, target vehicle tracking, occlusion handling (if any). Vehicle can be classified as unidentified by a license plate reader as discussed in [19] which has been an established and widely used technology. The proposed algorithm, offers the approach how to detect, track and handle occlusion irrespective of presence of other vehicles in the video-stream. The problem at hand has been solved by adopting Kanade-Lucas-Tomasi (KLT) feature tracker with Cascaded Object Detector (COD) assisted by Kalman Filter (KF) which switches over the control of tracking according to the situations encountered. KLT tracker controls the tracking during no occlusion situations whereas COD takes over the flow when the target vehicle is under occlusion assisted by KF. This approach thus facilitates the realization of a smart vehicle tracking system and handles occlusion also. The algorithm has been tested successfully for tracking a single vehicle in 6 different cases of increasing complexity to test detection, tracking and occlusion handling capability of algorithm. Noise and Illumination variations for all cases have also been considered to test overall robustness of the algorithm.

The remainder of this paper is organized as follows. Section II describes related work in the area of the object tracking. Section III describes different approaches in vehicle tracking. Section IV introduces our proposed system. Experimental results are discussed in Section V. Finally, Section VI offers our conclusions.

II. Related Work

In [20], Borisova *et al*., proposed target tracking based method on object's shape when target has low contrast when compared to background. Maresca *et al*., in [21] proposed the Matrioska tracking framework using Oriented FAST and Rotated BRIEF (ORB) features thus minimizing computational cost over Speeded Up Robust Features (SURF) features. Vehicle tracking using Fractional Feedback Kalman Filter was proposed by Kaur *et al*., in [22] to improve the Kalman gain over traditional Kalman Filter. Feature tracking has been discussed by Ali *et al*., in [23], to achieve multi object tracking of humans. In [24], Baheti *et al*., discussed an automatic object tracking by a combination of Scale Invariant Feature Transform (SIFT) and Random Sample Consensus (RANSAC) which makes tracking invariant to translation and geometric transformations. Xia *et al*., in [25] reliably track a target object in far field using SIFT features and RANSAC.

The real-time tracking with high accuracy in surveillance system is a challenging task. Most of the methods such as SANet [22], HCF [26], MDNet [27], SRDCF [28], MEEM [29], SINT [30], though give high distance precision but due to the computational load, these cannot be utilized in real-time tracking scenarios. In [31], a tracking method is proposed based on the object matching in every frame. It gives highly accurate results but comes with limitations on tracking speed which is due to the computational load in deep feature extraction requirements in each frame. On the contrary, Kernelized Correlation Filter (KCF) [32] tracks fast but it is unable to give much accurate results in short period of time [33]. Similarly, the tracker in [34] requires many frames (normally 6) for initialization and tracks at low speed of 2 fps [35].

Fan and Ling [33] presented a parallel tracking and verifying framework, which consists of two major components, tracker and verifier. Tracker tracks in real time and verifier which runs at specific frame interval instead of each frame, check the tracking results and correct tracker if it is needed. Tracker adjusts the results according to the feedback provided by the verifier. But this method is highly dependent on value of frame interval for verification. If it is small computational load increases otherwise if it is large, the chances to deal the case of occlusion reduces. But our proposed algorithm, take necessary steps at the instant when object goes through occlusion. Kalal *et al.* [35] divided the tracking problem into tracking, learning and detection. Tracker provides the training data from learning component which estimates detection error, is used for updating the detector and the detector initiates again when tracker fails.

Correlation filter requires the information of object as well as the background or negative training data. In [36], to start a tracking process and to gather more information, an image patch is first cropped from the first frame as a sample patch got from initial object position and size information. There patch size is set to 2.5 times of the object size to provide context information about the foreground and background.

Li *et al.* [37] proposed a correlation filter tracker which also considers background hard-negative patches. Fully adaptive cluster method- Affinity Propagation (AP), is applied as a Background patch selection strategy. For AP method, clustering is executed with the help of real-value similarity matrix. In iteration process, cluster numbers and cluster centers are given. It uses one layer of Convolutional Neural Network (CNN). Although the accuracy is high in terms of success and precision rate, the computational load in training process is very high.

A tracking system has been discussed in [38] where a mixture of particle filter, and background modeling to track a group of people is used. Background modeling is helpful when application requires multiple targets to be tracked. An occlusion handling problem has been discussed in [39] using Gaussian background estimation but it does not track a single vehicle under occlusion. Single vehicle tracking under occlusion was also attempted to be solved in [40] where single targeted tracking has been achieved using color segmentation for segmentation of a particular vehicle and locking mechanism. Color segmentation however poses for high possibility of false matches as there can be multiple targets of same color.

Object tracking and data association comprise a set of computer vision techniques which deal with generation of the path and its trajectory in image plane by finding and locating the position of the object in every frame of the video sequence. A considerable research work has been done in both transportation and non-transportation applications. Different vehicle tracking approaches can be classified as follows.

A. Model Based Tracking

Three-dimensional model-based vehicle tracking algorithms have previously been discussed extensively in [41] and [42]. The techniques rely on recovering trajectories and models with high accuracy for a small number of vehicles. A major disadvantage of these approaches is the reliance on mathematically detailed geometric object models. It is not feasible to expect to have detailed models for all vehicles that could be found on the roadway as they can be so diverse in their geometry. A more practical approach will be to use template matching which has been discussed later as it relies on correlation and relative scores.

B. Region Based Tracking

In this approach, tracking system identifies a coherently moving connected region in the image, a `blob', associated with each vehicle

and then tracks it over time [43]. Typically, the process is facilitated by the popular background subtraction technique. Foreground vehicles are detected by subtracting the input image frame from the current background estimate, looking for pixels where difference image is above some threshold and then finding connected components. This technique to track targets is widely used due to its low complexity.

However, this technique is more suited for the applications where all the moving targets are to be tracked, for example in studying flow of traffic. The problem discussed in this paper requires a single vehicle to be tracked even if other vehicles exhibit motion in the frame.

C. Active Contour Based Tracking

This tracking technique is based on active contour models, also known as snakes. This approach is similar to region based tracking. The underlying principle is to retrieve a bounding contour of the object and update it dynamically [44] [45]. According to the contour position in previous and current frame the object is tracked. This technique requires some overlapping of the object region in between previous frame and the current frame. This technique, however, generates significant measurement errors when the target undergoes an occlusion. Also, the problem of tracking the targeted vehicle due to the presence of other vehicles still remains unresolved since active contour method extract fine details about the boundary of the objects with respect to all the background disturbances.

D. Kernel Based Tracking

Kernel by definition means “the central part of something”. When it comes to object tracking, kernel is the central component of the target being tracked. A kernel based tracking algorithm is discussed in [46] in which an isotropic kernel is marked with the target spatially to generate the similarity function and according to that object is localized in an image. Recognition and classification based methods such as template matching, cascade object detection and support vector machine are among the kernel based tracking methods. All of these methods provide the center point of the rigid body which is to be tracked.

These methods provide an attractive option for tracking a single vehicle as they return a single best match; however, this best match is heavily based on the appearance of the target. Since, vehicles like cars have similar appearance, the possibility of false match increases. Also, applying search on every frame over all the pixels increases the computational cost unless search for the target can somehow be made local.

E. Feature Based Tracking

A different approach for tracking an object is feature based tracking which involves finding distinguishable lines, points and corners, considered as features of the target object in the current frame and then matching these features in the next frame. This matching allows the computation of motion model of the vehicle which can be used for tracking the vehicle [47][48]. This method has many advantages. It stabilizes the computational cost of the system. In the presence of partial occlusion some features of the moving vehicle remain visible. Also, feature tracking is possible in low illumination.

From the review of related work, it is found that it is difficult to achieve high precision in real-time tracking applications. Moreover, post occlusion detection accuracy is further affected in such situations. From the forgoing discussion we also concluded that a simultaneous operation of different algorithms is required to track a single targeted vehicle. On comparing all the approaches of tracking approaches feature based tracking is a promising choice as it stabilizes computational cost, does not require foreground extraction and doesn’t require detailed geometric models.

III. Proposed System

The paper presents an approach to track an already classified unidentified vehicle and to handle its occlusion. Our paper proposes utilization of IP camera using a local dedicated network having wireless routers and mobile hotspots to capture the scene unlike in [42] where frame capturing is done via IP camera using an internet connection which makes the system speed and network connectivity dependent.

We propose to detect the target using spatially local template matching technique as discussed later which returns only one detected target whose features are then tracked without any locking mechanism thus decreasing the computational cost of the system. Another tracking technique has been discussed in [49] using KLT feature tracker over SURF features. Though SURF features are reliable, single targeted tracking requires maximum features to be extracted. As discussed in [50], a much higher number of features are extracted using Minimum Eigen Value of interest points used in our algorithm as compared to the number of features extracted using SURF algorithm.

In Fig. 1, the flow chart elaborates the algorithm operation, comprehensively.

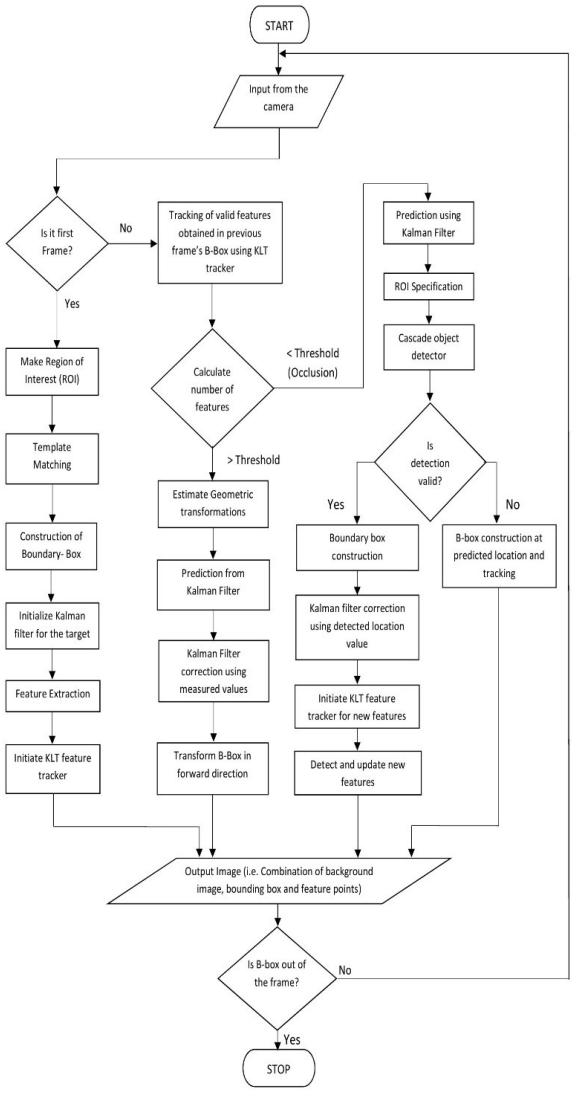


Fig. 1. Flow Chart of Proposed Method.

In the last condition of flow chart, STOP signifies that the bounding box around the targeted vehicle goes out of the frame. Then the system is reset and all the previous values are deleted.

Unidentified vehicle tracking problem is divided in terms of different situations that are required to be handled.

A. Detection In First Frame

In case of vehicle classified as unidentified, camera suitably positioned takes over its tracking operation. In order to be tracked it is essential to identify the location of this vehicle in the first frame correctly. The detection is done using template matching method by dividing the frame into a number of square windows and obtaining some score for each window such as suggested in [51]. However, searching the whole frame for the template can result in false detection due to possible presence of other vehicles. This issue can be resolved by searching the template locally instead of searching it globally. Thus template search is applied only in the region of pixels where geographical location of the entrance is correlated in the frame, as it can be safely assumed that the target car in the first frame will be in vicinity of the entrance. The template matching algorithm used in our study is based on the sum of absolute differences to find the best match given by (1).

$$d(A,B) = \sum_i \sum_j |A(i,j) - B(i,j)| \quad (1)$$

where A = Input frame, and B = Template and i,j = coordinate of pixel.

The tracking bounding box is constructed around the kernel for the window where $d(A,B)_{\min}$ is recorded.

B. Tracking in Subsequent Frames

Once the position in first frame is detected successfully, bounding box is constructed on the detected location to indicate the car in output frame. Feature tracking has been recognized as very reliable method to track targets in computer vision. Feature tracking relies on the concept of finding distinguishable points, corners & lines, called features, of the target. These features are then matched in every subsequent frame and a geometric transformation motion model is estimated which facilitates tracking of the target.

The proposed algorithm also detects features inside the bounding box, i.e. on the car, by finding distinguished points and corners. As bounding box is constructed with the help of template matching approach, so it works better than those methods such as TLD [35], LOT [52], DFT [53], CXT [54] which has reliability on exact boundary around the target in first frame or fixed-size patches. These methods degrade their performance on increasing scaling of the patch at initialization and are more sensitive to the background clutter [55]. Features can be detected by various methods as in [56] [57] [58] [59]. Good features are located by examining the minimum Eigen value of each 2 x 2 gradient matrix of every pixel in region of interest as proposed in [60]. This set of feature points is then tracked by the Pyramidal Implementation of the KLT feature tracking algorithm [61] in each subsequent frame which minimizes the residual function $\boldsymbol{e}$, throughout the down-sampled image pyramids, defined in (2) as follows

$$e(\mathrm{d}) = e(d_x, d_y) = \sum_{x=u_x-w_x}^{u_x+w_x} \sum_{y=u_y-w_y}^{u_y+w_y} (I(x,y) - J(x+d_x, y+d_y))^2 \quad (2)$$

where, $u = [u_x, u_y]^T$ is a point on image I, whose location, $v = u + \mathrm{d} = [u_x + d_x, u_y + d_y]^T$ in subsequent frame J is to be found out.

This algorithm is often used for short-term tracking as a part of larger tracking framework. Number of image pyramids formed by down-sampling previous frame levels is chosen by default to be three. This algorithm also has the capability of marking each feature as either valid or invalid for each frame. Once the feature tracker maps the features from one frame to next, geometric transformation is computed using the similarity projection as given by (3) [62].

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a\cos\theta & -a\sin\theta & t_x \\ a\sin\theta & a\cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (3)$$

Here, $[x \;\; y]$ are the coordinates of the feature tracked in first frame, $[x' \;\; y']$ are the coordinates of the same feature in the next frame, t_x and t_y are the translation in x and y axis respectively, θ is the projection angle and a is the scaling factor. For similarity geometric transformation, at least 2 pairs of matched features are required to compute all 4 degree of freedom. Four degree of freedom meant for translation (t_x, t_y), scaling (a) and rotation (θ). For affine transformation two more degree of freedom added i.e. aspect ratio and shear. On the basis of transformation model thus computed the tracking marker/bounding box is also transformed forward. Only the valid feature points are considered to estimate the geometric transform of the target.

In the first frame the size of the bounding box is directly correlated to the size of template which in turn is dependent on the geographical position where the camera is installed i.e., the size of template in first frame will be decided on case to case basis during the installation of the system, while in subsequent frames by finding the geometric transform between the feature points using the similarity projection (which encompasses scaling transform) the size of bounding box is adjusted, hence even if scale of vehicle size changes during tracking bounding box size is also scaled. As the tracking progresses over time, points can be lost due to occlusions. If a condition arises where number of valid points fall below a threshold, they need to be reacquired to track the object further. This is the condition of occlusion and needs a different technique to be handled.

C. Occlusion Handling

In computer vision, occlusion is the condition when the target being tracked is hidden by another object in the frame. The problem becomes more complicated when the target disappears and reappears after a brief occlusion and has to be tracked again. Partial occlusion (target size > occluding object's size) and complete occlusion (target size < occluding object's size) cases are shown in Fig. 2.

(a) (b)

Fig. 2. Vehicle under (a) partial occlusion (b) complete occlusion.

In feature tracking based algorithm like ours, occlusion is characterized by an extensive loss of features as the target hides. To solve the problem of occlusion it is important to redetect and specify new features if the target re-appears again. Redetection in our algorithm is in the form of a degenerate decision tree, as shown in Fig. 3, of classifiers cascaded together and applied to every sub-window of the frame to classify them as a positive or a negative result as proposed in [63].

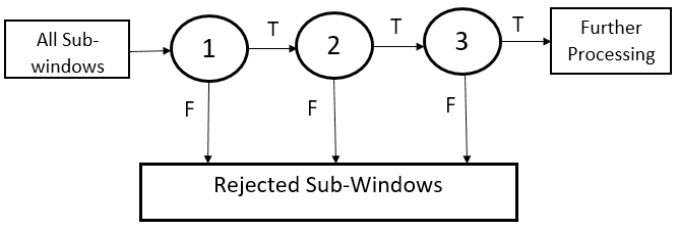


Fig. 3. Schematic representation of the Degenerate Decision Tree based Cascade Object Detector as proposed in [63].

The classifiers are trained with a dataset of cars named GRAZ-02 as developed in [64]. However, just like the first frame detection, the probability of a false match due to the presence of other cars when applied to the frame globally is high. Again, to eliminate the possibility of a false match it is important to search the target after occlusion locally, only at the expected location of target in the next frame, instead of searching it globally in the whole frame.

However, to predict this expected location in next frame it is necessary to compute a motion model of the target according to the history of its motion. The position of the car is a time dependent state vector. Kalman filter [65] is an efficient recursive computational solution to track a time dependent state vector, like position vector of a car, with equations of motion using least-squares method. Kalman filter assists and computes a state transition model to predict the dynamic position of the tracked target defined by following equations.

$$x_k = Ax_{k-1} + w_{k-1} \tag{4}$$

$$z_k = Hx_k + v_k \tag{5}$$

Here, x_k = predicted position of target using previous state; A = State Transition Matrix; x_{k-1} = previous position of target; w_{k-1} = process noise; H = Measurement Matrix; v_k = Measurement Noise; z_k = Measurement.

For each frame x_k is estimated depending on x_{k-1}. The model is then corrected using the measurement z_k. This process is then performed recursively to enhance the model's accuracy.

Condition of occlusion is characterized by the absence of z_k. In this condition, x_k is computed for next frame and the classifier searches for positive result only in the enlarged region of interest in the local neighborhood of x_k which makes our algorithm capable of tracking and detecting the object which is moving with slow or fast speed w.r.t. predicted speed and can also handle the drift or movement in upper and lower coordinated points as shown in Fig. 4. The region of interest (ROI) is greater than the maximum size of any vehicle that may be viewed by the camera and it is just enough in size so that it recognizes the vehicle coming out of the occlusion. ROI constructed in our proposed algorithm is 2.5 times greater than the bounding box.

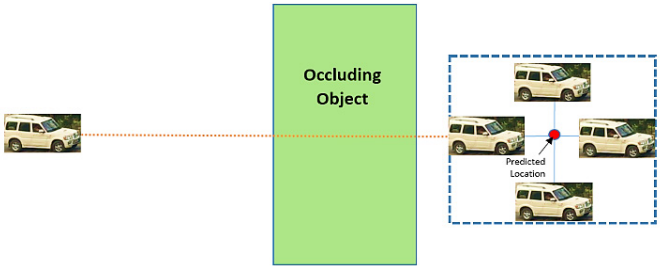


Fig. 4. Enlarged ROI at the time of occlusion.

If a positive result is retrieved new features are detected in the returned bounding box and tracking continues. As positive result will be achieved for the vehicle, most of the features are of vehicle. But in some critical cases, it can be possible that at the end of occlusion there is the inclusion of features of the occluding objects inside the bounding box, but those features are eliminated as *correction* is done in terms of KLT feature updating and, at the end, it will track only the object features and static object or occluding object's features get lost. If a positive result is not retrieved, prediction for next frame is computed and classifier searches again in the new region of interest. The process is recursive until vehicle is detected again after occlusion. If in case of occlusion, COD is not applied, then it can track accurately only those vehicles which are exact at the point of prediction xk. But such case is not always possible and extended ROI concept will also not be applicable. One solution is to make buffer storage of cropped image at certain interval, around the bounding box, and match that patch at the instant when occlusion occurs around the predicted point or on the whole frame. But additionally, computational overload will also be increased.

Vehicle can also be tracked as usual if there are object's features, even in those cases where it exhibits the non-linear motion or when linear motion of vehicle in ground plane is projected non-linear in image plane due to the distance between vehicle and image plane.

In Fig. 5, working of KLT tracker, KF and COD is shown in 2 different cases, i.e. with and without occlusion.

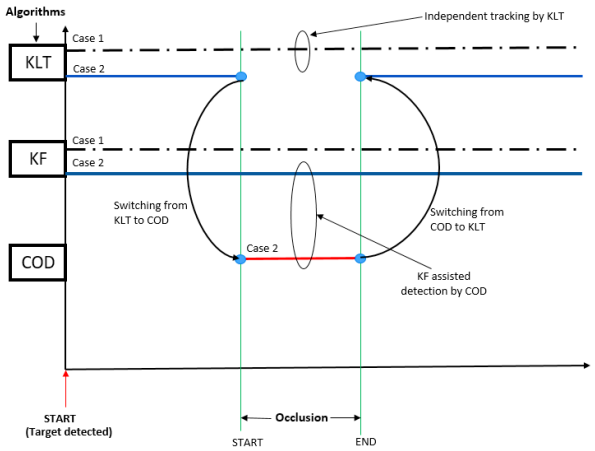


Fig. 5. State Diagram of KLT, KF, COD working together in 2 different cases, i.e. Case 1: without occlusion, Case 2: occlusion (partial and complete).

The next objective is to track the vehicle under different scenarios.

IV. Results

The proposed algorithm has been tested on six different cases of increasing complexity. We have used an Android mobile phone camera (13 megapixels) with IP cam application for capturing the video which is connected to a server cum processing system operating on a 1.5 GHz AMD processor, with 4GB RAM and 64 bit operating system, via Mobile Hotspot. Image processing is done on MATLAB 2016 version. Frame size of the video is 1280×720 and more.

Parametric studies of different Gaussian noise variance and different illumination level have also been performed for the no occlusion, partial occlusion and complete occlusion cases (in all cases where other vehicles are also present). In all the cases variance of noise and illumination level have been varied, the error between the coordinates of tracked vehicle in the case under study and the coordinates obtained for the reference conditions, i.e. zero noise and normal day light illumination (approximately 111,000 lux); have been plotted for each frame. The variance has been kept in increasing trend from 0.01 to 0.125, whereas illumination with respect to the reference condition is varied synthetically by reducing it from 100% illumination to 10% illumination. The closer the recorded tracking error is to zero, the better is the tracking for that frame.

We have taken such cases where there are less number of frames (on an average 50 frames) before occlusion to make cases more complex to check robustness of algorithm. Fig. 6 shows the images with varying Gaussian noise variance 0.01(default value), 0.075 and 0.125 with respect to the original scene.

Although the bounding box size (width) will vary in some cases due to the distance of the camera and the vehicle, in our case it was observed that an average size of 150 pixels is the size of the bounding

box during tracking. In all the cases, the deviation (error) from ideal case was not more than 14% and therefore a tolerance level has been set at (±) 20 pixels. Even in the case when the bounding box size was as large as 300 pixels, the tracking error was found to be less than (±) 20 pixels which shows the high accuracy of our algorithm.

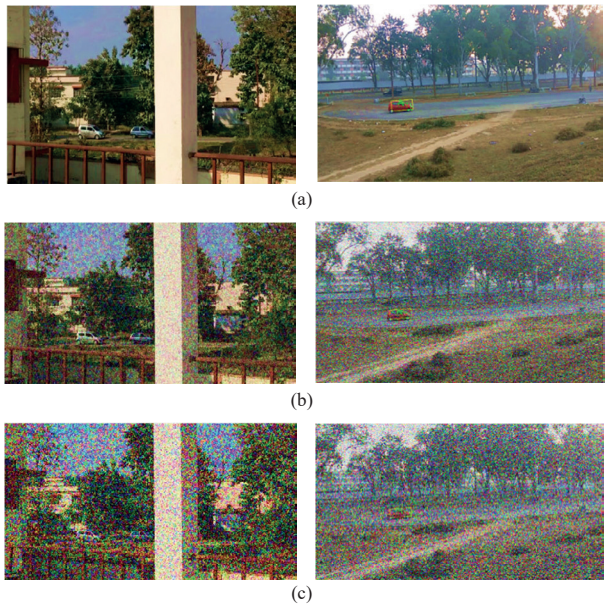


Fig. 6. Images at (a) default 0.01 Gaussian noise variance, (b) with 0.075 Gaussian noise variance, (c) with 0.125 Gaussian noise variance.

Fig. 7, 8, 10, 11, 13, 14 show tracking of the vehicle in six different cases as seen on the monitor with increasing complexity levels.

A. Single Vehicle Without Occlusion

Fig. 7. Tracking of the unidentified target vehicle.

B. Vehicle In Presence of other Vehicles Without Occlusion

Fig. 8. The tracked vehicle in presence of other vehicle.

In the case of Vehicle tracking without occlusion for which error plots are depicted in Fig. 9, system worked perfectly fine when Gaussian noise variance was less than 0.125.With the noise variance crossing 0.125 tracking error increased beyond 20 pixels. However, tracking of vehicle under different illumination levels was found largely illumination invariant under the no occlusion condition as seen in Fig. 9(b).

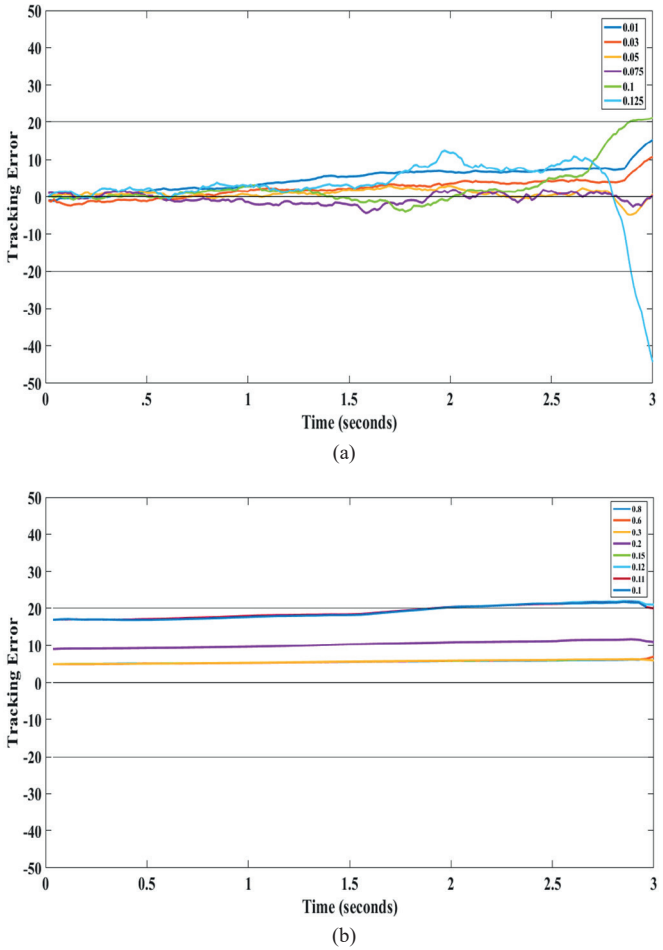


Fig. 9. (a) Tracking Error with different Gaussian Noise Variance (without occlusion) (b) Tracking Error with different Illumination Level (without occlusion).

C. Single Vehicle Undergoing Partial Occlusion

Fig. 10. Tracking of the vehicle before, during and after undergoing partial occlusion.

D. Vehicle in Presence of Another Vehicle Undergoing Partial Occlusion

Fig. 11. Tracking of the vehicle before, during and after undergoing partial occlusion in presence of other vehicle.

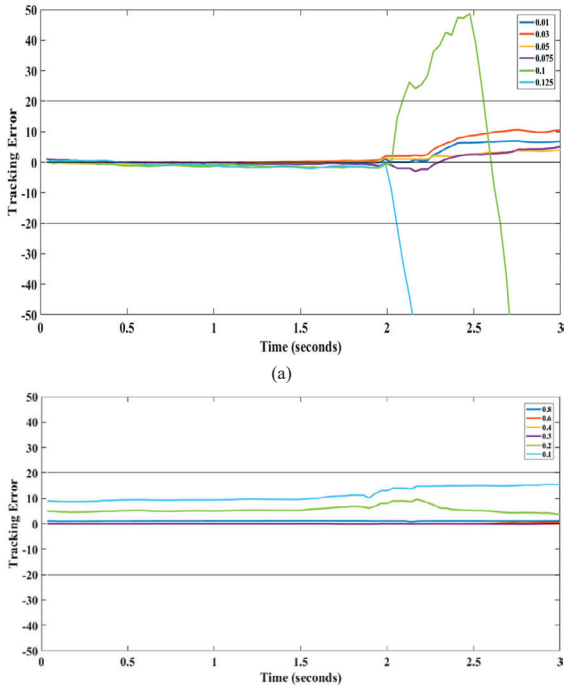


Fig. 12. (a) Tracking Error with different Gaussian Noise Variance (partial occlusion) (b) Tracking Error with different Illumination Level (partial occlusion).

In the case of Vehicle tracking with partial occlusion, system worked well between the ±20 pixels limit until the noise variance was below 0.075 as illustrated in Fig. 12(a), at 0.1 variance; the system tracks the vehicle accurately before the occlusion but the tracking error increases beyond the ±20 pixels limit after occlusion which indicates that algorithm is more susceptible to errors after occlusion due to classification by cascade classifiers. At 0.125 variance, the system fails to track after occlusion, which is an expected result. As in the previous case, system again proved to be illumination invariant, shown in Fig. 12(b) with a slight error after occlusion but still inside the ±20 pixels limit even when illumination is as low as 10% or 0.1 times the reference illumination. The scene is very complex as it consists of a tree as an occluding object instead of a simple pole. In this scene features not reached threshold instantly, as features not wiped out on a vertical line. Hence, command transfer to the COD is delayed by few frames. In this case, our expanded ROI approach is able to handle these type of cases.

E. Vehicle Alone Undergoing Complete Occlusion

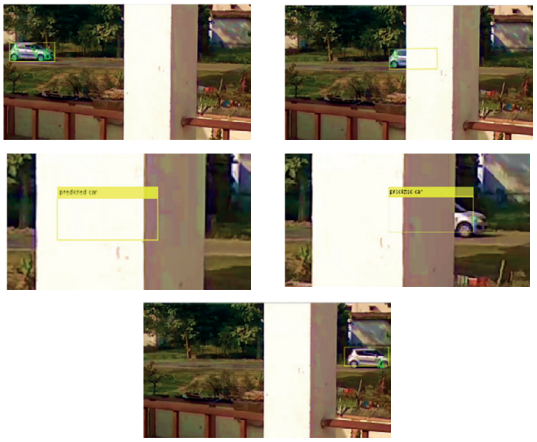


Fig. 13. Tracking of the vehicle before, during and after undergoing complete occlusion.

F. Vehicle in Presence of other Vehicle Undergoing Complete Occlusion

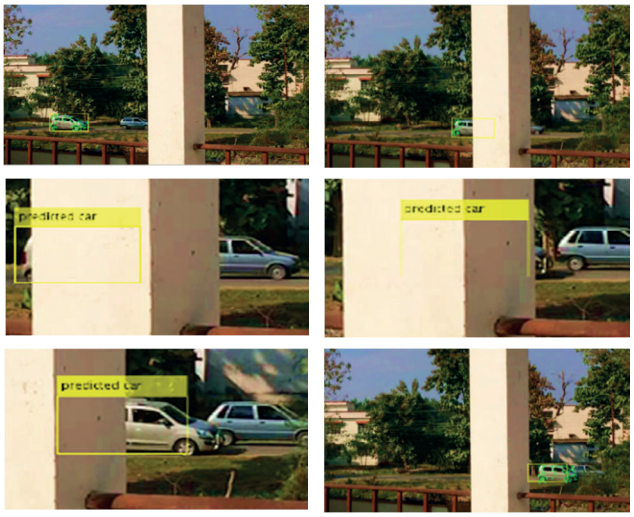
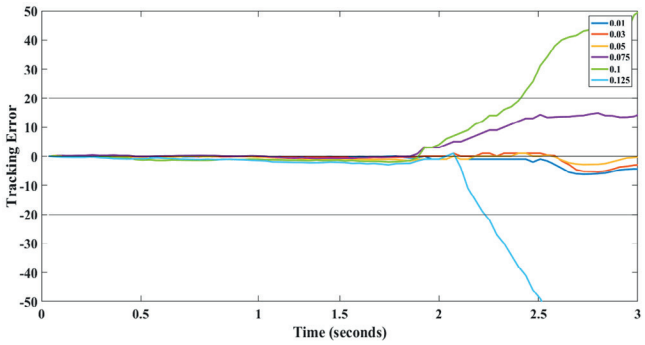


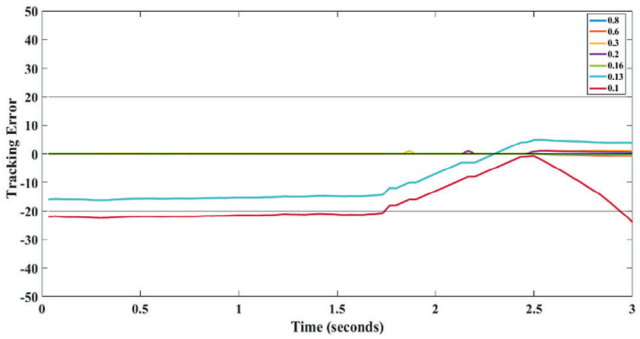
Fig. 14. Tracking of the vehicle before, during and after undergoing complete occlusion in presence of other vehicle.

In the case of Vehicle tracking with complete occlusion, our algorithm is capable to handle the occlusion even after short duration of learning and the system is robust to noise until variance value of 0.075. Although the tracking error increases after occlusion, it is in between the ±20 pixels limit for variance value of 0.075. For variance equal to 0.1, the system loses tracking earlier and shows an erratic behavior after occlusion, i.e. obtained position of the target leads the reference position and then after some frames, obtained position starts lagging the reference position of the target, for 0.125, system fails after occlusion due to very high noise as depicted in Fig. 15(a). On varying the Illumination, the system worked fine with a slight tracking error after occlusion until 17% illumination, at 15% illumination level; there was a tracking error of approximately (-)18 pixels from the first frame, which was reduced by Kalman filter during complete occlusion and was reduced to (+)7 pixels after occlusion, as shown in Fig 15(b). At 10% illumination or 0.1 times reference illumination

level, tracking error was slightly higher than (-)20 pixels and during occlusion; Kalman filter reduced the tracking error to approximately (-)6 pixels. The tracking error recorded can be attributed to excessively low illumination level.



(a)



(b)

Fig 15. (a) Tracking Error with different Gaussian Noise Variance (complete occlusion) (b) Tracking Error with different Illumination Level (complete occlusion).

The tracking is judged basically on two parameters- Success and Precision rate. Precision plot represents the percentage of frames which are within the estimated ground truth threshold varying from 0 to 50. Precision score is calculated where difference of tracked object location and ground truth location is within the distance of 20 pixels. Success plot represents the percentage of frames where overlap score is greater than the threshold. Overlap score, S is calculated as-

$$S = \frac{|B_t \cap B_g|}{|B_t \cup B_g|} \tag{6}$$

where, B_t and B_g are the tracked and ground truth bounding box respectively, $|\cdot|$ represents the number of pixels in a region [55][66]. Success score is calculated for 0.5 overlap threshold. Precision plots and success plots are shown in Fig. 16 and 17 respectively.

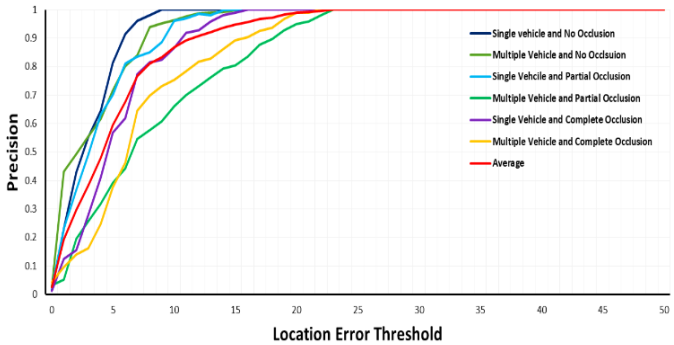


Fig. 16. Precision Plot.

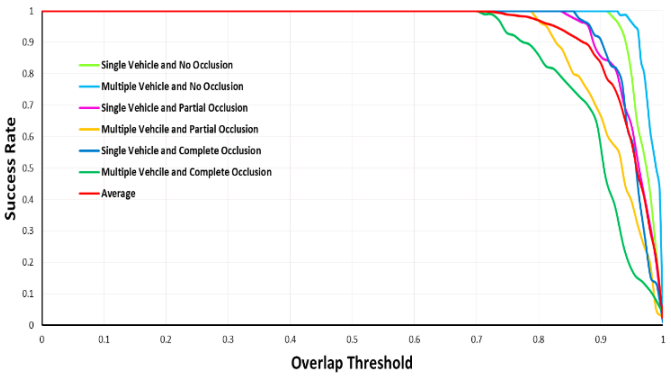


Fig. 17. Success Plot.

The precision score and success score are resulted as 0.9861 and 1.0 respectively. Hence, tracking is done very well with high accuracy.

Table I and Table II show the Maximum Tracking Error; the peak value of tracking error in different cases.

TABLE I. MAXIMUM TRACKING ERROR IN 3 DIFFERENT CASES WITH INCREASING NOISE VARIANCE

Gaussian noise variance	Maximum tracking error (pixels)		
	Without occlusion	Partial occlusion	Complete occlusion
0.01	12	6	(-)6
0.03	8	10	(-)5
0.05	(-)5	4	(-)3
0.075	(-)7	6	16
0.1	22	48	43
0.125	(-)25	Above (-)50	Above (-)50

TABLE II. MAXIMUM TRACKING ERROR IN 3 DIFFERENT CASES WITH DECREASING ILLUMINATION LEVEL

Illumination level (with respect to reference level)	Maximum tracking error (pixels)		
	Without occlusion	Partial occlusion	Complete occlusion
80%	5	3	(-)3
60%	6	2	(-)2
40%	Nil	2	Nil
30%	5	3	(-)4
20%	11	8	3
17%	Nil	Nil	3
15%	21	Nil	(-)4
13%	Nil	Nil	(-)16
12%	22	Nil	Nil
11%	21	Nil	Nil
10%	22	16	(-)23

From an overall analysis, it can be deduced that the proposed algorithm works best in the conditions where Gaussian noise variance is lower than or equal to 0.075 and illumination is greater than 10% or 0.1 times the reference illumination level.

V. Conclusion

To commercially develop a smart surveillance systems that can track an unidentified vehicle and can also detect if it is passing through some occlusion, it is important to design an algorithm which has the capability to track and analyze a single moving vehicle through a video stream even when other vehicles are also present in the stream. This paper proposed an algorithm which solves the above problem by a coherent operation of different tracking and recognition techniques such as Feature Tracking, Template Matching, Cascade Object Detection and Motion estimation using Kalman Filter. The algorithm was tested successfully for 6 cases of different complexity in real time. The performance was analyzed at high noise and low illumination and the algorithm was found to be exhibiting good performance even at high Gaussian noise with a variance of 0.075 and low illumination of 10%. It is able to handle occlusion in a short interval of learning. The precision and success score are 0.9861 and 1.0 respectively which signifies very good tracking.

Our proposed system is completely wireless and works on real time, even more frames can be handled if we reduce the frame size. It is easy to install and cost effective as it is not using other things such as FPGA boards or server based system. System is resistive to motion blur; capable of handling high frame rate (30 fps or more) depending on the connectivity between IP camera and router/hotspot. We have the flexibility to use it with internet or without internet i.e. If we want to use for the local surveillance we can use Wi-Fi network or mobile hotspot without the internet and if we require the surveillance from far place or any part of the world then we can use the network with internet. Surveillance system is secured as the network made through mobile hotspot/ Wi-Fi is password protected (WPA security).

VI. Future Scope

A series of modifications have to be made in above algorithm to make it more efficient to be implemented in a surveillance system to solve other problems. These include solving the problem of inter vehicle occlusion, introducing PTZ (Pan-Tilt-Zoom) camera to make more efficient tracking so that less number of cameras will be required, communication with other cameras when the tracked object goes out of the field of view of one camera to other camera's field of view. Our developed algorithm can also be utilized to track human and other objects during occlusion.

Acknowledgment

This work was supported by the scholarship provided to the first author for postgraduate studies by Ministry of Human Resource Development (MHRD), Government of India.

References

[1] Bulling and H. Gellersen, "Toward Mobile Eye-Based Human-Computer Interaction," in IEEE Pervasive Computing, vol. 9, no. 4, pp. 8-12, October-December 2010.

[2] T. Ko, "A survey on behavior analysis in video surveillance for homeland security applications," 2008 37th IEEE Applied Imagery Pattern Recognition Workshop, Washington DC, 2008, pp. 1-8.

[3] A. Ratre and V. Pankajakshan, "Tucker tensor decomposition-based tracking and Gaussian mixture model for anomaly localisation and detection in surveillance videos," in IET Computer Vision, vol. 12, no. 6, pp. 933-940, 9 2018.

[4] E. Soyak, S. A. Tsaftaris and A. K. Katsaggelos, "Low-Complexity Tracking-Aware H.264 Video Compression for Transportation Surveillance," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 21, no. 10, pp. 1378-1389, Oct. 2011.

[5] H. Kim, J. L. Gabbard, A. M. Anon and T. Misu, "Driver Behavior and Performance with Augmented Reality Pedestrian Collision Warning: An Outdoor User Study," in IEEE Transactions on Visualization and Computer Graphics, vol. 24, no. 4, pp. 1515-1524, April 2018.

[6] D. Beymer, P. McLauchlan, B. Coifman and J. Malik, "A real-time computer vision system for measuring traffic parameters," Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, 1997, pp. 495-501.

[7] Q. Zou, H. Ling, Y. Pang, Y. Huang and M. Tian, "Joint Headlight Pairing and Vehicle Tracking by Weighted Set Packing in Nighttime Traffic Videos," in IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 6, pp. 1950-1961, June 2018.

[8] A. Ghoneim, G. Muhammad, S. U. Amin and B. Gupta, "Medical Image Forgery Detection for Smart Healthcare," in IEEE Communications Magazine, vol. 56, no. 4, pp. 33-37, April 2018.

[9] S. Harput et al., "Two-Stage Motion Correction for Super-Resolution Ultrasound Imaging in Human Lower Limb," in IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 65, no. 5, pp. 803-814, May 2018.

[10] W. Paier, M. Kettern, A. Hilsmann and P. Eisert, "A Hybrid Approach for Facial Performance Analysis and Editing," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 27, no. 4, pp. 784-797, April 2017.

[11] Feng Wang, Chong-Wah Ngo and Ting-Chuen Pong, "Gesture tracking and recognition for lecture video editing," Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004, pp. 934-937 Vol.3.

[12] A. Y. Yang, S. Iyengar, S. Sastry, R. Bajcsy, P. Kuryloski and R. Jafari, "Distributed segmentation and classification of human actions using a wearable motion sensor network," *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Anchorage, AK, 2008, pp. 1-8.

[13] A. Jalal and S. Kamal, "Real-time life logging via a depth silhouette-based human activity recognition system for smart home services," *2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Seoul, 2014, pp. 74-80.

[14] Y. Song, J. Tang, F. Liu and S. Yan, "Body Surface Context: A New Robust Feature for Action Recognition From Depth Videos," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 6, pp. 952-964, June 2014.

[15] A. Jalal, S. Kamal and D. Kim, "Shape and Motion Features Approach for Activity Tracking and Recognition from Kinect Video Camera," *2015 IEEE 29th International Conference on Advanced Information Networking and Applications Workshops*, Gwangiu, 2015, pp. 445-450.

[16] T. Senst, V. Eiselein and T. Sikora, "Robust Local Optical Flow for Feature Tracking," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 9, pp. 1377-1387, Sept. 2012.

[17] J. B. Kim, C. W. Lee, K. M. Lee, T. S. Yun and H. J. Kim, "Waveletbased vehicle tracking for automatic traffic surveillance," Proceedings of IEEE Region 10 International Conference on Electrical and Electronic Technology. TENCON 2001 (Cat. No.01CH37239), 2001, pp. 313-316 vol.1.

[18] W. Kim and C. Jung, "Illumination-Invariant Background Subtraction: Comparative Review, Models, and Prospects," in IEEE Access, vol. 5, pp. 8369-8384, 2017.

[19] Y. Wen, Y. Lu, J. Yan, Z. Zhou, K. M. von Deneen and P. Shi, "An Algorithm for License Plate Recognition Applied to Intelligent Transportation System," in IEEE Transactions on Intelligent Transportation Systems, vol. 12, no. 3, pp. 830-845, Sept. 2011.

[20] V. Borisova and G. D. Yakovlev, "Target tracking based on object's shape," 2016 11th International Forum on Strategic Technology (IFOST), Novosibirsk, 2016, pp. 397-400.

[21] M. E. Maresca and A. Petrosino, "The Matrioska Tracking Algorithm on LTDT2014 Dataset," 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, 2014, pp. 720-725.

[22] H. Kaur and J. S. Sahambi, "Vehicle Tracking in Video Using Fractional Feedback Kalman Filter," in IEEE Transactions on Computational Imaging, vol. 2, no. 4, pp. 550-561, Dec. 2016.

[23] M.M. Naushad Ali, M. Abdullah-Al-Wadud, Seok-Lyong Lee, Multiple object tracking with partial occlusion handling using salient feature points, Information Sciences, Volume 278, 2014, Pages 448-465, ISSN 0020-0255, https://doi.org/10.1016/j.ins.2014.03.064.

[24] Baheti, U. Baid and S. Talbar, "An approach to automatic object tracking system by combination of SIFT and RANSAC with mean shift and KLT," 2016 Conference on Advances in Signal Processing (CASP), Pune, 2016, pp. 254-259.

[25] Y. Xia, G. Li, R. Chen and P. Yu, "A object tracking method in two cameras with common view field," 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Shanghai, 2017, pp. 1-5.

[26] H. Fan and H. Ling, "SANet: Structure-Aware Network for Visual Tracking," *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, HI, 2017, pp. 2217-2224.

[27] C. Ma, J. Huang, X. Yang and M. Yang, "Hierarchical Convolutional Features for Visual Tracking," *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, 2015, pp. 3074-3082.

[28] H. Nam and B. Han, "Learning Multi-domain Convolutional Neural Networks for Visual Tracking," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 2016, pp. 4293-4302.

[29] M. Danelljan, G. Häger, F. S. Khan and M. Felsberg, "Learning Spatially Regularized Correlation Filters for Visual Tracking," *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, 2015, pp. 4310-4318.

[30] Zhang J., Ma S., Sclaroff S. (2014) MEEM: Robust Tracking via Multiple Experts Using Entropy Minimization. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8694. Springer, Cham.

[31] R. Tao, E. Gavves and A. W. M. Smeulders, "Siamese Instance Search for Tracking," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 2016, pp. 1420-1429.

[32] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista. Highspeed tracking with kernelized correlation filters. IEEE TPAMI, 37(3):583–596, 2015.

[33] H. Fan and H. Ling, "Parallel Tracking and Verifying: A Framework for Real-Time and High Accuracy Visual Tracking," *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, 2017, pp. 5487-5495.

[34] Q. Yu, T.B. Dinh, and G. Medioni, "Online Tracking and Reacquisition Using Co-Trained Generative and Discriminative Trackers," Proc. 10th European Conf. Computer Vision, 2008.

[35] Z. Kalal, K. Mikolajczyk and J. Matas, "Tracking-Learning-Detection," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1409-1422, July 2012.

[36] Q. Hu, Y. Guo, Z. Lin, W. An and H. Cheng, "Object Tracking Using Multiple Features and Adaptive Model Updating," in IEEE Transactions on Instrumentation and Measurement, vol. 66, no. 11, pp. 2882-2897, Nov. 2017.

[37] J. Li, X. Zhou, S. Chan and S. Chen, "Robust Object Tracking via Large Margin and Scale-Adaptive Correlation Filter," in IEEE Access, vol. 6, pp. 12642-12655, 2018.

[38] J. S. Kim, D. H. Yeom and Y. H. Joo, "Fast and robust algorithm of tracking multiple moving objects for intelligent video surveillance systems," in IEEE Transactions on Consumer Electronics, vol. 57, no. 3, pp. 1165-1170, August 2011.

[39] S. Zhao, S. Zhang and L. Zhang, "Towards Occlusion Handling: Object Tracking With Background Estimation," in IEEE Transactions on Cybernetics, vol. 48, no. 7, pp. 2086-2100, July 2018.

[40] Prasanna Kumar Reddy Katuru , VivekJosyula , Praveen Samudrala, M. Lokanath, P. Ashish , 2017, "Color Based Vehicle Detection and Tracking using Kalman Filter with a Fractional Feedback Loop", Advances in Computational Sciences and Technology(ACST), Volume 10, Year: 2017,Pages: 2565-2585 , Research India Publications, ISSN: 0973-6107

[41] Worrall A.D., Marslin R.F., Sullivan G.D., Baker K.D. (1991) Model-based Tracking. In: Mowforth P. (eds) BMVC91. Springer, London, https://doi.org/10.1007/978-1-4471-1921-0_39

[42] Koller, D., Daniilidis, K. & Nagel, H.H. Int J Comput 11263 on (1993) 10: 257. https://doi.org/10.1007/BF01539538

[43] KaewTraKulPong P., Bowden R. (2002) An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection. In: Remagnino P., Jones G.A., Paragios N., Regazzoni C.S. (eds) Video-Based Surveillance Systems. Springer, Boston, MA, https://doi.org/10.1007/978-1-4615-0913-4_11

[44] M. Yokoyama and T. Poggio, "A contour-based moving object detection and tracking," 2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005, pp. 271-276.

[45] Kass, M., Witkin, A. & Terzopoulos, D. Int J Computer Vision (1988) 1: 321. https://doi.org/10.1007/BF00133570

[46] Comaniciu, V. Ramesh and P. Meer, "Kernel-based object tracking," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 5, pp. 564-577, May 2003.

[47] E Ali, EU Khan, EZ Mahmudi, R Ullah. "A Comparison of FAST, SURF, Eigen, Harris, and MSER Features." International Journal of Computer Engineering and Information Technology 8, no. 6 (2016): 100-105.

[48] N N. Saunier and T. Sayed, "A feature-based tracking algorithm for vehicles in intersections," The 3rd Canadian Conference on Computer and Robot Vision (CRV'06), 2006, pp. 59-59.

[49] V. Buddubariki, S. G. Tulluri and S. Mukherjee, "Multiple object tracking by improved KLT tracker over SURF features," 2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), Patna, 2015, pp. 1-4.

[50] E Ali, EU Khan, EZ Mahmudi, R Ullah. "A Comparison of FAST, SURF, Eigen, Harris, and MSER Features." International Journal of Computer Engineering and Information Technology 8, no. 6 (2016): 100-105.

[51] H. Schweitzer, R. Deng and R. F. Anderson, "A Dual-Bound Algorithm for Very Fast and Exact Template Matching," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 3, pp. 459-470, March 2011.

[52] S. Oron, A. Bar-Hillel, D. Levi and S. Avidan, "Locally Orderless Tracking," *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, 2012, pp. 1940-1947.

[53] L. Sevilla-Lara and E. Learned-Miller, "Distribution fields for tracking," *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, 2012, pp. 1910-1917.

[54] T. B. Dinh, N. Vo and G. Medioni, "Context tracker: Exploring supporters and distracters in unconstrained environments," *CVPR 2011*, Colorado Springs, CO, USA, 2011, pp. 1177-1184.

[55] Y. Wu, J. Lim and M. Yang, "Object Tracking Benchmark," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1834-1848, 1 Sept. 2015.

[56] S. D. Wei and S. H Lei., "Fast template matching based on normalized cross correlation with adaptive multilevel winner update," IEEE Transactions on Image Processing, vol. 17 no. 11, pp. 2227-2235, 2008.

[57] D. G. Lowe, "Distinctive Image Features from Scale Invariant Keypoints," International Journal of Computer Vision, 2004, vol. 60, no. 2, pp. 91-110.

[58] Calonder M., Lepetit V., Özuysal M., Trzcinski T., Strecha C. and Fua P.,"BRIEF: Computing a Local Binary Descriptor Very Fast", IEEE Trans. Pattern Anal. Mach. Intell., 34(7):1281-1298, 2012.

[59] Bay H., Tuytelaars T. and Gool L.V., "SURF: Speeded Up Robust Features", Comput. Vis. Image Understand., vol. 110, no. 3, pp. 346-359, 2008.

[60] Jianbo Shi and C. Tomasi, "Good features to track," 1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, 1994, pp. 593-600.

[61] Tomasi, Carlo & Kanade, Takeo. (1991). Detection and Tracking of Point Features. International Journal of Computer Vision. 9. 137-154.

[62] Hartley, R., & Zisserman, A. (2004). Multiple View Geometry in Computer Vision. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511811685

[63] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, 2001, pp. I-511-I-518 vol.1.

[64] A. Opelt, A. Pinz, M. Fussenegger and P. Auer, "Generic object recognition with boosting," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 3, pp. 416-431, March 2006.

[65] Greg Welch and Gary Bishop. 1995. An Introduction to the Kalman Filter. Technical Report. University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

[66] Y. Hua, K. Alahari and C. Schmid, "Online Object Tracking with Proposal Selection," *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, 2015, pp. 3092-3100.

Rakesh Chandra Joshi

Rakesh Chandra Joshi was born in Champawat, India in 1994. He received the B.Tech. degree from Invertis University, Bareilly, India in 2015. He has completed the M.Tech. degree from Govind Ballabh Pant University of Agriculture and Technology, Pantnagar, India in 2018 and currently, he is perusing independent research in area of image processing. His research interests include Digital Image Processing, Machine Learning and Surveillance Systems.

Adithya Gaurav Singh

Adithya Gaurav Singh has completed the B.Tech. degree from Govind Ballabh Pant University of Agriculture and Technology, Pantnagar, India in 2018. Currently, he is associated with BaseApp Systems & Software Pvt. Ltd. His research interests include Computer Vision, Robotics and Image processing.

Mayank Joshi

Mayank Joshi has completed the B.Tech. degree from Govind Ballabh Pant University of Agriculture and Technology, Pantnagar, India in 2018 and currently, he is in the process of joining for higher studies. His research interests include Robotics, Computer Vision and Embedded Systems.

Sanjay Mathur

Sanjay Mathur received the B.E. degree from Madan Mohan Malaviya Engineering College, Gorakhpur, India in 1985, the M.E. degree from the University of Roorkee, presently IIT Roorkee, India in 1996, and the Ph.D. from Govind Ballabh Pant University of Agriculture and Technology, Pantnagar, India in 2009. Currently he is a Professor in the Department of Electronics and Communication Engineering, Govind Ballabh Pant University of Agriculture and Technology, Pantnagar. His research interests include Communication Systems, Signal Processing, Image and Audio Processing, Neural Networks and Smart Antenna. He is a Fellow of Institution of Engineers (I) and Chairman of Pantnagar Local Centre. He is also a Life member of Indian Society of Technical Education (ISTE).

Genetic Operators Applied to Symmetric Cryptography

Jefferson Rodríguez, Brayan Corredor, César Suárez *

Systems Engineer, District University Francisco José de Caldas, Bogotá (Colombia)

Received 28 November 2018 | Accepted 14 June 2019 | Published 31 July 2019

Abstract

In this article, a symmetric-key cryptographic algorithm for text is proposed, which applies Genetic Algorithms philosophy, entropy and modular arithmetic. An experimental methodology is used over a deterministic system, which redistributes and modifies the parameters and phases of the genetic algorithm that directly affect its behavior, carrying out a constant evaluation using the fitness function, in order to optimize the results. An independent encryption is established for the auxiliary key, using a main key, in charge of increasing security. The tests are performed over different text sizes, manipulating the parameters and criteria proposed to obtain their appropriate values. Finally, a comparison is presented against the following cryptographic algorithms DES (Data Encryption Standard), RSA (Rivest, Shamir and Adleman) and AES (Advanced Encryption Standard), exposing factors such as processing time, scalability, key size, etc. It is shown that the proposed algorithm has a better performance.

Keywords

Genetic Algorithms, Symmetric Cryptographic, Entropy, Modular Arithmetic, Computer Security.

DOI: 10.9781/ijimai.2019.07.006

I. Introduction

COMPUTER security has always been the discipline responsible for the protection of data stored in a physical or logical computer system. In recent years, technological growth has been such that this security has focused on minimizing any risk to information, combating all types of vulnerabilities that may exist in a network environment. Cryptographic algorithms confront computer attacks that are increasingly complex, forcing them to evolve by participating in recent and reliable methods [1].

There are many algorithms of this type, categorized in Symmetric (private key) and Asymmetric (public key). In symmetric cryptography, a key is used to encrypt and decrypt data, while in asymmetric cryptography two keys are used to perform these tasks; a public figure and a private decipher (e.g., RSA [Rivest, Shamir and Adleman]). Encryption is based on intensively computed mathematical functions and deciphering is usually the reverse process using the key(s) [2], for this DES (Data Encryption Standard) uses a 64-bit key, while AES (Advanced Encryption Standard) uses keys of 128, 192 and 256 bits [3].

In recent years, several investigations have been presented that relate cryptography with genetic concepts, making this field known as an alternative to solve computer security problems. B. Beegom and S. Jose [4] present an asymmetric cryptographic model based on a genetic approach and expose an efficient method in which they make use of the complexity of the DNA chains. N. Srilatha and G. Murali [5] define their work as an efficient three-level cryptographic technique, based on processes that use DNA sequences to transmit information.

Nowadays, for data encryption through the Internet, the HTTPS (Hypertext Transfer Protocol Secure) protocol uses SSL/TLS-based encryption to create a secure channel to shared data [6]. The cryptographic protocols TLS (Transport Layer Security) and SSL (Secure Sockets Layer) used by HTTPS use asymmetric cryptography which uses a pair of keys for sending information, authenticating the receiver more reliably [7].

The presented proposal uses symmetric cryptography and supports its security on the keys more than on the same algorithm, since under an attack it is useless to have knowledge of the algorithm used in the encryption if the key(s) are unknown [13]. Therefore, the use of the phases of the GA is proposed, taking advantage of numerous initial conditions that are included in the key, achieving a quite secure encryption. It is important to add that the security increases based on the length of the key, but at the same time, its access is slowed, which causes more processing time [14].

The proposed algorithm is based on operators used in Genetic Algorithms (GA), adapted to encrypt and decrypt text. The principles of GA were exposed by Holland in 1975 [8], and described more broadly by Goldberg in 1989 [9], Poli, Langdon, McPhee, Michelle and Davis [23].

The main limitation at the time of the creation of a cryptographic algorithm is the idea of innovation, since, currently, there are large amounts of algorithms of this type. With this motivation and in order to generate an original proposal, we decided to take advantage of GA techniques in cryptographic development, making some modifications without neglecting its philosophy. These alterations arise from the interest of creating a different work, which through a clear exemplification demonstrates that it is possible to create developments in the area of computer security using basic concepts. Our work takes the randomness and the order of operation of the genetic phases to arrive at a deterministic development, proposing in addition the use of two keys (auxiliary and main) without ceasing to be a symmetric algorithm.

II. Theoretical Background

A. Genetic Algorithms

The Genetic Algorithms are an adaptive method used to solve search and optimization problems, inspired by biological evolution

* Corresponding author.

E-mail addresses: jefsrodriguezr@correo.udistrital.edu.co (J. Rodríguez), bjcorredorp@correo.udistrital.edu.co (B. Corredor), casuarezp@udistrital.edu.co (C. Suárez)

[10] and based on molecular genetics; a genetic algorithm makes use of terms specific to this field, as well as its main phases: Selection, Crossing and Mutation [11].

Selection: The selection is the stage in which each chromosome (representing a potential solution to the problem) goes through a process of evaluation on a certain fitness value, where some of them are chosen to be later transformed by the crossing and/or mutation operators. In this process the number of chromosomes, genes and alleles of the genotype is kept constant [12].

Crossing: The cross is a genetic operator that allows information to be exchanged between two chromosomes to form a new one. For binary-chain individuals, ring, one-point, two-point, and uniform crossings are often used [12]. Fig. 1 shows the offspring that occurs with a crossing point 4 between two parents of length 5.

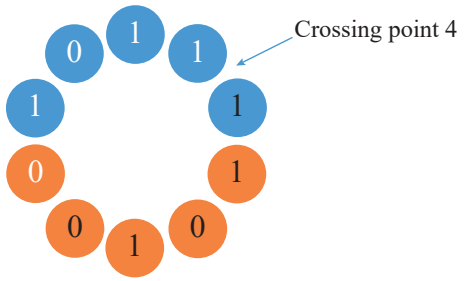


Fig. 1. Example of ring Crossing example.

Mutation: The mutation changes the value of the selected allele [12]. In Fig. 2 the mutation of the selected allele is shown as an example.

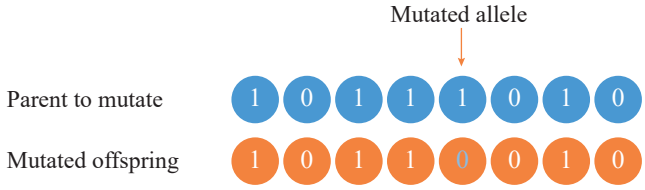


Fig. 2. Illustration of Mutation.

Fitness Function: It is the quality control within the GA and plays a vital role in the guide of the same, it helps to explore the search space more effectively and efficiently [12].

The main idea of the GAs is to reproduce the random nature where the population of individuals adapts to their environment through natural selection, as well as the behavior of the ecosystem. Once the genetic representation of the initial population has been defined, a set of stochastic operators are applied iteratively: selection, crossing and mutation; under certain quality criteria called fitness function. The application of the GAs to optimization problems provides flexibility and adaptability, combined with the robustness and the advantages of the global search [16].

In the philosophy of genetic algorithms, a set of terms of the genetic language have been adopted to clarify and unify the concepts of development of this type of algorithms, as follows: Allele: Each bit is called Allele. Gene: Each group of alleles is called Gen. Chromosome: Each group of genes is called Chromosome.

B. Computer Security

Cryptographic algorithms, have mainly three measurable characteristics: Capacity, Security and Robustness. Capacity is related to the amount of information that the algorithm can process. Security refers to the protection that data receive against possible attacks. Robustness is the resistance that the method has in its entirety against external attacks [15].

There are different types of computer attacks that try to obtain unauthorized access to a network service, among the most common is the Brute Force attack, which makes repeated and systematic attempts using possible credentials, based on different parameters that usually come from sets of credentials set by default, commonly used or valid in previous attacks [17].

C. Related Concepts

The work below has a strong relationship with Entropy, known as the measure of the uncertainty associated with a random variable [18], conceived as a measure of disorder, as well as the repetition of certain combinations. It is measured in bits, where the number of information bits of each character is given by:

$$S = log_2 k \quad (1)$$

where k is the total number of characters. The entropy of a random source is the expected information content of the symbols it has, that is, the expected uncertainty of each symbol, knowing only the distribution according to the symbol [19].

As a last concept, due to its constant use in the different phases of the proposed algorithm, we have the Modular Congruence or Modular Arithmetic, defined as an arithmetic system for whole-number equivalence classes introduced by Carl Friedich Gauss in his book Disquisitiones Arithmeticae (1801) [20]. It is defined as follows:

Let a and b be any integer, and n a positive integer. If n | (a-b) we say that a and b are congruent modulo n and we write [21]:

$$a \equiv b \ (mod\ n) \quad (2)$$

III. Methodology

An investigation-action is established, whose objective is the development of a cryptographic text Algorithm based on principles of computer security and genetic algorithms. The development and the tests are carried out in MATLAB R2017a, with academic license. The chosen methodology is experimental and is summarized in the diagram shown in Fig. 3.

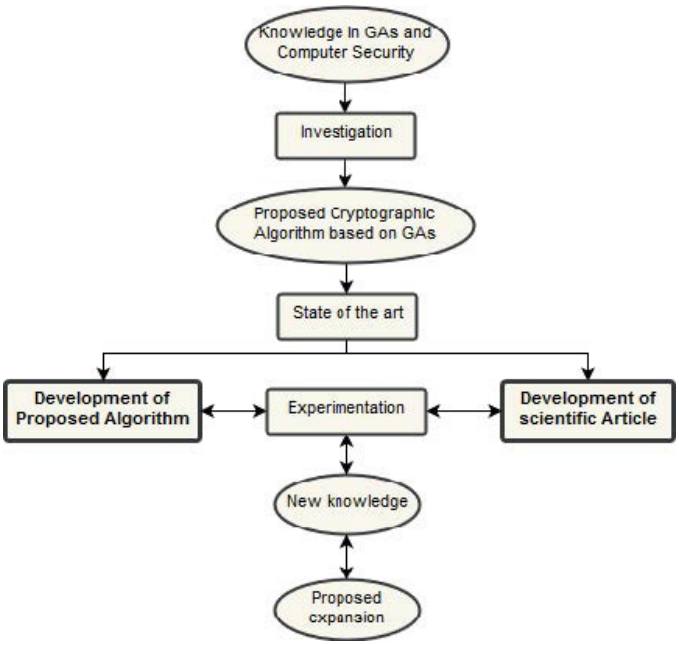


Fig. 3. Applied Methodology.

The methodology is based on previous knowledge in the area of GA and Computer Security, which is accompanied by a deep research and a state of the art description, which aim to establish a clear idea of the development of a symmetric cryptographic Algorithm that takes advantages of the phases of the GA and other related concepts in the matter.

IV. Proposed Cryptographic Algorithm

The proposal below is part of a deterministic system, which implies elimination of randomness and full knowledge of the input variables. It is important to consider that in the nomenclature worked below; the parameters are variables established by the user under certain conditions, while the criteria refer to constant data used to optimize the algorithm and established under a range of tests. The flow diagram (Fig. 4) explains the operation of the proposed algorithm.

A. Initial Population Preparation

It starts from a random message to be encrypted, as a string of characters based on the Latin alphabet and modern English. It is random in nature since it is unknown. As an example, we have the text "Hola mundo" (10 characters counting the space). Table I (annexes) shows the conversion of the entire message to American Standard Code for Information Interchange (ASCII) and then to binary, in this strict order and **taking the interspaces**.

B. Encryption Process

Once the Initial Text has been converted, the Initial Population is taken as the starting point of the GA. The ring-type crossing of the population is carried out with a single point, taking said point from the initial conditions. One child is generated for each pair of parents (genes). Each parent crosses 2 times, with a different partner, which originates a new generation, replacing the previous one. In order to achieve a greater diversity of genes, it is proposed to use the reversalCriterion, which allows to revert the order of the second parent digits of each crossing (or iteration) depending on the value of the criterion. Applying a crossParameter of 6 and a reversalCriterion of 3 (highlighted in bold) the offspring shown in Table I (annexes) is obtained.

The Last Generation is taken to continue with the Selection of the individuals, using the proposed Selection Equation for the Mutation, also called Mutation Clock, taken from the definition of modular congruence (equation 2), and defined as follows:

$$i \equiv 0 \ (mod\ k) \tag{3}$$

Where i is the position of each allele A_i being $1 \leq i \leq n$, where n is the Population Total; and k is he mutationParameter defined between $0 < k < n$ in the initial conditions.

The mutation clock selects the alleles to be mutated, taking the last generation as a string of bits **without counting spaces between genes**.

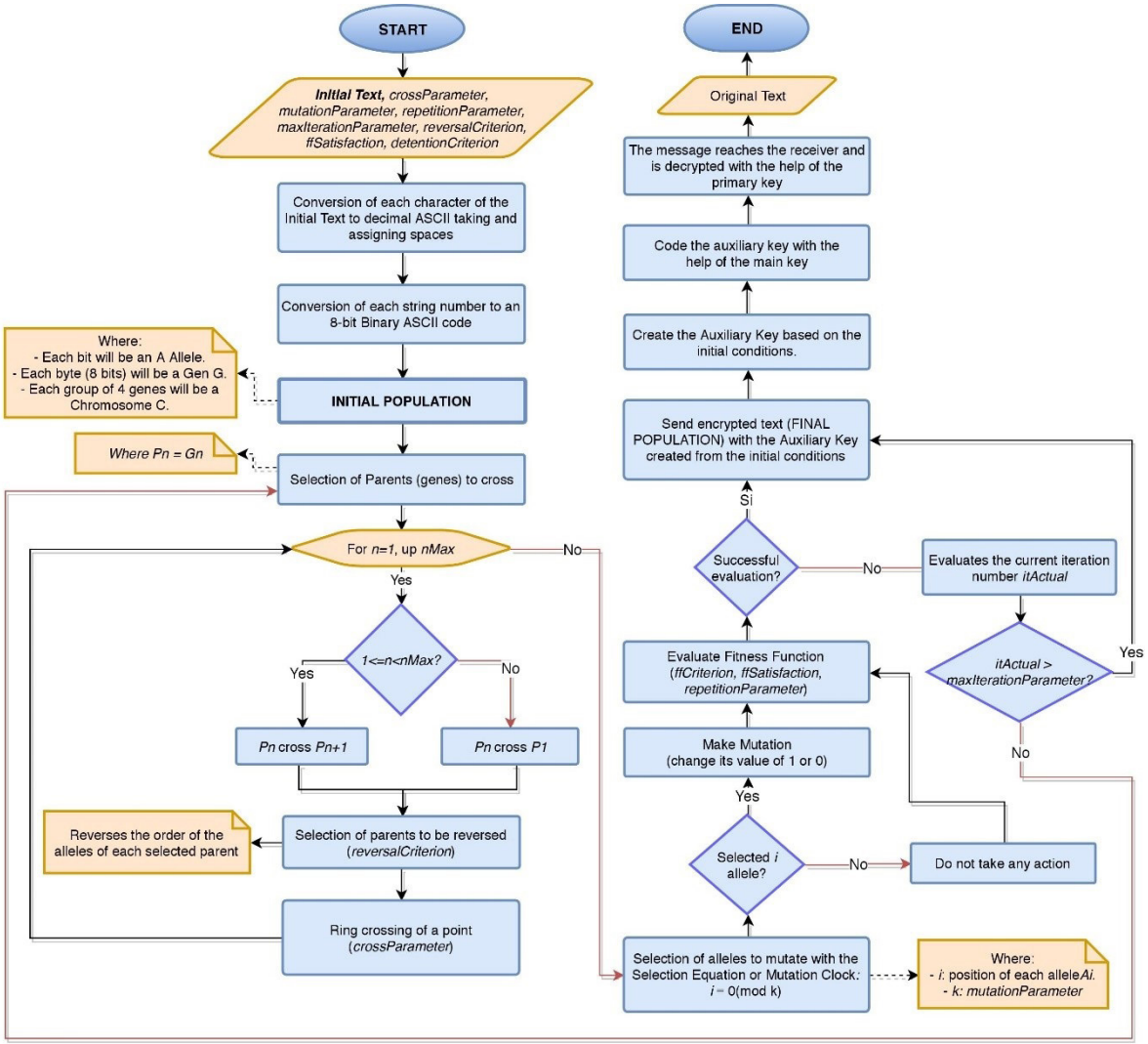


Fig. 4. Flow Diagram of the proposed cryptographic algorithm.

Thus, the algorithm continues with the alleles Mutation, to change their values and generating a new population. For example, the value of the mutation parameter k is taken arbitrarily equal to 5, whose process is evidenced (in bold) in the column Alleles selected by the Mutation Clock, in Table II (annexes).

The last population is taken as the Mutated Population, that at the same time is the **Population to Evaluate**. It is evaluated with the proposed Fitness Function based on Entropy. This function compares groups or allele chains (not necessarily the size of the original genes) to find their frequency within the population. Table III (annexes) shows the value that *ffCriterion* must take based on the length of the population to be evaluated.

In some cases, the number of genes is not an integer number, for these cases, the last chain that is not the size of the *ffCriterion* value is omitted and is not evaluated in the Fitness Function.

For the example case, the range of genes is from 32 to 63, since the total number of genes is 39; therefore, the population is divided in such a way that each gene has 6 alleles, given the value of ffCriterion. It is important to know the total number of alleles in the population:

$$totalAlle = 39 * 8 = 312 \ alleles \tag{4}$$

Next, *totalAlle* is divided into *ffCriterion* to know the number of genes that the Population to Evaluate will have.

$$\frac{totalAlle}{ffCriterion} = \frac{312}{6} = 52 \ genes \tag{5}$$

C. Fitness Function Evaluation

As a result, the population to be evaluated consists of 52 genes, each one of 6 alleles, for a total of 312 alleles. Knowing these data, we proceed to perform the evaluation with the Fitness Function exposed in the pseudocode.

```
function fitness(popToEval, repetitionParameter)
  validChainsCount ← 0
  // C is a different chain's array
  // ia is an array of the different chain's positions
  // ic is an array of the first coincidence's positions
  [C ia ic] ← function FindUniques(popToEval)
  for i ← 0 to length(popToEval)
    count ← length(find(in ic all the positions of i))
    if count <= RepetitionParameter
      countValidChains ← countValidChains + 1
    End
  End
  return [length(C), countValidChains]
End
[DifferentChains, ValidChains] = fitness(popToEval, DetentionCriterion)
SatisfactionFF = (ValidChains/DifferentChains) * 100
DetentionCriterion = 95
If DetentionCriterion <= SatisfactionFF
  Return new generation
If Not
  Continue to the algorithm
```

Where *populationToEvaluate* is the current population, divided into *ffCriterion* size chains; *repetitionParameter* corresponds to a given value in the initial conditions that defines the maximum number of times that each of the chains of the Population to be Evaluated can be repeated; *unique*, *length* and *find* are functions of MATLAB R2017a; *differentChains* are all different chains within the population to be evaluated; *validChains* are strings that meet the value of *repetitionParameter*; *ffSatisfaction* is the percentage of satisfaction to the fitness function given by the *validChains* divided by *differentChains*; *detentionCriterion* is the value of the percentage that is needed to deliver the optimal population (See section Tests and Results, section criteria).

To elucidate the fitness function procedure, we have the following data as input:

$$\boldsymbol{populationToEvaluate} = [0001, 0010, 0001, 0100, 0000, 0010, 0010]$$
$$\boldsymbol{repetitionParameter} = 2$$
$$\boldsymbol{detentionCriterion} = 80\%$$

And as output data:

$$\boldsymbol{differentChains} = [0001, 0010, 0100, 0000]$$
$$\boldsymbol{Quantity\ different\ chains} = 4$$
$$\boldsymbol{validChains} = [0001, 0100, 0000]$$
$$\boldsymbol{Quantity\ valid\ chains} = 3$$
$$\boldsymbol{ffSatisfaction} = \frac{3}{4} * 100\% = 75\%$$

Given the previous case, we obtain a percentage of satisfaction to the Fitness Function of 75%, that represents the total of different chains in the population to be evaluated, 75% is repeated 2 times or less as it is restricted by *repetitionParameter*. The value of *detentionCriterion* established in 80% is greater than *ffSatisfaction*, for this reason, the population does not pass the assessment and the algorithm continues with a new iteration. When the percentage of *ffSatisfaction* is greater than the *detentionCriterion*, the iteration will be considered optimal and the final population is delivered in a large gene chain of 8 alleles, as it was originally stated.

Further, the variable *maxIterationParameter* is defined in the initial conditions, which limits the maximum number of times the algorithm repeats its entire process. If the number of iterations becomes equal to said parameter and an optimal population is not found, the algorithm delivers the population with the best performance.

D. Decryption Process

The decryption of the cypher text is considered as the inverse process of the encryption using the key, in this way the initial text is delivered in a secure way to the receiver. Fig. 5 exposes the diagram of this process.

We propose a function that computes the approximate number of total alleles *at* as a function of the characters *c* of the initial text:

$$at \approx 8 * (4c - 1) \tag{6}$$

V. Keys Generation

The Keys of a cryptographic process are pieces of information whose main objective is to allow the encryption and decryption of data. For the present proposal we have an auxiliary key and a main one.

A. Main Key

It is previously created by the receiver and used by the algorithm (like private key). It consists of 16 characters included in ASCII code (which has 95 printable characters) arranged in a table. The selection of this key depends on the receiver, who creates it under his own criteria,

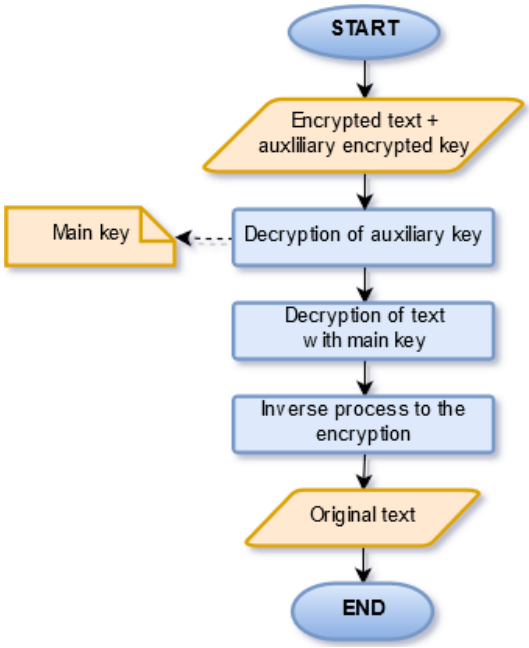


Fig. 5. Flow Diagram of the Decryption process.

always fulfilling that **there cannot be repeated characters**

$$mainKeyExamp = \boldsymbol{Tr}3\boldsymbol{V}QW90/\boldsymbol{am2}PLu* \quad (7)$$

For greater understanding, observe Table IV (annexes), which shows the correct order to locate the key.

B. Auxiliary Key

It is generated in a strict order that only the algorithm knows, it is created as follows:

auxKey=

[crossParameter,mutationParameter,

repetitionParameter,maxIterationParameter,

reversalCriterion,ffSatisfaction,

detentionCriterion,optimalIteration

As an example, we have:

$$auxKeyExamp = [6, 5, 3, 200, 3, 6, 95, 2] \quad (8)$$

Based on the above, we proceed to perform the conversion of each number to ASCII (taken from '0' ASCII code 48, to '9' ASCII code 57) and then to binary. As a result, the Auxiliary Key is obtained together with the information on the length of each of its elements (highlighted in bold).

$auxKeyExamp =$ [00110101, 00110100, 00110101, 00110011,
00110101, 00110001, 00110101, 00110000,
00110100, 00111000, 00110100, 00111000,
00110101, 00110001, 00110101, 00110100,
00110101, 00110111, 00110101, 00110110,
00110101, 00110000, **00110001**, **00110001**,
00110001, **00110011**, **00110001**, **00110001**,
00110010, **00110001**] (9)

C. Auxiliary Key Encryption

The generated Auxiliary key is encrypted with the help of the Main Key, as the first measure the gene is separated into two parts of four alleles each one; the first sub-chain is taken and the first two alleles of it are allocated in the values of the first column of the Table, subsequently the next two alleles are allocated in the first row of the table and the sub-chain is assigned the letter corresponding to the matching coordinate. For the first gene we have:

$$gene1 = \mathbf{0011}0101 \rightarrow gene1_1 = \mathbf{00\ 11}, gene1_2 = 01\ 01$$

The character corresponding to $gene1 = 0011$ is *V*, since it is the point of intersection between row 00 and column 11. In the same way, the process is carried out with the other genes that are part of the main key, obtaining for each gene a pair of characters. The next step is to convert each character to ASCII and then to binary code. Table V (annexes) shows the information corresponding to the first two genes. As a result, we obtain the *encryptedAuxKey* that consists of 180 genes, that is, 1440 alleles.

Each digit in the criteria and parameters represent 2 characters in the auxiliary key, which are taken from the main key and then converted into 8 numbers in ASCII code, or 12 genes in the encrypted auxiliary key, that is, 96 alleles; In addition, the values representing the total number of digits of each condition in the initial auxiliary key always add 384 alleles in the encrypted auxiliary key. The minimum number of possible initial digits is 9 (*crossParameter* 1 digit, *mutationParameter* 1 digit, *repetitionParameter* 1 digit, *maxIterationParameter* 1 digit, *reversalCriterion* 1 digit, *ffSatisfaction* 1 digit, *detentionCriterion* 2 digits, *optimalIteration* 1 digit), that is, 864 alleles in the encrypted auxiliary key, which means that the key will be at least 864 + 384 = 1248 alleles. Given this value, the probability of being guessed is:

$$\frac{1}{2^{1248}} \approx 2{,}06331 * 10^{-376}\% \quad (10)$$

The purpose of encrypting the auxiliary key is to increase security, because if it becomes a victim of a *Brute Force* computer attack, the encrypted string but not the original key could be discovered. This key goes from being a chain composed of 30 genes to have 180 after the encryption process, which implies a significant advance for the algorithm in terms of the security, since its length is considerably increased.

D. About the Keys

The amount of information in bits that each key carries, is based on the possible combinations that the same one can have (value that differs from its representation in bits). Each parameter and criterion takes a value or another depending on its own restrictions (if it has any). The parameter value of crossParameter ranges from 1 to 7, *mutationParameter* between 1 and 999, *repetitionParameter* between 1 and 99, *maxIterationParameter* between 1 and 999, *reversalCriterion* between 1 and 999, *ffSatisfaction* between 1 and 99, *detentionCriterion* between 72 and 98, optimalIteration between 1 and 999. According to the above, the length of the auxiliary key is determined as follows -based on equation (1), where *k* would be the number of total combinations-:

$$log_2\,(7 * 999 * 99 * 999 * 999 * 99 * 27 * 999) \approx 60.67\ bits \quad (11)$$

Furthermore, given that the possibilities of each of the 16 characters in the **main key** is 95 (characters printable in ASCII code) its information in bits is determined as follows:

$$log_2\,(95^{16}) \approx 105.11\ bits \quad (12)$$

VI. Quality Analysis

For the quality analysis we have the data generated of the algorithm developed, the tests and discussion about it, a comparative table of speed between DES and RSA algorithms, and finally a comparison of different factors between the AES and DES algorithms.

A. Test and Results

The tests below are carried out based on the development of the

proposed algorithm, it allows to conclude on the results thrown by the same. Variations are made in some criteria (treated under the tests, however, are established under the user's criterion) to optimize the encryption process aimed at security and performance, the treatment of restricted parameters within their ranges is also performed.

crossParameter: This parameter is used by the algorithm in the crossing stage, it defines the start point to select the offspring between two genes, the maximum value that can take must not exceed the length of a gene. Since there is always an initial population composed of genes of 8 alleles (length 8), this variable must be contained within a range between 1 and 7.

mutationParameter: The mutation parameter defines the alleles to be selected in order to modify them, allowing the algorithm to diversify the population. It is suggested to set this variable to an odd value, because the genes are of length 8 (even number). The number of mutated alleles varies according to the value of the parameter, if this is greater than the size of the population, the algorithm does not make any mutation.

repetitionParameter: This variable defines the maximum number of times that each of the chains of the population to be evaluated can be repeated. Its value is established according to the performance of the algorithm, **note that if the text size is long, more repeataded chains are found**.

reversalCriterion: It allows to invert the digits of the second father of each crossing (or iteration), its value is selected based on the results thrown by the tests itself. For these, a value is taken between 1 to 6 and its performance is evaluated in 3 different cases according to the Fitness Function. It is iterated 4 times and the average satisfaction percentage is found. Table VI (annexes) shows the values that each initial condition takes for three different cases. Table VII (annexes) shows the results obtained by the tests carried out on the algorithm for each case.

It is observed that for Case 1 the best result (97%) is obtained in the values of *reversalCriterion* of 1, 2 and 4, implying a great homogeneity and allowing to conclude that if the initial text is short (until 4 words or 25 characters), the value of the criterion can vary without affecting the efficiency of the algorithm. For Cases 2 (90%) and 3 (97%) the best performance is achieved in the value of 3, it is suggested that this should be taken as *reversalCriterion* when the string is greater than 4 words.

detentionCriterion: The detention criterion establishes the percentage of different chains that comply with the *repetitionParameter*, it is used in the fitness function to evaluate if an iteration passes the evaluation. Based on the tests carried out for the *reversalCriterion*, it is suggested to establish the value of this variable in a range between 72 and 98, since this was the minimum and the maximum percentage obtained, respectively.

B. Runtime and Call Functions

Some compilation data of the algorithm in MATLAB are presented in Table VIII (annexes), where the number of calls, the time spent by each of the present functions and the total value are described. For these tests, the cases described in Table VI are taken.

Runtime depends directly on the iterations made by the algorithm to arrive at the optimal solution; for the tests, it was established a *detentionCriterion* of 98, it carried out 2, 52 and 10 iterations for Case 1, 2 and 3 respectively. It is also observed that the function with the highest number of calls is *Convert* for all cases, this function aims the conversion of ASCII code to binary code and is used to transform the initial text into chains that the algorithm manipulates to perform the encryption, besides it is used in the process of creating keys.

It is also noted that the function that covers the longest runtime (a little more than the half) is A*lgorithmEncryption*, since it complies the task of carrying out the encryption process, starting from the initial population previously converted and getting a final population, ready for the evaluation.

VII. Comparison

In order to establish objective conclusions about the performance of the proposed cryptographic algorithm, a comparison is made (at encryption execution time) against DES and RSA [22], allowing a comparison to be made as shown in Table IX (annexes). The value of *Total Time Algorithm Proposed* takes into account the time of the evaluation of the Fitness Function, additionaly to the creation and encryption of the keys, but not the time of decryption.

The text size for the proposed algorithm is taken as the bit representation of the original text, that is, the so-called *Initial Population*. The values of *detentionCriterion* and *reversalCriterion* were established in 97 and 3 respectively for the tests, while the values of *crossParameter*, *mutationParameter*, *repetitionParameter* and *maxIterationParameter*, respectively: 6, 5, 3 and 200 were set for text sizes of 128 and 256 bits; 5, 7, 5 and 100 for the text sizes of 512, 1000, 2000 and 5000 bits and 5, 17, 10 and 100 for the text size of 10000.

The key length (understood as the amount of information it carries, but not the size of its representation in bits) is 56 bits for DES and 22 bits for RSA. For the proposed Algorithm it is 105.11 bits for the main key and 60.67 bits for the auxiliary key.

It is remarkable the performance against the DES and RSA algorithms, the superiority in runtime against RSA is shown in Table IX. On the other hand, the Proposed Algorithm is superior to DES in execution time, for texts greater than 512 bits. In addition, it is observed that each of the total times of the proposed algorithm, mostly exceed that of the RSA encryption (except for 128 bits) and are very close to the DES encryption time.

In Table X (annexes), different features of the AES, RSA and DES algorithms [2] are exposed, to compare against the proposed Algorithm.

VIII. Conclusion

The proposed algorithm modifies the order and process of the phases of the genetic algorithms, by applying a deterministic system, leaving aside some random procedures.

When comparing the proposed algorithm against RSA and DES, satisfactory performance is evidenced in several factors, demonstrating that Genetic Algorithms are a good alternative to face problems in computer security.

The proposed algorithm manages to disrupt the information through entropy, evidenced in the fitness function as the different chains.

The length (amount of information that it transports) of the auxiliary key is of 60.67 bits and of the main key is of 105.11 bits, overcoming in this aspect the cryptographic algorithm DES and approaching considerably to AES.

The present work exposes a development based on basic concepts like GA, entropy, modular congruence and determinism, that together make an efficient cryptographic process.

IX. Considerations

The code of the algorithm developed is in a private *github* repository, with the option of being visible for those who request access to any of the contact emails. In addition there is a demo in *heroku* that performs the whole process of encryption of a given text: https://symmetric-cryptography-genetic.herokuapp.com/

It is expected to be able to use the principles of the proposed

algorithm in the encryption of images and audio. On the other hand, there is a possible application in data compression.

Annexes

TABLE I. Conversion to ASCII and Binary Code. Crossing of the Initial Population (First Iteration)

Initial Message	Conversion to ASCII	Conversion to Binary (Initial Population)	Position of each parent or Gene	Crossed Parents	Offspring (First Generation)
H	0	**00110000**	1	1 y 2	00001101
	7	00110111	2	2 y 3	11001100
	2	00110010	3	3 y **4**	10000001
	(space)	**00100000**	4	4 y 5	00001100
o	1	00110001	5	5 y 6	01001100
	1	00110001	6	6 y **7**	01100011
	1	**00110001**	7	7 y 8	01001000
	(space)	00100000	8	8 y 9	00001100
l	1	00110001	9	9 y **10**	01000011
	0	**00110000**	10	10 y 11	00001110
	8	00111000	11	11 y 12	00001000
	(space)	00100000	12	12 y **13**	00000011
a	0	**00110000**	13	13 y 14	00001110
	9	00111001	14	14 y 15	01001101
	7	00110111	15	15 y **16**	11000001
	(space)	**00100000**	16	16 y 17	00001100
(space)	0	00110000	17	17 y 18	00001100
	3	00110011	18	18 y **19**	11010011
	2	**00110010**	19	19 y 20	10001000
	(space)	00100000	20	20 y 21	00001100
m	1	00110001	21	21 y **22**	01000011
	0	**00110000**	22	22 y 23	00001110
	9	00111001	23	23 y 24	01001000
	(space)	00100000	24	24 y **25**	00100011
u	1	**00110001**	25	25 y 26	01001100
	1	00110001	26	26 y 27	01001101
	7	00110111	27	27 y **28**	11000001
	(space)	**00100000**	28	28 y 29	00001100
n	1	00110001	29	29 y 30	01001100
	1	00110001	30	30 y **31**	01000011
	0	**00110000**	31	31 y 32	00001000
	(space)	00100000	32	32 y 33	00001100
d	1	00110001	33	33 y **34**	01000011
	0	**00110000**	34	34 y 35	00001100
	0	00110000	35	35 y 36	00001000
	(space)	00100000	36	36 y **37**	00100011
o	1	**00110001**	37	37 y 38	01001100
	1	00110001	38	38 y 39	01001100
	1	00110001	39	39 y **1**	01000011

Source: Authors.

TABLE II. Population's Selection and Mutation

Selected alleles by the mutation clock	Mutated Population
0000**1**101 1**1**0011**0**0 100**0**0001 **0**0001**1**00 01**0**0110**0** 0110**0**011 0**1**001**0**0 000**0**1100 **0**1000**0**11 00**0**0111**0** 0000**1**000 0**0**00001**1** 000**0**1110 **0**1001**1**01 11**0**0000**1** 0000**1**100 0**0**0011**0**0 110**1**0011 **1**0001**0**00 00**0**01100 0100**0**011 0**0**0011**1**0 0100**1**000 **0**0100**1**11 01**0**0110**0** 0100**1**101 1**1**00000**1** 000**0**1100 **0**1001**1**00 01**0**0001**1** 00001000 0**0**0011**0**0 010**0**0011 **0**0001**1**00 00**0**0100**0** 0010**0**011 0**1**0011**0**0 010**0**1100 **0**1000**0**11	00000101 1**1**001**1**10 100**1**0001 **1**0001**0**00 01**1**01**1**01 0110**1**011 00**0**01**1**0 0001**1**100 **1**1000**1**11 00**1**01**1**1 00000000 0**1**000**0**01 0001**1**110 **1**1001**0**01 11**1**00000 00000100 0**1**001**1**10 11**0**00011 **0**0001**1**00 00**1**011**0**1 0100**1**011 0**1**001**1**00 0101**1**000 **1**0100**1**11 01**1**01101 0100**0**101 1**0**00001**1** 000**1**1100 **1**1001**0**00 01**1**0001**0** 00000000 0**1**001**1**10 010**1**0011 **1**0001**0**00 00**1**01001 00101011 00**0**01**1**10 0101**1**100 **1**1000**1**11

Source: Authors.

TABLE III. *ffCriterion* Value According to the Range Where the Population Size Is Located

Genes range of numbers in the population	Value *n*	Value according to the range	Alleles chains length to compare (ffCriterion)
4 - 7	2	$2^2 = 4$	3
8 - 15	3	$2^3 = 8$	4
16 - 31	4	$2^4 = 16$	5
32 - 63	5	$2^5 = 32$	6
2^n - $2^{n+1}-1$	n	2^n	$n+1$

Source: Authors.

TABLE IV. Main Key Distribution

	00	**01**	**10**	**11**
00	**T**	**r**	**3**	**V**
01	Q	W	9	0
10	/	**a**	**m**	**2**
11	P	L	u	*

Source: Authors.

TABLE V. Corresponding Characters for the First Two Genes of Auxiliary Key

Gene	$Gene_n$	Substring value	Corresponding character (according to table 4)	Conversion to ASCII	Conversion to binary (*encryptedAuxKey*)
gene1	$gene1_1$	0011	V	0	00110000
				8	00111000
				6	00110110
	$gene1_2$	0101	W	0	00110000
				8	00111000
				7	00110111
gene2	$gene1_3$	0011	V	0	00110000
				8	00111000
				6	00110110
	$gene1_4$	0100	Q	0	00110000
				8	00111000
				1	00110001

Source: Authors.

TABLE VI. Initial Conditions Value for the Three Tests Cases

Initial condition	Case 1	Case 2	Case 3
Initial text	Hola mundo	Esta es una cadena mas larga que la anterior	See **Text 3**
Initial population (bits representation)	39	175	2051
Population to Evaluate	52	175	1367
crossParameter	6	5	5
mutationParameter	5	5	17
repetitionParameter	3	4	7
maxIterationParameter	200	200	200

Source: Authors.

TABLE VII. TESTS ON *REVERSALCRITERION* FOR CASE 1, CASE 2 AND CASE 3

reversalCriterion	**Different strings**			**Valid strings**			**Satisfaction percentage**			**Average Satisfaction percentage**		
Case	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
1	36	68	552	36	59	516	100%	87%	93%	**98%**	83%	94%
	33	57	569	32	45	536	97%	79%	94%			
	32	56	547	32	44	511	100%	79%	93%			
	34	65	546	32	58	513	94%	89%	94%			
2	36	72	606	36	63	578	100%	88%	95%	**98%**	81%	95%
	35	63	576	34	48	550	97%	76%	95%			
	33	67	607	32	57	575	97%	85%	95%			
	35	58	583	34	43	555	97%	74%	95%			
3	32	78	659	28	68	635	88%	87%	96%	93%	**90%**	**97%**
	33	70	704	33	59	684	100%	84%	97%			
	36	84	645	33	77	625	92%	92%	97%			
	34	85	628	32	81	602	94%	95%	96%			
4	37	54	528	36	52	497	97%	96%	94%	**98%**	82%	95%
	35	49	590	34	33	567	97%	67%	96%			
	33	62	579	33	50	548	100%	81%	95%			
	27	65	596	26	55	564	96%	85%	95%			
5	29	50	606	26	39	585	90%	78%	97%	91%	72%	96%
	33	42	633	31	27	606	94%	64%	96%			
	30	47	641	28	34	619	93%	72%	97%			
	28	47	636	24	35	616	86%	74%	97%			
6	34	64	622	32	53	594	94%	83%	95%	96%	83%	95%
	31	63	631	31	50	601	100%	79%	95%			
	31	68	641	30	55	611	97%	81%	95%			
	31	72	624	29	65	596	94%	90%	96%			

Source: Authors.

TABLE VIII. ALGORITHM PERFORMANCE FOR CASE 1, CASE 2 AND CASE 3

Function's name	**Function calls**			**Time (seconds)**		
	Case 1	**Case 2**	**Case 3**	**Case 1**	**Case 2**	**Case 3**
AlgorithmEncryption	1	1	1	0,027	0,133	0,261
AuxiliaryKeyEncrypt	60	64	72	0,011	0,012	0,012
FitnessFunction	2	52	10	0,002	0,011	0,028
Unique	2	52	10	0,008	0,023	0,018
Convert	249	399	2299	0,004	0,004	0,008
Strlength	12	13	15	0,001	0,001	0,001
NewDivisionPopulationToEvaluate	2	52	10	0,001	0,029	0,045
TOTAL	**328**	**633**	**2417**	**0,055**	**0,215**	**0,374**

Source: Authors, MATLAB R2017a.

TABLE IX. EXECUTION TIME IN SECONDS FOR THE ENCRYPTION DES, RSA AND PROPOSED ALGORITHM

Text size (bits)	**DES Encrypt**	**RSA Encrypt**	**Proposed algorithm Encrypt**	**Total time of the Proposed algorithm**
128	0.054945	0.0549	0.055	0.100
256	0.054946	0.1098	0.058	0.100
512	0.070976	0.2197	0.083	0.137
1000	0.1418	0.3846	0.125	0.187
2000	0.2835	0.7142	0.157	0.228
5000	0.6816	1.7032	0.479	0.704
10000	1.3601	3.402	1.000	1.441

Source: [22] and Authors.

TABLE X. Comparison Between AES, DES, RSA and Proposed Algorithm

Factors	AES	DES	RSA	Proposed algorithm
Development year	2000	1977	1978	2017
Key's length	128, 192, 256 bits	56 bits	>1024 bits	105.11 bits
Key for encrypt and decrypt	Same password	Same password	Different password	Same passwords
Block size	128 bits	64 bits	least 512 bits	8 bits
Scalability	Not scalable	Scalable (It depends on the block size and the password)	Not scalable	Scalable (It depends on the initial text's size)
Algorithm's kind	Symmetric	Symmetric	Asymmetric	Symmetric
Execution time	Fast	Moderate	slow	Fast
Key's tank	Necessary	Necessary	Necessary	Necessary
Inherent vulnerability	Force attack	Force attack, linear and differential cryptanalysis attack	Force attack and "Oracle attack"	Force attack, linear and differential cryptanalysis attack
Rounds	10/12/14	16	1	Depends on the maximum iteration parameter value or the optimal iteration number

Source: [3] and Authors.

References

[1] Carle, G. (2003). *Network Security Chapter 7: Cryptographic Protocols*. In: http://www.ccs-labs.org/~dressler/teaching/netzsicherheit-ws0304/07_CryptoProtocols_2on1.pdf.

[2] Abd, D., Abdual, H. and Hadhoud, M. (2010). *Evaluating the Performance of Symmetric Encryption Algorithms.* International Journal of Network Security, Vol. 10, No. 3, pp. 213–219.

[3] Mahajan, P. and Sachdeva, A. (2013). A Study of Encryption Algorithms AES, DES, and RSA for Security. *Global Journal of Computer Science and Technology*, vol. 13. In: https://computerresearch.org/index.php/computer/article/view/272/272.

[4] Beegom, B. y Jose, S. (2017). "*An Enhanced Cryptographic Model Based on DNA Approach*". Presented at: *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*, Idukki, India. Doi: 10.1109/ICECA.2017.8212824.

[5] Srilatha, N. and Murali, G. (2016). "*Fast three level DNA Cryptographic technique to provide better security*". Presented at: *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, Bangalore. Doi: 10.1109/ICATCCT.2016.7912037.

[6] Lawrence, E., (2006). *HTTPS Security Improvements in Internet Explorer 7*. In: https://msdn.microsoft.com/en-us/library/bb250503(v=vs.85).aspx.

[7] Dierks, T. and Rescorla E., (2008). *The Transport Layer Security (TLS) Protocol, Version 1.2.* In: https://tools.ietf.org/html/rfc5246.

[8] Holland, J. (1975). *Adaptation in Natural and Artificial Systems*. London, England: MIT Press, pp. 6-16.

[9] Goldberg, D. (1989). Genetic Algorithms in Search, Optimization and Machine Learning. Boston, USA: Addison-Wesley, pp. 1-68.

[10] Hermawanto, D. (2013). Genetic Algorithm for Solving Simple Mathematical Equality Problem. *e-print arXiv:* 1308.4675.

[11] Syswerda, G. (1991). *Schedule optimization using genetic algorithms*. New York, USA: Lawrence Davis, pp. 322-349.

[12] Jhajharia, S., Mishra, S., and Bali, S. (2013). "Public key cryptography using neural networks and genetic algorithms". Presented at: 2013 Sixth International Conference on Contemporary Computing (IC3), Noida, India. Doi: https://doi.org/10.1109/IC3.2013.6612177

[13] Simmons, G.J. (1988). A survey of Information Authentication. *IEEE Proceedings*, vol. 76, pp. 603-620, doi: https://doi.org/10.1109/5.4445

[14] Bhowmik, S. and Acharyya, S. (2011, June). "*Image Cryptography: The Genetic Algorithm Approach*". Presented at: *International Conference on Computer Science and Automation Engineering*, Shanghai, China. Doi: 10.1109/CSAE.2011.5952458

[15] Conci, A., Brazil, A.L., Leal, S.B. and MacHenr, T. (2015, November). "*AES Cryptography in Color Image Steganography by Genetic Algorithms*". Presented at: *2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*, Marrakech, Morocco. Doi: https://doi.org/10.1109/AICCSA.2015.7507100

[16] Al-Tabtabai, H., and Alex, A.P. (1999). Using Genetic Algorithms to solve optimization problems in construction. *Engineering, Construction and Architectural Management*, vol. 6, Issue: 2, pp. 121-132, doi: https://doi.org/10.1108/eb021105

[17] Traberg, G., Moliari, L., Venosa, P. and Macia, N. (2015). "*Automatizando el descubrimiento de portales de autenticación y evaluación de la seguridad mediante ataques de fuerza bruta en el marco de una auditoría de seguridad".* Presented at: *XXI Congreso Argentino de Ciencias de la Computación,* Argentina. In: http://sedici.unlp.edu.ar/handle/10915/50589

[18] Shannon, C.E. (1949). Communication Theory of Secrecy Systems. *The Bell System Technical Journal*, vol. 28, Issue: 4, doi: http://dx.doi.org/10.1002/j.1538-7305.1949.tb00928.x

[19] Othman, H., Hassoun, Y. and Owayjan, M. (2015, October). "*Entropy Model for Symmetric Key Cryptography Algorithms Based on Numerical Methods*". Presented at: *2015 International Conference on Applied Research in Computer Science and Engineering (ICAR)*, Beirut, Lebanon. Doi: 10.1109/ARCSE.2015.7338142

[20] Gauss, C.F. (1801) and Clarke, A.A. (1965). *Disquisitiones Arithmeticae Translated.* New Haven, Connecticut, USA, pp. 1-4.

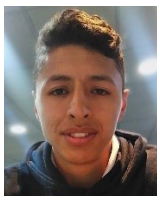
[21] Jimenez, R., Gordillo, E. and Rubiano, G. (2012). *Teoría de Números (para principiantes)*. Bogotá, Colombia: Pro–Offset Editorial Ltda., pp. 98-104.

[22] Narasimham, C. and Jayaram, P. (2008). *Evaluation of performance characteristics of Cryptosystem using text files.* Journal of Theoretical and Applied Information Technology, pp. 55-59.

[23] Zaldaña, HR. and Castañeda E. (2015). *The Use of Genetic Algorithms in UV Disinfection of Drinking Water*. International Journal of Interactive Multimedia and Artificial Intelligence, vol. 3, pp. 43-48.

Jefferson Rodríguez Rodríguez

Jefferson Rodríguez Rodríguez was born in Bogotá, Colombia in 1994. Aspiring to Computer Engineer in Universidad Distrital Francisco José de Caldas. He works in Exsis SAS since 2017 as a systems analyst. His area of interest includes AI, computer security and backend development.



Brayan Julián Corredor

Brayan Julián Corredor was born in Bogotá, Colombia in 1994. Computer Engineer from Universidad Distrital Francisco José de Caldas since 2017. He works in Stefanini IT since 2017 as an Engineer jr. His area of interest includes AI, computer security and frontend development.

César Suárez Parra

César Suárez Parra was born in Bogotá, Colombia in 1956. He got his industrial engineering degree from Fundacion Universidad Incca de Colombia in 1980, and his mechanical engineer degree from Fundacion Universidad Incca de Colombia in 1992. He received his specialist degree in applied mathematics from Universidad Sergio Arboleda in 2003 and his master degree in materials and manufacturing processes from Universidad Nacional de Colombia in 1998. He is a teacher in Universidad Distrital Francisco José de Caldas, dedicated to higher education in engineering and investigations groups. He has a strong research record in cryptography, complexity and genetic algorithms.

Humanoid Localization on Robocup Field using Corner Intersection and Geometric Distance Estimation

M. N. Sudin[1], S. N. H. Sheikh Abdullah[1], M. F. Nasudin[2] *

[1] Center for Cyber Security, Faculty of Information System and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor (Malaysia)
[2] Center for Artificial Intelligence Technology, Faculty of Information System and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor (Malaysia)

Received 18 February 2019 | Accepted 13 March 2019 | Published 9 April 2019

Abstract

In the humanoid competition field, identifying landmarks for localizing robots in a dynamic environment is of crucial importance. By convention, state-of-the-art humanoid vision systems rely on poles located outside the middle of the field as an indicator for generating landmarks. However, in compliance with the recent rules of Robocup, the middle pole has been discarded to deliberately provide less prior information for the humanoid vision system to strategize its winning tactics on the field. Previous localization method used middle poles as a landmark. Therefore, robot localization tasks should apply accurate corner and distance detection simultaneously to locate the positions of goalposts. State-of-the-art corner detection algorithms such as the Harris corner and mean projection transformation are excessively sensitive to image noise and suffer from high processing times. Moreover, despite their prevalence in robot motor log and fish-eye lens calibration for humanoid localization, current distance estimation techniques nonetheless remain highly dependent on multiple poles as vision landmarks, apart from being prone to huge localization errors. Thus, we propose a novel localization method consisting of a proposed corner extraction algorithm, namely, the contour intersection algorithm (CIA), and a distance estimation algorithm, namely, analytic geometric estimation (AGE), for efficiently identifying salient goalposts. At first, the proposed CIA algorithm, which is based on linear contour intersection using a projection matrix, is utilized to extract corners of a goalpost after performing an adaptive binarization process. Then, these extracted corner features are fed into our proposed AGE algorithm to estimate the real-word distance using analytic geometry methods. As a result, the proposed localization vision system and the state-of-the-art method obtained approximately 3-4 and 7-23 centimeter estimation errors, respectively. This demonstrates the capability of the proposed localization algorithm to outperform other methods, which renders it more effective in indoor task localization for further actions such as attack or defense strategies.

Keywords

Analytic Geometry, Localization, Robotics, Single Camera, Visual Intelligence.

DOI: 10.9781/ijimai.2019.04.001

I. Introduction

During the last 65 years when computer processing remained in its infancy, the application of computer processing could not be extended to visual recognition because it merely consisted of a simple input and output operation. Hence, computer vision and recognition represents one of the culminating achievements of the aforementioned research domain, although a huge challenge has yet to be overcome. Any area of computer processing research is geared toward mimicking human-like visual perception and interpretation to translate an image into recognizable objects [1]. Currently, in the wake of rigorous computer hardware and software advancements, image processing and recognition no longer remains as a mere theoretical possibility. Moreover, it is deemed as an essential advantage to incorporate such technologies into a variety of industrial enterprises, as in visual recognition systems for quality control and assessment, for instance, prior to public consumption [8][9]. Evidently, the application of the technology is in high demand in the manufacture of mobile phones, compact cameras and robot soccer.

In robot soccer, artificial vision processing plays a crucial role in recognizing goal posts, teammates, opponents and balls. Based on prior information, the above process is of great utility in localizing humanoid robots in field play to prevent rogue and random movements. In addition, it is indeed expedient for any robot to be able to recognize its location on a field when employing attacking or defending strategies. Based on the 2013 rules of the Robot Soccer competition, poles colored blue-yellow-blue on the left and yellow-blue-yellow on the right, which functioned as landmarks, were set up outside the middle field to assist in robot self-localization on the field [5][16]. However, a different set of rules are currently in place whereby the poles are to be replaced with the goalpost as the landmark. Consequently, a serious challenge must be confronted as a result of the over- reliance on these poles as landmarks by the majority of preceding studies. This calls for a new localization technique that can substitute the missing poles at the middle of the field with goalposts as the reference points. One way to enable recognition of goalposts prior to localization is to consider prior information, namely the properties or features of the object to be recognized, including the shape, color, size and total number of corners.

Corner detection is the most used feature in the object recognition process, as evidenced by the range of corner detection techniques available for various applications [2][7]. Recently, various corner

* Corresponding author.
E-mail addresses: nudin@siswa.ukm.edu.my (M. N. Sudin), snhsabdullah@ukm.edu.my (S. N. H. Sheikh Abdullah), mfn@ukm.edu.my (M. F. Nasudin).

detection methods such as JUDOCA [13], CPDA [2], ANDD [19] and MPT [17] have been developed. Nonetheless, most of these techniques are very time consuming for merely extracting corners and are overly sensitive to the dynamics of the environment. Distance estimation processes for localization are in need of a new technique in response to the changes in the rules and to the loss of prior information while overcoming one of the biggest challenges in robot soccer: performing distance estimation processes with minimal equipment. Most state-of-the-art techniques produce non-negligible errors in distance estimation, which in turn necessitates an additional device, such as an odometer, to increase accuracy [15][13] or the use of an RGB-D camera with OpenPTrack [18][20]. Another approach used a simple localization with a camera [22][23][24]. With this approach, a 3 robot need to put extra sensor (such as wheel odometers and inertial navigation sensors) and making it accurate and desirable for our system. However, in our application, no extra sensors can be used other than camera based on Robocup rules. Thus, a new and improvised technique of visual recognition is to be developed with the objective of higher efficiency in distance estimation, lower processing time and greater responsiveness to the dynamics of the environment.

II. The Proposed Method

There are two main tasks in developing a new vision system for a robot to self-localize in field play during robot soccer competitions. First, a technique to extract corner features is to be applied, followed by the use of these features to estimate distances between the robot and the goalpost as the main landmark. Fig. 1 shows the previous match rules.

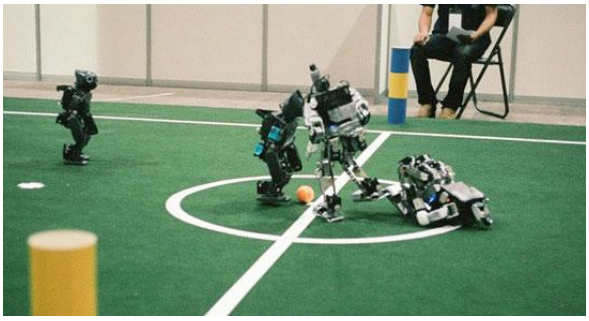


Fig. 1. Robot soccer competition game play with middle pole as a landmark in 2012.

A. Line Intersection

The raw images suffer from superfluous information and dynamics. Accordingly, before a corner is extracted, the raw images must first be converted to a binary image containing only values of 0 and 1 [3] [4]. Because the testing data set is in grayscale, adaptive thresholding is used as a binarization method. Equations (1) and (2) show the binarization method:

$$f'_{(x,y)} = \begin{cases} 1 & if\ f_{(x,y)} > T_{(x,y)} \\ 0 & else \end{cases} \quad (1)$$

$$T_{(x,y)} = \frac{1}{2\pi\sigma^2} e^{-[(x-u_x)^2+(y-u_y)^2]/2\sigma^2} \quad (2)$$

Based on Equation (1), $f_{(x,y)}$ is a pixel value in the raw images, and $f'_{(x,y)}$ is a pixel value upon binarization. $T_{(x,y)}$ is a threshold value derived from the Gaussian equation for a distribution function for non-correlated variables with variations x and y having normal distributions for the same standard deviation. In this research, the threshold value is the middle value of the 3×3 neighboring pixels. Fig. 2 shows goal images before and after thresholding.

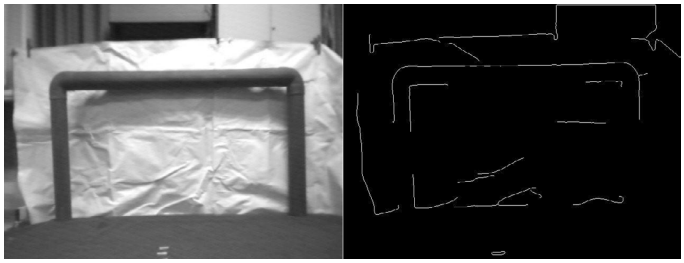


Fig. 2. (a) Grayscale image for goalpost; (b) Binary image for goalpost.

Upon applying the conversion method, the binary images must be analyzed pixel by pixel. Hence, square matrices (kernel) of size 3×3,5×5,8×8 and 11×11 are generated to process the binary images line by line. By using these kernels, the x and y corners will be extracted when the kernel processes the corners in the image. Four coordinates are recorded if the corner is detected, with two coordinates on the x – axis and two coordinates on the y - axis. The next equations show the selection process for the row and column selected by each kernel to generate the four coordinates in the kernel.

$$k = \{a, (s-1) - a\} \quad (3)$$

Based on Equation (3), k is a row or column index for a kernel of size *s*,{3,5,8,11}, and a is a constant value {0,1,2,3} incremented by the kernel size. The blue line in Fig. 3 shows the horizontal and vertical lines chosen for each kernel size.

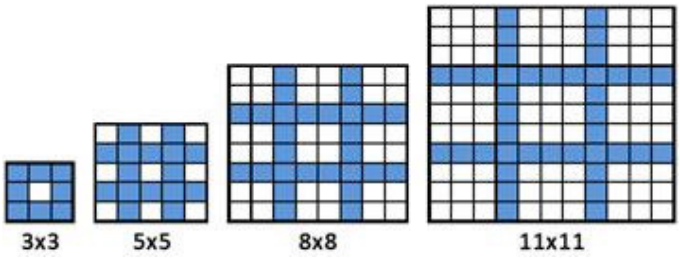


Fig. 3. Horizontal and vertical lines for each kernel.

Based on Fig. 3, if any part of the blue lines includes a white pixel (value of 1), then the coordinate of the pixel is recorded and processed when all four coordinates are recorded. The vertical blue line is utilized to obtain the coordinates from the horizontal contour, and the horizontal blue line is utilized to obtain the coordinates from the vertical contour. Equations (4) and (5) show the process of white pixel detection for each line assuming P_l for the horizontal contour and P_t for the vertical contour.

$$P_l(x,y) = \sum_{x=k}^{2} \sum_{y=0}^{L} \left[\begin{cases} x = k, y = L & if\ f'_{(x,y)} = 1 \\ L = s & if\ f'_{(x,y)} = 0 \end{cases} \right] \quad (4)$$

$$P_t(x,y) = \sum_{x=k}^{2} \sum_{y=0}^{L} \left[\begin{cases} x = k, y = L & if\ f'_{(x,y)} = 1 \\ L = s & if\ f'_{(x,y)} = 0 \end{cases} \right] \quad (5)$$

After four coordinates are obtained, we can approximate the gradient value from both contours, namely, the vertical and horizontal contours. Then, using the gradient value from both contours, any intersection point from these contours can be assumed to be a corner. To extract the coordinates of the corner, linear equation theory and analytic geometry theory are applied by manipulating all four coordinates obtained previously.

$$P_l(y) = m_l(x) + c_l \quad (6)$$

$$P_t(y) = m_t(x) + c_t \quad (7)$$

We assume that the first point is on the horizontal contour, P_{l1} *(x,y)*, and that the second point is on the horizontal contour, P_{l2} *(x,y)*, on the same contour line; hence, they share the same gradient value, and m_l and the intersection points are at y-axis, c_l. This is also the case with the vertical contour, where the first point, P_{t1} *(x,y)*, and the second point, P_{t2} *(x,y)*, are assumed to be on the same vertical contour and share the same gradient value, m_t, and intersection point on the y-axis, c_t. To determine the gradient value for each contour, Equations (6) and (7) were placed into a new form.

$$m_l = \frac{P_{l2}(y) - P_{l1}(y)}{P_{l2}(x) - P_{l1}(x)} \tag{8}$$

$$m_t = \frac{P_{t2}(y) - P_{t1}(y)}{P_{t2}(x) - P_{t1}(x)} \tag{9}$$

Once the gradient value is obtained, the value of the intersection point on the y-axis for both contours, horizontal and vertical, can be obtained using the following equations.

$$c_l = P_l(y) - (m_l P_l(x)) \tag{10}$$

$$c_t = P_t(y) - (m_t P_t(x)) \tag{11}$$

When all the values are obtained, the intersection between the vertical contour and the horizontal contour is classified as a corner. If the vertical and horizontal contours are on the same contour, then the gradient value is equal to zero. Accordingly, if no corner is detected in the current kernel, then it will move on to the next kernel. The equations used to acquire the corner are defined below.

$$x_{corner} = \frac{(c_t - c_l)}{(m_t - m_l)} \tag{12}$$

$$y_{corner} = (m_l \times x_{corner}) + c_l \tag{13}$$

Finally, the corner coordinates are recorded and marked for further processing by the robot during localization. Fig. 4 shows the process for each kernel. The blue points represent the detected corner, and the red pixels represent the four pixels obtained on the vertical and horizontal contours.

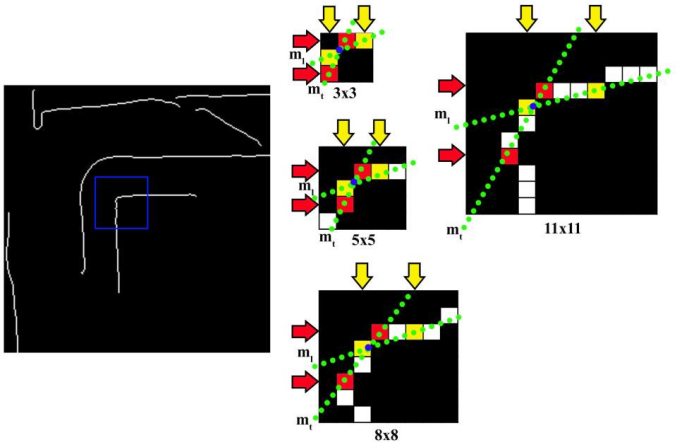


Fig. 4. Corner detection performed using line intersections. Two red pixels in P_t *(x,y)* and two yellow pixels in P_l *(x,y)* are used as coordinates to find the intersection point (blue pixel) of two green lines, m_t and m_l.

B. Analytic Geometry for Distance Estimation

In previous studies on camera calibration, using prior information such as the focal length of the camera, constant x and y values were extracted [12][13][14][20]. All these values can be obtained as intrinsic camera values. Aside from these intrinsic values, extrinsic values are also present and are of great utility in reducing distortion in an image captured by a camera using a convex lens. To use analytical geometry approaches, the produced image must closely resemble that in real time [12]. The distortion of an image produces errors in distance estimation. To address convex-lens-induced distortions in images, 10 images with a chessboard were captured at random orientations to determine the distortion difference relative to the real time imagery as shown in Fig. 5.

Fig. 5. Self-collection data from MVLab used to obtain the intrinsic and extrinsic values.

Based on Zhang's methods [21], the intrinsic, extrinsic, and pinhole values of the camera are extracted and processed. These processes only need to be performed once as long as the same camera and lens are used. The equation shows the data value and the intrinsic value obtained from the first calibration. Given that R is an image rotation and T is an image transformation, we have

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}, T = \begin{bmatrix} t_{11} \\ t_{21} \\ t_{31} \end{bmatrix} \tag{14}$$

Upon repairing the image, some of the features of the goal post are extracted. Based on Robocup rules, the pole had a height of 10 cm and a width of 60 cm. The goalpost should be yellow and blue only, and the shape of the goal should be rigid and hard. Therefore, the color and shape of the goal post are suitable for extraction. To extract the color, HSV color range values are considered due to their higher robustness to light intensity variations compared to the RGB color range. RGB colors are more sensitive to even the slightest change in color. Moreover, RGB color ranges produce more noise in the data and hence this makes difficult to clearly distinguish the goal post shape [10]. Fig. 6 shows the color of the goal post extracted using the HSV color range.

Fig. 6. Color extracted from a goal post.

After the color is extracted, the shape of the goal was determined. The purpose of the shape detection is to ensure that only the complete shape of the goal is input to the next process. If the camera is too close to the goal post, the shape of the goal post becomes larger than the camera frame size, and vice versa. An improper goal post size renders the next process more cumbersome and prone to error. Hence, measurements such as density, elongation, roughness, convexity, height and width were defined beforehand. If the goalpost does not meet the predefined criteria, then it is considered to be incomplete, resulting in the termination of the process. In turn, a re-calibration of the camera and redefinition of the color have to be performed. Fig. 7 shows an incomplete goal post.

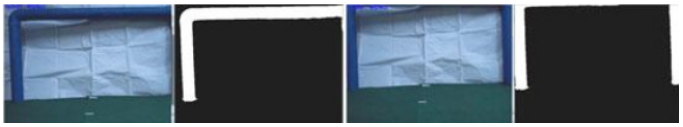


Fig. 7. Example of an incomplete goal shape from self-collection data set in MVLab.

If the goal shape can be viewed completely, the goal corner will be extracted using the above corner detection method. The purpose of this action is to compare and relate actual and virtual distances between two corners. According to Siswantoro et al. [14], the object size in a camera image is relatively equivalent to the actual object size in real time. Using a similar approach, distance estimation can be performed after considering the relative factor that causes differences in measurement in real time and in virtual images. The relation between images from the camera and those obtained in real time can be determined via prior camera calibration along with the intrinsic values. The equation below shows the relation between the image camera coordinates and the real-time coordinates.

$$\begin{bmatrix} f_x & 0 & C_x \\ 0 & f_y & C_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c/z_c \\ y_c/z_c \\ 1 \end{bmatrix} = \begin{bmatrix} x_{img} \\ y_{img} \\ 1 \end{bmatrix} \tag{15}$$

In Equation (15), the left-hand side represents the coordinates in the image plane, and the right-hand side represents rotation and transformation coordinates in the camera plane. The values of f_x and f_y are the focal length of the lens on the x and y axis. C_x, C_y and Z_c are confession numbers from the intrinsic values. For distance estimation, the image coordinate should be obtained beforehand so that it can subsequently be related to actual parameters. The next equations demonstrate how the method obtains the coordinates in the camera plane.

$$X_{cam} = Z_c \left[\frac{x_{img} - C_x}{f} \right] \tag{16}$$

$$Y_{cam} = Z_c \left[\frac{y_{img} - C_y}{f_y} \right] \tag{17}$$

$$Z_{cam} = \frac{f_x + f_y}{2} \tag{18}$$

As stated previously, the goal post is taken as the main landmark for parameter assessment. For the distance estimation process, the height of the left and right goal posts are selected as a reference to compare with the actual height and image height. Fig. 8 shows the location of the camera coordinate, image coordinate and real-time coordinate.

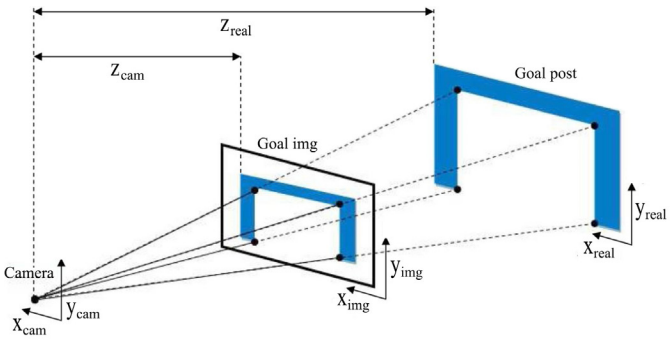


Fig. 8. Relation between camera, image and real-time coordinate.

Based on Fig. 8, it is observed that the goal post image is reduced in size compared to the real goalpost. By relating the camera coordinate to the real coordinate, the value of Z_{cam} becomes a scaled version of Z_{real}, which represents the actual distance between the camera and the goal post. Hence, the distance estimation can be performed using this equation,

$$\frac{l}{a} = \frac{L}{a+b} \tag{19}$$

where l is the distance between two points in an image, a is the value of Z_{cam}, L is an actual two-point distance in real time of the goalpost, and a + b is a Z_{real} or the estimated distance. In this paper, l and L are representing two points as the right and left-hand sides of the goalpost. The calculation is performed twice to increase the accuracy of the estimated distance.

III. Result and Discussion

One of the objectives of this research was to develop a new corner detection technique for goalpost corner extraction. The technique must be efficient, robust and less time consuming to process. Accordingly, in assessing the performance, the proposed corner detection is to be compared with state-of-the-art corner detection methods such as JUDOCA [13], CPDA [2], ANDD [19] and MPT [17]. A common test for corner detection is the localization error test, which attempts to obtain the average error for each point within 3 pixels in each corner in an image. The smaller the localization error produced by a technique, the more accurately the technique can detect corners.

$$L_e = \sqrt{\frac{1}{N_r} \sum_{i=1}^{N_r} [(x_{oi} - x_{ti})^2 + (y_{oi} - y_{ti})^2]} \tag{20}$$

Based on Equation (20), x_{oi} and y_{oi} are the original coordinates of a corner in an image, and x_{ti} and y_{ti} are detected coordinates at the current $point_i$. N_r is the total number of corners detected in the same image. The benchmark data set for corner detection from Awrangjeb research [2] was commonly used to measure accuracy as show in Fig. 9 and Fig. 10 shows the results when we apply localization error to the current state-of-the-art corner detection methods.

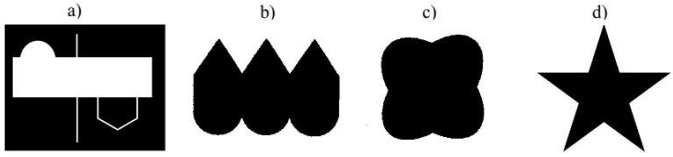


Fig. 9. Benchmark data set for corner detection.

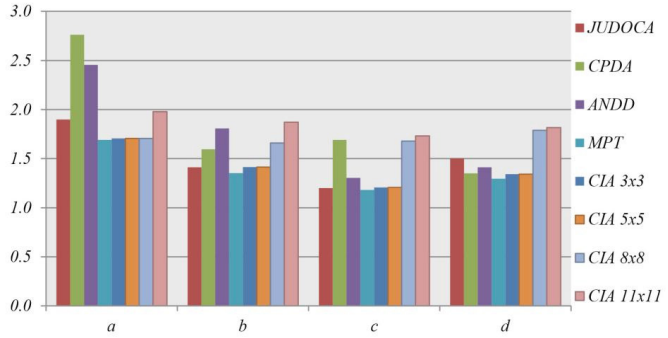


Fig. 10. The localization errors for JUDDOCA, CPDA, ANDD, MPT and our proposed CIA corner detection techniques when the kernel size is 3x3, 5x5, 8x8 and 11x11.

Based on Fig. 10, MPT [17] yields the best results by providing the lowest localization average error across the entire dataset, followed by the line intersection for the 5×5 kernel, JUDOCA, 3×3 kernel, 8×8 kernel, ANDD, CPDA and 11 × 11 kernel. The MPT method obtains the highest accuracy for extracting corners from images; however, the process requires approximately 2 seconds for each picture, a severe disadvantage for Robocup competitions. Hence, the line intersection method is selected because it is the second-most accurate method and requires only 0.049 seconds for extracting a corner.

For the distance estimation test, the proposed method is compared with the dual camera stereo vision method for the image quality assessment [15]. Fig. 11 shows the test setup on the robot. The use of a stereo camera is advantageous due to its similarity to human vision. The test was performed three times during actual game play at distances between 350-190 cm, each 10 cm with straight vision to the landmark (0 degree), 20 degrees to the left of the landmark, and 20 degrees to the right of the landmark to represent vision from various sides of the field. Table I shows a comparison of the results.

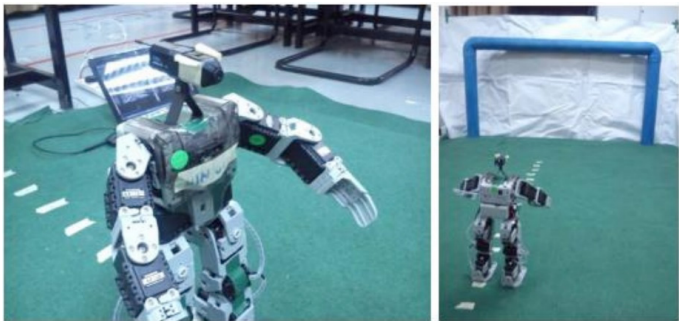


Fig. 11. Wireless camera attached to a robot for the distance estimation test.

TABLE I. Difference Between Stereo and Single Camera Via Proposed AGE Method

	Stereo						Proposed (AGE)					
Actual	20°	error	0°	error	-20°	error	20°	error	0°	error	-20°	error
190	199.94	9.94	199.94	9.94	199.94	9.94	184.96	5.04	179.08	10.92	185.55	4.45
200	207.34	7.34	207.34	7.34	207.34	7.34	196.89	3.12	196.48	3.53	197.30	2.70
210	215.32	5.32	215.32	5.32	215.32	5.32	208.69	1.31	209.18	0.82	209.18	0.82
220	233.26	13.26	233.26	13.26	233.26	13.26	219.35	0.65	218.82	1.19	219.26	0.75
230	239.16	9.16	239.16	9.16	239.16	9.16	228.73	1.27	228.73	1.27	228.67	1.33
240	243.40	3.40	243.40	3.40	243.40	3.40	239.80	0.20	238.21	1.79	238.20	1.80
250	244.48	5.52	244.48	5.52	244.48	5.52	243.20	6.80	248.03	1.97	245.49	4.51
260	254.47	5.53	254.47	5.53	254.47	5.53	256.24	3.76	256.95	3.05	256.77	3.23
270	259.87	10.13	259.87	10.13	259.87	10.13	265.47	4.53	265.09	4.91	268.74	1.26
280	266.58	13.42	266.58	13.42	266.58	13.42	276.03	3.97	276.03	3.97	275.92	4.08
290	279.91	10.09	279.91	10.09	279.91	10.09	281.38	8.62	279.87	10.13	286.46	3.54
300	283.72	16.29	283.72	16.29	283.72	16.29	295.64	4.36	290.24	9.76	297.51	2.49
310	308.22	1.78	308.22	1.78	308.22	1.78	306.74	3.26	306.79	3.21	305.75	4.25
320	311.01	8.99	311.01	8.99	311.01	8.99	311.52	8.48	313.54	6.46	318.04	1.96
330	325.31	4.69	325.31	4.69	325.31	4.69	325.80	4.20	325.99	4.01	325.68	4.32
340	329.31	10.69	329.31	10.69	329.31	10.69	335.77	4.23	337.80	2.20	335.65	4.35
350	349.89	0.11	349.89	0.11	349.89	0.11	349.65	0.35	346.26	3.74	346.33	3.67
average		7.98		7.98		7.98		3.77		4.29		2.91

Based on Fig. 12, the field simulation shows the actual radial distance (black line) from the center of the goal post and the goal post clearly observed by the camera on the robot. In Table I, both techniques (stereo and proposed) are able to estimate the distance the around field with straight to the landmark and 20 degreed offset to the left and right. On average, for each angle, the stereo technique produces constant mean errors of approximately 7.980 cm, demonstrating that the technique was robust and stable at any angle. The proposed technique produces inconsistent mean distance errors for certain angles. However, the margin errors remain less than those of the stereo technique. Overall, in distance estimation, the AGE technique produces an improved accuracy of 4.322 cm compared to the stereo-camera-based technique according to the summary in Table II. Concerning processing time, the AGE technique requires 0.54672 milliseconds to estimate the distance based on Fig. 13, whereas in Fig. 14, the error percentage is initially large, wherein the proximity to the landmark is the closest but undergoes substantial decrements as the distance increases. Summarizing, Fig.15 shows that the stereo technique (green) produces huge min and max error values but produces consistent median errors at any angle given; meanwhile, the AGE technique produces smaller min and max error values but does not produce very consistent errors at some offsets. It addition, the AGE technique is less accurate for straight lines, demonstrated by its larger error range compared to the offset value; however, the error is still small compared to the stereo technique. The error obtained using the AGE technique may be smaller because this technique only uses a single image from a single camera, whereas the stereo camera technique must patch together both images from both cameras.

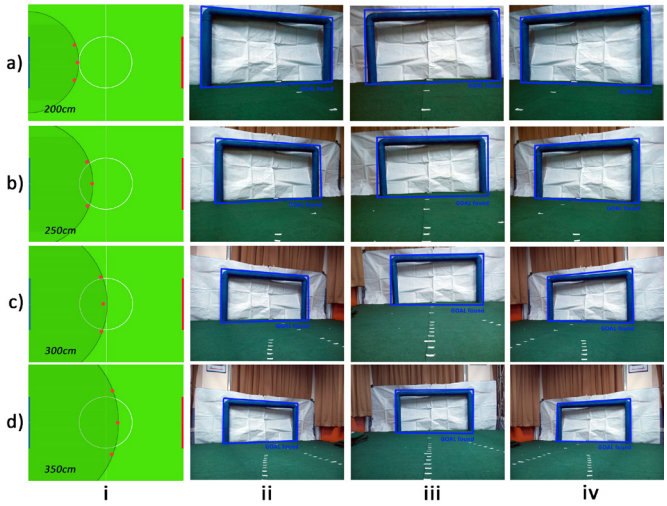


Fig. 12. The actual distance (black line) and distance estimation simulator results (i) based on images captured by our humanoid robot at (ii)+20◦, (iii) 0◦ and (iv) -20◦. The red dot in the simulator represents the proposed detected point based on our proposed (AGE) technique from a radial distance of a) 200 cm b) 250 cm c) 300 cm and d) 350 cm.

TABLE II. Summary Difference Between Stereo and Single Camera Via Proposed AGE Method

Technique	Data	Variants	Mean error (cm)	Percentage (%)
Stereo	51	18.04454331	7.980	2.955
Proposed (AGE)	51	6.30001502	3.658	1.355

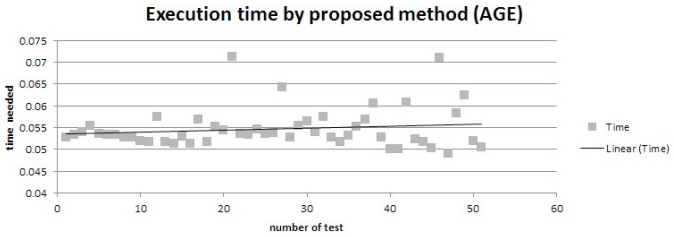


Fig. 13. Time required for proposed method (AGE) to estimate distance.

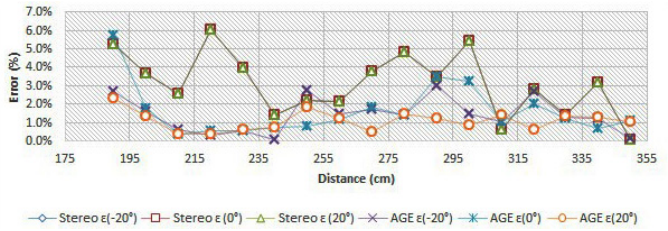


Fig. 14. Fluctuating relation between error and distance produced by both methods (AGE & stereo).

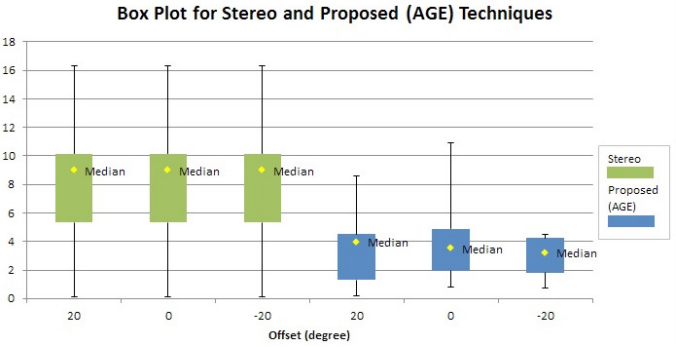


Fig. 15. Box plot for stereo and proposed (AGE) technique.

Furthermore, our proposed AGE method is simple and practical because it only uses a predetermined camera information and rules given. Considering those steps, this technique is more desirable to incorporate into the Robocup competition due to its minimal error in robot localization.

IV. Limitations and Conclusions

Having performed the test, it is reassuring to conclude that this technique provides highly accurate corner detection and distance estimation compared with the other techniques. This evidently attests to the capability of the proposed technique in surmounting the most persistent predicament of object recognition at a significantly reduced execution time. However, this method continues to face a few setbacks, wherein this technique is unable to estimate distances of less than 190 cm due to the rather enlarged size of the goalpost image relative to the camera frame. This results in the incomplete detection of the shape of the goal post. Despite that, localization remains applicable at other areas within 8.0 meters, making it suitable for competition because of its accuracy, speed and robustness.

Acknowledgment

This research was funded by an FRGS/1/2014/CT07/UKM/02/5 grant entitled Overlapped Irregular Shape Descriptor based on Non-Linear Approach and AP-2017-005/2 grant entitled Using STEM Data Through Smart Self-Crime Prevention at Schools for Open Data Readiness. We would like to thank Zamalah: Fellowship at Universiti Kebangsaan Malaysia (UKM).

References

[1] Nur Ariffin M. Z. 2013. Incorporating Camshift Algorithm and Support Vector Machine for Object Tracking. Master's Thesis in Universiti Kebangsaan Malaysia.

[2] Awrangjeb, M. Lu, G. 2008. Robust image corner detection based on the chord-to-point distance accumulation technique. Multimedia, IEEE Transactions on. 10(6): 1059-1072.

[3] Birchfield, S. T. Rangarajan, S. 2005. Spatiograms versus histograms for region-based tracking. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2: 1710-1714.

[4] Birchfield, S. T. Rangarajan, S. 2007. Spatial histogram for region-base tracking. Electronics and Telecommunications Research Institute (ETRI) Journal. 29(5): 697-699.

[5] Tian, B., Ng, C. L. Chew C. M. 2010. Self-localization of humanoid robots with fish-eye lens in a soccer field. Robotics Automation and Mechatronics (RAM), 2010 IEEE Conference on, pp. 522-527.

[6] Bouthemy, P. Francois, E. 1993. Motion segmentation and qualitative dynamic scene analysis from an image sequence. International Journal of Computer Vision. 10(2): 157-182.

[7] Bowyer, K., Kranenburg, C., Dougherty, S. 1999. Edge detector evaluation using empirical ROC curves. Computer Vision and Pattern Recognition, IEEE Computer Society Conference on. 1: 354-359.

[8] Bradski, G. R. Kaehler, A. 2008. Learning OpenCV: Computer Vision with the OpenCV Library. California: OReilly Media Inc.

[9] Bradski, G. R. 1998. Computer vision, face tracking for use in perceptual user interface. Intel Technology Journal. 2(2): 12-21.

[10] Chen, W., Shi, Y. Q. Xuan, G. 2007. Identifying computer graphics using HSV colour model and statistical moments of characteristic functions. Proceeding of IEEE International Conference on Multimedia and Expo, hlm. 1123-1126.

[11] Cheng, Y. 1995. Meanshift, mode seeking and clustering. IEEE Transaction on Pattern Analysis and Machine Intelligence. 17(8): 790-799.

[12] Davison A. J. 2003. Real-time simultaneous localisation and mapping with a single camera. Proceedings. Ninth IEEE International Conference on Computer Vision, vol. 2, pp. 1403-1410.

[13] Elias, R. Laganiere, R. 2012. JUDOCA: JUnction Detection Operator Based on Circumferential Anchors. Image Processing, IEEE Transactions on, Vol. 21, no. 4, pp. 2109-2118.

[14] Siswantoro, J, Prabuwono, AS Abdullah, A 2013. 'Real World Coordinate from Image Coordinate Using Single Calibrated Camera Based on Analytic Geometry'. in Communications in Computer and Information Science. vol. 378 CCIS, Springer Verlag, pp. 1-11, 2nd International Multi-Conference on Artificial Intelligence Technology, M-CAIT 2013, Shah Alam, 28-29 Ogos., 10.1007/978-3-642-40567-9-1.

[15] Pirahansiah, F.; Abdullah, S.N.H.S.; Sahran, S.," Camera calibration for multi-modal robot vision based on image quality assessment," in Control Conference (ASCC), 2015 10th Asian, pp.1-6, May 31 2015-June 3 2015 doi: 10.1109/ASCC.2015.7360336.

[16] Jens-Steffen Gutmann, Dhiraj Goel, Philip Fong, Mario E. Munich, Extensions to vector field SLAM for large environments, Robotics and Autonomous Systems, vol. 62, no. 9, September 2014, pp. 1248-1258, ISSN 0921-8890, http://dx.doi.org/10.1016/j.robot.2014.03.021.

[17] Kahaki, Seyed Mostafa Mousavi, Md Jan Nordin, and Amir Hossein Ashtari. "ContourBased Corner Detection and Classification by Using Mean Projection Transform." Sensors 14.3 (2014): 4126-4143.

[18] Matteo Munaro, Filippo Basso, Emanuele Menegatti, OpenPTrack: Open source multi-camera calibration and people tracking for RGB-D camera networks, Robotics and Autonomous Systems, vol. 75, Part B, January 2016, pp. 525-538, ISSN 0921-8890, http://dx.doi.org/10.1016/j.robot.2015.10.004.

[19] Shui, P. Zhang, W. 2013. Corner detection and classification using anisotropic directional derivative representations. Image Processing, IEEE Transactions on. 22(8): 3204-3218.

[20] Xiang Gao, Tao Zhang, Robust RGB-D simultaneous localization and mapping using planar point features, Robotics and Autonomous Systems, vol. 72, October 2015, pp. 1-14, ISSN 0921-8890, http://dx.doi.

org/10.1016/j.robot.2015.03.007.

[21] Zhang, Z. 2000. A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 11, pp. 1330-1334.

[22] P. An, Y. Liu, W. Zhang and Z. Jin, "Vision-Based Simultaneous Localization and Mapping on Lunar Rover," 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC), Chongqing, 2018, pp. 487-493. doi: 10.1109/ICIVC.2018.8492755.

[23] V. P. Kirnos, V. A. Kokovkina, V. A. Antipov and A. L. Priorov, "Simultaneous Localization and Mapping Using the Support Image Base Algorithm," 2018 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO), Minsk, 2018, pp. 1-4. doi: 10.1109/SYNCHROINFO.2018.8456965.

[24] X. Guo, S. Zhu, L. Li, F. Hu and N. Ansari, "Accurate WiFi Localization by Unsupervised Fusion of Extended Candidate Location Set," in IEEE Internet of Things Journal. doi: 10.1109/JIOT.2018.2870659.

Muhammad Nuruddin bin Sudin

First Degree in Computer Science major in Artificial Intelligence at The National University of Malaysia (UKM), Malaysia. He furthered his master study in the Robotic area in the same university. Runner up in Robot Soccer Worldcup competition under FIRA in 2012 and become a winner in 2013 for 5v5 mirosot category. Currently, he continued his Phd Study in the area of Cyber Security at Faculty of Information Science and Technology, UKM. His research focuses are Pattern Recognition, Robotic, Artificial Intelligence and Digital Forensics.

Siti Norul Huda Sheikh Abdullah

Received her first Degree in Computing at University of Manchester Institute of Science and Technology, United Kingdom. She furthered her master study in the area of Artificial Intelligence in Universiti Kebangsaan Malaysia. Later, she continued her Phd Study in the area of Computer Vision at Faculty of Electrical Engineering, Universiti Teknologi Malaysia. Starting the career, she involved in conducting several national and international activities such as Royal Police of Malaysia, Cyber Security Malaysia, Cyber Security Academia Malaysia, Federation of International RobotSoccer Association (FIRA), Asian Foundation, Global Ace Professional Certification Scheme, MIAMI, MACE and IDB Alumni. She is now holding a post as the Chairperson of Center for Cyber Security. Her research focuses are Digital Forensics, Pattern Recognition, and Computer Vision Surveillance System. She has published two books entitled "Pengecaman Pola" or "Pattern Recognition" and "Computational Intelligence for Data Science Application" and more than 50 and 100 of journal and conferences manuscripts correspondingly.

Mohammad Faidzul bin Nasrudin

He is an associate professor at the Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (UKM) and an active researcher at the Center for Artificial Intelligence Technology (CAIT), UKM. He obtained his Bachelor of Business Administration degree majoring in Computer Information System at the Western Michigan University, USA. He received a master's from UKM and a doctorate in Artificial Intelligence by joining the Malaysia - Imperial College London Doctoral Programme. He has coauthored several books and published more than 60 conference and journal papers. His current research focuses on document analysis and recognition, robot vision and metaheuristic optimization.

MWAND: A New Early Termination Algorithm for Fast and Efficient Query Evaluation

Zemani Imene Mansouria*, Zekri Lougmiri*, Senouci Mohamed

Department of Computer Sciences, LAPECI Laboratory, Ahmed Ben Bella Oran1 University (Algeria)

Received 22 October 2018 | Accepted 22 March 2019 | Published 10 April 2019

Abstract

Nowadays, current information systems are so large and maintain huge amount of data. At every time, they process millions of documents and millions of queries. In order to choose the most important responses from this amount of data, it is well to apply what is so called early termination algorithms. These ones attempt to extract the Top-K documents according to a specified increasing monotone function. The principal idea behind is to reach and score the most significant less number of documents. So, they avoid fully processing the whole documents. WAND algorithm is at the state of the art in this area. Despite it is efficient, it is missing effectiveness and precision. In this paper, we propose two contributions, the principal proposal is a new early termination algorithm based on WAND approach, we call it MWAND (Modified WAND). This one is faster and more precise than the first. It has the ability to avoid unnecessary WAND steps. In this work, we integrate a tree structure as an index into WAND and we add new levels in query processing. In the second contribution, we define new fine metrics to ameliorate the evaluation of the retrieved information. The experimental results on real datasets show that MWAND is more efficient than the WAND approach.

Keywords

Evaluation Measures, Information Retrieval, Large Inverted List, MWAND, Query Processing, Top-k, WAND.

DOI: 10.9781/ijimai.2019.04.002

I. Introduction

PRATICAL web search engines are very complex with the goal of returning fast and precise results. The result must be both effective and efficient. These search engines use techniques and algorithms of query processing, such as WAND algorithm, to return a set of ranked documents results named Top-k. These algorithms are executed on a data structure called inverted index [1]. Such structure gives for every term the set of documents in which it appears with additional information, like the term frequencies TF, the list of positions in every document, the format and the size in which it is written. Such construction generates a very large index. In fact, its size is larger than the set of original documents. As a consequence, traversing this index becomes the major bottleneck in query processing. In fact, it is not suitable, really not practical or impossible, to sweep all posting lists. An early termination algorithm is so recommended for such situation. It can return the exact Top-k without scanning the entire posting list. Note here, that a posting list is the part of the inverted index charged in the memory for treatment. We must note also here that the lists are ordered in an ascendant order according to documents numbers [2][3] or on descendant order on TF [4][5]. The choice is done according what the algorithm designer wants. In order to reduce the information representation in the posting lists, a set of compressing techniques have been proposed [6][7][8].

In information retrieval, two major and basic alternatives have been proposed for traversing the posting lists. It is about TAAT (Term-At-A-Time) and DAAT (Document-At-A-Time) strategies [9].

* Corresponding author.

E-mail addresses: imenezemani@yahoo.fr (Z. I. Mansouria), lougmiri@yahoo.fr (Z. Lougmiri)

In fact, SAAT(Score At A Time), GAAT(Graph At A Time) and RAAT(Rank At A Time) and JASS(SAAT) are additional strategies which are proposed for remedying the first strategies weaknesses [10] [11][12].

Since the first works in the field of information retrieval [13], the stopping condition is an interesting part of every early termination algorithm. It consists in ending the execution if k responses are computed even if there are more important results with ranks greater than k. A document is considered relevant if its score is greater or equal to a certain bound. The threshold algorithm TA of Ronald Fagin [14] is one of the most popular algorithms in the context of databases. As information systems are so large and as search engines must deal with large dataset, the WAND algorithm has became unavoidable. It has been used in a number of commercial search engines [15]. It has the ability to skip in an intelligent manner some documents and parts of posting lists according to a precise test, as we will see in next sections.

Really, the strongest weakness of early termination algorithms resides in the lack of precision in their responses. For Top-k and for a query q of length Lq, it is usual that WAND misses in Top-r, with $r \leq k$, some documents which share a high number of terms with the query. It is about this ascertainment that we built our solution. In this paper, we focus on early termination and we propose a new extension to algorithm WAND. Our aims are: a) To return all totally relevant documents that contain query terms. b) To reduce the operations number in query processing. c) To ameliorate the results quality by ameliorating the responses precision. In particular, we propose new fine metrics to measure the relevance degree of the returned documents. Compared to the naive approach that contains at least one of the query terms, our approach returns the relevant documents ranked first, without any loss in precision or recall or in new proposed metrics.

The reminder of the paper proceeds as follows: Section II gives a

representation of index structure, TAAT / DAAT strategies and early termination. The WAND approach is detailed in section III. In section IV, we quote our contributions and section V presents the details of our propositions. In section VI, we describe our experimental results. Finally, we conclude in section VII and discuss our future work.

II. Background

In this section, we provide a background on index structure, compression, early termination and index traversing strategies.

A. Index Structure

In order to evaluate queries in search engines, a well structure is constructed; it gives with precision all information about every term in the set of documents. This structure is called inverted index as it captures the list of documents which contain this term with other information. So, the term is used as a key access to such structure. Every line from this structure is called a posting list when it is charged in memory. This structure was largely presented and explained in literature. Works [7][16][17][18][1] have presented all what can be related to the construction and the use of such structure.

For a term t_i which appears, a record of its inverted list can have the next format:

d_j	TF_{ij}	$pos_1, pos_2, \ldots$

Where d_j is a unique document number, called also document identifier, in the set of documents collection, this number is assigned during the crawling process. TF_{ij} is the number of occurrences of t_i in d_j. The list $pos_1, pos_2,\ldots$ contains all positions of t_i in this document. It is useful essentially for searching phrases. The set of terms composes what is so-called the vocabulary or the dictionary, it can be so large, especially when different languages are considered. As the web and information systems are dynamic, the number of documents (documents, web pages,…) is growing perpetually, which induces that the size of posting lists grows exponentially also. As a result of such evolution systems behavior, the length of inverted lists becomes an important bottleneck [19].

In order to exploit efficiently the inverted lists, many organizations have been proposed. It is possible to present them briefly as follows [8]:

Document-Sorted Index: It is the first representation of the inverted index. In this situation, the inverted lists are sorted by document identifier [8][20][3].

Impact-Sorted Index: In this situation, the importance is given to the terms. Their scores will contribute to computing the similarity between a query q and the set of documents. The inverted lists are sorted by their impact; BM25 [21] is the most popular method which has implemented such idea. [22] and [10] have proposed and compared another method for retrieving information based on impact-sorted index.

Impact-layered index: This is a combination of the two precedent presentations. It computes the impact of every term according to a function similarity and each list is partitioned into a number of layers, such that the scores of one layer are higher than the next [8].

It is important to see that these organizations are built during the indexing process. Every organization is suitable for the solutions designers; it depends on what these ones want.

With the growing amounts of data on the internet, every time new documents (books, papers, tweets, Web pages, comments ...) are inserted. The preprocessing step will take a huge time for preparing indexes. The major problem will be finding a suitable manner for representing information. As the number of documents is too large, it will be essential to reduce the size of the inverted index. Here a compressing method will be helpful. Every document is designed by a unique integer. The most used method is to compute the gap between successive numbers of documents that contain a term *t*. For example, suppose $d_1, d_2, d_3, \ldots, d_k$ are the identifiers of the documents which contain the term *t*, the gap method is to store them in the inverted list as follows [6]:

$d_1; d_2\text{-}d_1; d_3\text{-}d_2; \ldots; d_k\text{-}d_{k-1}$

It has been shown in [24] that if the database can fit in the memory, so the accesses to it will be faster than the use of indexes. This result is so easy to see. Besides the inverted index contains all the vocabulary of the document collection, it contains more and more information, like TFs and positions of terms, so its size is larger than document collection's size. So, it will be impossible to fit it in the memory without compressing it. For this object, the first step is to compute the d_gaps, as shown before. The second step is to implement a well method for compressing it and getting information without decompressing. Sholer [23] has shown that for documents identifiers it is well to use the Elias [24] code when these numbers are too large; but for the positions, as they are little numbers, it is better to code them with the Gamma code.

In the same logic, a high number of methods of compression have been proposed in the literature. Everyone depends on the situation. It is possible to cite for classical methods: Elias encodings [24] and Golomb/Rice's encoding [25]. Newer methods are VByte [26], Simple [27], Interpolative [28], PForDelta [29]. Other techniques are proposed in [30][31]. Trotman [32] gives a comparison between different compressing methods. More material can be found in [7].

B. Early Termination

As we have said before, it is impossible to scan all posting lists for evaluating a query. Since first works, it has been questioned to stop execution if a certain condition is verified [13].The ranking operations usually apply Early Termination algorithms to avoid fully processing of documents. We say that the set of responses is not exhaustive; so many important responses can be avoided. There are three cases in early termination:

Stop early: In this case, the most promising documents are ranked first in posting lists and the execution will stop once k documents are obtained according to a well defined monotonic function. The algorithms like TA, FA, and NRA algorithms of Fagin [33][34][35] are examples of such type. They can be used for computing the most important objects, like Web pages or cars, in a database, but they cannot be used in information retrieval of context as the lists here cannot have the same format.

Skips: By defining a certain criteria, it will be possible to skip a set of documents if the criteria are not respected. WAND [20][11] are examples of such algorithms.

Score only partially: Algorithms evaluate documents by computing only approximate scores if their score is lower than a certain value [36].

Algorithms like WAND, BM25, BMW, BMI use early termination techniques avoiding processing complete lists [36][37][38][39].

C. Traversing Index Strategies

In the query processing domain, Turtle and Flood [9] classified evaluation into two main classes.

Term-At-A-Time (TAAT): The strategy traverses query terms term-by-term, while partial documents score are cumulated [40][10]. In this strategy, more information is stored in inverted lists.

Document-At-A-Time (DAAT): This strategy is very fast and processes all posting lists in a parallel manner, with respect to a single document before moving to the next one. DAAT is destined for big

collections processing. Examples: WAND [20], MaxScore [41].

In 2011, Fontoura et al. [42] compared TAAT and DAAT, to show efficiency gains in DAAT strategy [20]. Moreover, TAAT strategy happens to be more complex than that of the DAAT. At the opposite, [43] has shown that TAAT outperforms DAAT. We think the performance depends on how every strategy is implemented.

Recall that query processing use two query models: conjunctive (AND) and disjunctive (OR) queries. In general, conjunctive queries are more significant and disjunctive queries are more expensive than conjunctive queries.

It is important to report that additional methods have been proposed, even if they are deduced from the above strategies.

SAAT (Score At A Time): This strategy is considered when impact-based indexing is used. It searches all lists and processed the posting lists in the decreasing order of the impact values [10][11]. Hao [12] uses a different version of previous works for applying this technique for pseudo-relevance feedback. Zhang [44] made a revision of this technique and compares it to TAAT and DAAT. Joel [45] proposed a set of heuristics for studying efficiency and effectiveness of this technique. This method was implemented over JASS system [46]. This one is dedicated to evaluate queries according to the impact-sorted indexes. This method was compared to JAAS.

RAAT (Rank At A Time): This strategy was proposed by [47]. It is of type impact scores and it combines Boolean intersection queries. The lists are presented in a descending order. The head of every list is the term with the high score. Every list is considered as a query.

GAAT (Graph At A Time): This technique is not in topics here, but we want to cite it as a manner for searching information where information can be presented as a graph, such as RDF (Resource Description Framework), citation between co-authors in DBLP and protein interactions [48].

D. Works Around WAND

In this section, we will present the important works that have studied WAND; we have discarded works which have general aspects of information retrieval as they are not the subject of this paper. We want to show that WAND is really at the state of the art since it was published in 2003. It is cited as the most method which presents correctly the DAAT technique. It has the ability of skipping parts if, at a certain moment, the next documents scores cannot exceed an upper bound. [8] is an enhancement of WAND-style. Ding and Suel defined BMI method, also called BMW. It has augmented the inverted lists by well defined blocks. Every block is delimited by an upper bound. If the score of a document does not exceed, a skip is performed. So BMI will be faster than WAND. [3] defines new bloc indexes which on authors define a new hierarchical algorithm. This work has presented also a comparison between WAND and MaxScore [11]. Rojas [19] proposes a 2-steps method for parallelizing WAND. The objective is to reduce the inter-processor communication and running time cost. This work invokes a multithreading approach to exploit the multi-core parallelism. This same work was resumed in [15] by new formulas. The objective here is to reduce memory usage and computation cost based on WAND and BMI. Andrei Broder, the WAND's designer, proposes SWAND (for Sampling WAND) in [49]. SWAND aims to obtain performances in order to be inserted in any search engine. An object oriented approach is followed and other logic operators are inserted. The comparison between SWAND, IBM's Trevi and JURU search engines is done. SWAND has presented better performances.

In order to couple efficiency and effectiveness, [50] proposes a new method for dynamic pruning. Authors made an adjustment on the threshold and k of WAND in order to recover missed responses. This work has largely benefit from the safe-to-rank of WAND and from its manner of skipping lists. [51] proposes to save pairs of terms in posting lists in order to get more efficient WAND and MaxScore [41]. The new versions are called WANDP and MaxScoreP. [52] studied the effect of document identifier ordering on the dynamic pruning. Authors propose to apply random, document length and url ordering.

[53] examines multi-stage retrieval architecture. This one consists of a candidate generation stage, a feature extraction stage, and a reranking stage using machine-learned models. Authors studied the NDCG(Cumulative gain-based evaluation) metric according to WAND, BM25 and SvS [54]. Note that SvS (Small versus Small) is a method which searches the intersection between a set of ordered lists. This study has shown to be better than the two other methods. [55] has introduced a new bloom filter variation called Bloom filter chains on WAND for generation of a new method called Bloom WAND in order to retrieve tweets in real time.

[56] is an important work as authors work a lot on the aspects of language model. [56] has studied whether WAND is effective in this context. Experimentations have shown that it is not so helpful in this domain. Also, as authors of [57] work on selective search, they studied intensively whether WAND can be useful for selective search. They demonstrate than when indexes are well structured, WAND can be too effective. The effectiveness here is justified by the fact that indexes are grouped by subjects.

As BMW or (BMI) [8] proposes fixed sizes of blocks, [58] uses variable-sized blocks. Authors begun by partitioning blocks after what they define an algorithm for finding an approximate solution. This solution is named VBMW.

Bortnikov [59] defines conditional-skip iterator traversal strategy for pruning dynamically Top-k responses. This method can jump to a target document while skipping all matching documents preceding the target.

Daoud [60] presents a new system called WAVES which is multi-tier indexes for fast evaluation of queries. It is suitable to note here that this one is largely inspired by BMW and MMBW (Multi-tier BMW). WAVES uses BM25 for computing scores and stores upper bounds on blocks like BMW.

Based on WAND, MaxScore and BMW, authors of [61] have proposed a new method for computing pages's scores. The imperfection of BMW is that it does not support static scores. A static score is one given by a function like PageRank of Google.

Petri [62] gives a comparison between WAND, BMW as DAAT strategies with the system JASS which is SAAT strategy. Note here, that every year a set of works are published in order to compare DAAT, TAAT, SAAT and RAAT methods. Sometimes, conclusions between those works are contradictory.

Andrei Broder et al. in [63] have extended WAND by coupling it with the K-means clustering method. The cause is that this later does not scale with millions of documents. Thus this work gave the birth to WAND-k-means algorithm.

Recently in July 2018, by defining the safe-rank, Andrew Kane et al. [64] propose a new method called split-list WAND. Authors define an initial threshold and split lists on two layers according to scores. Authors shown that this proposal is better than WAND and BMW.

Also in 2018, [66] proposes a framework for predicting parameters during query-by-query evaluation. They demonstrate, by prediction, that they do not need experts' judgments. [66] cites WAND as a well understood algorithm for retrieving information in different situations.

III. The Two-Level Evaluation Process

This section describes the WAND mechanism.

As explained by Border in [20] and Oscar in [15]; WAND is a query processing method based on two levels, the first level is simple and the second level use a complex scoring to extract relevant documents. In the first level (named preliminary evaluation), WAND identifies candidate documents using approximate scoring, the advantage of the first level is that it is possible to skip evaluation of a number of documents. At the second level, the identified candidate documents are fully evaluated with precise metrics; and stored in top-k heap. The threshold is an important dynamic value used by WAND approach, varied during execution. The initial value of threshold depends on the type of query (disjunctive or conjunctive).In general, the initial value equals to zero as long as the top-k heap is not completely full, once the heap is full, threshold equals the minimum score in the top-k heaps.

A. First Level "Preliminary Evaluation"

Keeping in mind that WAND algorithm in the preliminary evaluation calculates an approximate score by summing UB (each term is associated with an Upper Bound):

$$UB(d, q) = \sum_{t \epsilon q \cap d} UB_t \quad (1)$$

$$UB_t = \alpha_t \max(w(t, d_1); w(t, d_2) \ldots) \quad (2)$$

a) α_t= tf * idf represents a function of the number of occurrences of term t in a document d, multiplied by the inverse document frequency.

b) W (t, d) represents a function of the term frequency of t in d divided by the document length |d|.

Every document with approximate score smaller than threshold will be skipped.

B. Second Level

In the second level, the candidate documents are evaluated with an exact score:

$$Score(d, q) = \sum_{t \epsilon q \cap d} w(t, d) \quad (3)$$

The top-k heap is a list of k documents results which is initially empty; an admission of a new document in top-k heap is done when the exact score of this document is greater than the threshold (threshold represents the minimum score in the heap). If the heap is full, the new document replaces the document associated with the minimum score. Documents with a score smaller than the minimum score in the heap will be skipped. Every document skipped will not be inserted in the heap results [20].

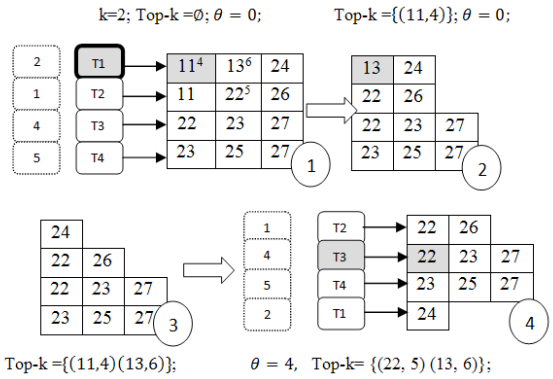


Fig. 1. A scenario of query processing "WAND" of a 4-termquery.

In Fig. 1 step (1), list of UB_t= (2, 1, 4, 5); top-k heap is empty; the initial threshold is 0 and the current pointers on the lists are (11, 11, 22, 23).Note that lists are sorted by their current documents. For (Current upper bound UB_{t1}=2, $UB_{t1} \geq \theta$) the pivot term is T1, and pivot document is 11 and score = 4≥ θ. Insert pivot document in top-k heap: Top-k = {(11, 4)}. As top-k is not completely full threshold equals to zero.

In step (2) list of UB_t= (2, 1, 4, 5), threshold equals 0 and the current pointers on the lists are (13, 22, 22, 23). For (Current upper bound UB_{t1}=2, $UB_{t1} \geq \theta$) the pivot term is T1; pivot document is 13 and score = 6≥ θ. Insert pivot document in top-k heap: Top-k = {(11, 4) ;(13, 6)}. As top-k is completely full, we update threshold θ = 4.

In step (4) lists are sorted by their current document (T2, T3, T4, T1) (22, 22, 23, 24). $UB_{t2}+UB_{t3}$=1+4=5 ≥ θ; the pivot term is T3, pivot document is 22 and score=5. Insert pivot document in top-k heap: Top-k = {(22, 5) ;(13, 6)}. Update threshold θ = 5.

IV. Our Contributions

As we have remarked, WAND is not so effective during queries' evaluation, as it can miss some important responses. We had the idea to propose a new method for recovering missed responses with the ability to propose an early termination algorithm. It is logic to see that when a user submits a query with length Lq, he will wait to responses with documents containing all, or a maximum of terms of this query; WAND can respond by a document containing few terms in first ranks, so WAND is missing effectiveness [65][66][50]. In theory of information retrieval, efficiency and effectiveness are major tradeoffs of search engines [50][67].Also it is important to see that effectiveness may have an impact on efficiency. Really all works that have extended WAND, cited in this paper, agree with us about the weakness of WAND in terms of effectiveness. In the next paragraph, we explain the cause of this imperfection.

During the steps of execution, and precisely when the threshold is equal to 0, an important time is spent as every document score is greater than this threshold; the insertions of these documents are implicit. So, it is unnecessary to test the relation between documents scores and the threshold as it will deal to a high number of fruitless rounds. This step is too hard and is the cause of the absence of effectiveness of WAND. More of this, it is possible to insert responses with high probabilistic scores but which share a part of terms with the query. Search engines always attempt to return responses in a semantic way such the responses contain a maximum terms in sharing with the queries.

In the next sections, we give in detail our proposal presented as follows:

(a) We propose modified algorithm based on WAND algorithm, which is composed of four levels of query processing;

(b) We propose to integrate a tree index;

(c) We maximize responses, which contain a maximal number of shared terms with the query;

(d) We evaluate our technique with classic metrics and with new fine proposed metrics named: fidelity, exact relevance degree, almost total relevance degree, and exact recall.

V. Our Approach

In this section we explain our algorithm and we give complete details of the four level evaluations of MWAND.

For large systems the full evaluation is an expensive task. The intention of our algorithm is to minimize the number of processing iterations and to select in first level the most relevant documents. In this paper we describe a novel extension to WAND and we employ DAAT strategy. MWAND is destined for both conjunctive and disjunctive queries that contain at least 3 terms.

Keeping in mind that query processing on WAND algorithm has two levels:

Level 1: Select a candidate document to be scored in the preliminary

evaluation.

Level 2: Select the Top-k document by exact scoring, if the score is larger than the current threshold.

Although the anticipation logic proposed by WAND has been very attractive and has inspired many researchers, there still are some palpable imperfections in the execution such as:

a) How to avoid the loss of relevant documents in the passage between operations?

b) The unnecessary approximate evaluation of several documents.

c) Are the returned documents, relevant documents?

d) Can we have a faster way, to speed up the search for terms in the index table?

A. Index Structure

Our aim is to resolve the time consuming problem, since the main task of search engines is by far the need to answer user queries within fractions of second, and since naively going through all the list of index terms can take hundreds of milliseconds.

More of this and as we have explained above, the vocabulary is too large, in order to reduce the space occupation, we choose to compute the longest common prefix between index terms [68]. This proposal reduces significantly the occupied space. According to each prefix the index is partitioned.

Our model integrates a tree index for IR systems into WAND and consists of a set of trees (Fig. 2). In which, every tree head represents an alphabet letter (a, b, c.....z) or a number (0, 1, 2.....9) or character (&.....#), and the descending nodes of the tree represent decompositions of terms in letter (Fig. 3). Each last letter term is associated with an inverted list. This index structure was chosen to optimize the search time and the indexation time.

Fig. 2. The set of index trees.

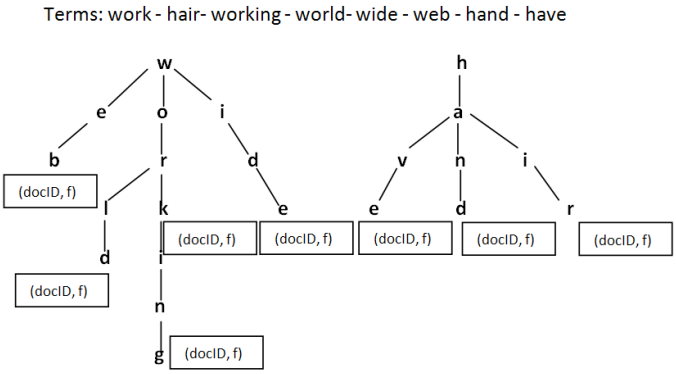


Fig. 3. Example of index tree.

Fig. 3 shows a scenario of tree index that is an example to index or to create the path of terms: t_1=work and t_2=world

*The first letter in t_1 is (w)

- Insert t_1 in tree(w);
- Create a node (o) in level m_1, and connect this node to tree head;
- Create a node (r) in level m_2, and connect this node to (o)m_1;
- Create a node (k) in level m_3, and connect this node to (r)m_2.

*The first letter in t_2 is (w)

- Insert t_2 in tree(w);
- Node (o) exist in level m_1;
- Node (r) exist in level m_2;
- Create a node (l) in level m_3, and connect this node to (r)m_2;
- Create a node (d) in level m_4, and connect this node to (l)m_3.

B. MWAND Algorithm

Our query processing algorithm is based on WAND and our scoring is based on formulas (1); (2) and (3).Our method is divided into 4 levels:

- First level "Intersection function";
- Second level "Filling function";
- Third level "Approximate evaluation of WAND";
- Fourth level "Exact evaluation of WAND".

1. First Level "Intersection Function"

The key idea of the first level is to avoid the loss of total relevant documents in the passage between operations of query processing. This intersection function takes as input parameter: The number of documents to be returned k and a query q of | q | terms (q > 2) using an index I which contains posting lists (It) sorted by DocID. The results parameters returned by intersection function are: a) The relevant top-k heap (top-k'). b) The number of documents returned k'. c) New posting lists destined to the second level. This function points on the first document in the first posting list and verifies its existence in the other lists; its role is to retrieve all the documents that appear in the | q | and | q-1 | postings. The top-k' heap returned represents the union of the set of documents that contain | q | terms ($D_{i|\,q\,|}$) and the set of documents that contain | q-1 | terms ($D_{j|\,q-1\,|}$),as explained below:

$$\text{Top-k'} = It_{|q|} \cap It_{|q|-1} \cap It_{|q|-2} \ldots\ldots \cap I_{t1}$$
$$= \{ D_{i|\,q\,|} \} \cup \{ D_{j|\,q-1\,|} \}$$

The objective of this function is to minimize the number of documents to be evaluated in the next levels.

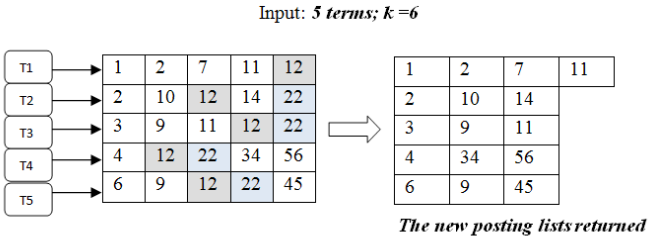


Fig. 4. Example of "Intersection function" and the results returned.

In Fig. 4 a scenario of intersection function is shown, the input parameters are a query of 5 terms (|q|=5) and k=6; parameters returned are:

- Top-k'=$\{D_{i|q|}\} \cup \{D_{j|\,q-1\,|}\}$=$\{12\} \cup \{22\}$=$\{12, 22\}$; Top-k' represents an union of sets of documents that contain 5 terms and 4 terms of query;
- k'= Card(Top-k')=2; k' represents the number of documents in Top-k';
- Intersection function returns new posting lists: I_{t1}=$\{1, 2, 7, 11\}$; I_{t2}=$\{2, 10, 14\}$; I_{t3}=$\{3, 9, 11\}$; I_{t4}=$\{4, 34, 56\}$; I_{t5}=$\{6, 9, 45\}$.

2. Second Level "Filling Function"

The basic WAND algorithm is proposed in [20], our MWAND algorithm improves it in the pivot term selecting. Critical of the computation of WAND approach is the definition of the top-k filling:

the initial value of threshold equals to zero as long as top-k heap is not completely full. As it is shown in Fig. 1: ∀ (k ∈ N); For k first iterations of query processing, WAND algorithm loses time passing through the two phases of approximate and exact evaluation. In this context, our approach adds a new level, that have to optimize the processing time, using a filling function to avoid the time lost in the approximate evaluation phase. This function takes as input parameters: a) the new posting lists returned by the first level b) the value of k' returned by the first level; and returns as result: a) initial top-k heap b) value of threshold c) new lists destined to WAND. In our function we eliminate the approximate evaluation and unnecessary tests; the sum of upper bound score is greater than the value of threshold that equals zero ($\forall$ *approximate score* $= UB\,(d,q) = \sum_{t \in q \cap d} UB_t \geq \theta \geq 0$) for k first iterations. Thus, our filling function of top-k heap selects rapidly a set of initial pivot documents. Fig. 5 shows an example on filling function, to select a set of pivot documents. In this example, all posting lists are sorted by their current DocID (1, 2, 3, 4, 6). We know that It_1 current DocID is (1), It_2 current DocID is (2), It_3 current DocID is (3), It_4 current DocID is (4) and It_5 current DocID is (6). The maximum current DocID is named "MaxCurDocID" and its value is equal to 6. In this scenario, we select all DocID smaller than maximum current DocID (candidate document = DocID \ if DocID ≤ MaxCurDocID). The set of documents selected is {It_1 (1, 2); It_2 (2); It_3 (3, 5); It_4 (4); It_5 (6)} = {1; 2; 3; 4; 5; 6}.This list is sorted in ascending order and the k" first documents are inserted in top-k heap and their exact scores can be calculated. Knowing that $k''=k-k'$. The initial *top-k"* heap selected by MWAND algorithm is similar to the initial top-k heap of WAND.

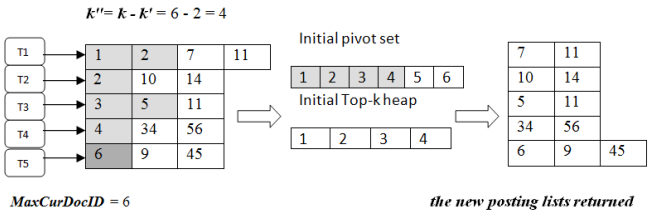


Fig. 5. Example of "Filling function" and the results returned.

In Fig. 5 filling function of MWAND sort lists by their current docIDs, and select all documents with docID ≤ MaxCurDocID. In this scenario initial k=6 and k'=2 then k"= k-k'=4.

- Initial pivot is {1, 2, 3, 4, 5, 6} ; select k first element, top-k heap is {1, 2, 3, 4};
- New posting lists returned are: It_1{7,11}; It_2{10, 14}; It_3{5, 11}; It_4{34, 56}; It_5{6, 9, 45}.

In cases where:

a) k" > n (n=Card(initial pivot set));we execute the filling function again, updating the value of k"=n - k", after we combine the lists.

b) k' equals to k; the execution stop in the first level and we proceed directly to the processing of the next query.

VI. Results

In this section, we will expose our experimental results. But, first we explain why we choose to compare MWAND only with WAND and not with other methods. It is logic to compare MWAND with those works, but we assume that MWAND and WAND work with the same logic. Works like BMW [8][50][51][52][58] or others in our related works, have extended WAND with supplementary structures; for example BMW propose to consult well defined blocks, besides it manipulates the inverted lists. Of course times spent by such proposition will be better than those consumed by WAND. The technique of augmenting solutions by supplementary structure is a common way in research. For example, in order to compute the Top-k in databases, besides it manipulates the same lists as TA (the Threshold Algorithm) of Fagin [34], BPA (the Best Positions Algorithm) of [69] proposes to use a supplementary list which saves the best positions of an object. Before the score of an object is computed, [69] consults the positions of an object. If it has good positions so it is in the Top-k. By this manner BPA will consume at the worst case the same time as TA. We can see the same remark between the algorithms TPUT [35] and HT (the Hybrid Threshold) [70]. This later defines more structures for computing the Top-k, despite it uses the same logic as TPUT. In our case, we want to show that with the same structure we can give a better solution. We do not use any extra-structure for our proposition.

A. Experimental Setup

Datasets. We have tested the WAND and the MWAND algorithms using Reuters collection-21578.

Query sets. We use a collection of 62350 queries which contains queries of lengths that vary between 3 and 10 terms and a list of 677 stop words (we remove stop words in the queries and in the inverted index).We returned top-20 results for each query.

Index structure. We compared the WAND and the MWAND approach on a tree index using DAAT query processing. Note that WAND and our approach by tree index return exactly the same results that WAND and our approach with classical index. In all our runs, we load index completely into the main memory. All codes are available by contacting the authors.

The experimentations were conducted on a machine with Intel (R) Core(TM) i7-4500U CPU @1,8Ghz with 8 GB of RAM. The programs were written in Java in Netbeans 7.0.

B. Results

In this section we compare our algorithm MWAND (using DAAT) and WAND on conjunctives and disjunctives queries. Then we measure the performance by six criteria: precision -- recall -- f-measure --Exact relevance degree of results --almost total relevance degree of results -- fidelity -- exact recall -- number of operations -- running time (in ms).

In the field of information science and systems of IR, the area of results quality is very important, that is why a set of tests must be applied and the question that may be asked is: should new measures be integrated in this domain to define the performance of systems?

1. Performance Evaluation

As explained Buttcher in [71] and Lewis in [72], the definition of precision and recall metrics is obtained by dividing the documents returned by SRI into two main categories: relevant documents and irrelevant documents, as it is shown in Fig. 6.

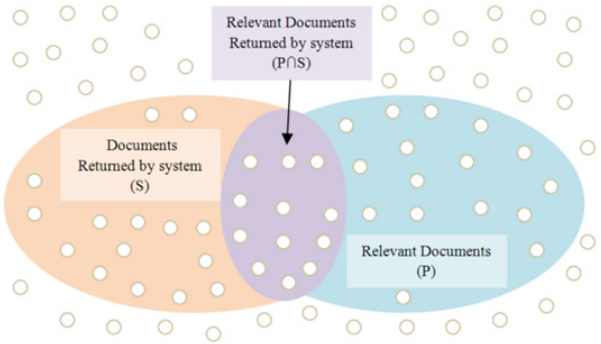


Fig. 6. Performance Evaluation.

At first, we compare WAND and MWAND algorithms on precision, recall and f-measure metrics:

$$\text{Precision} = \frac{|P \cap S|}{|S|} \quad (4)$$

$$\text{Recall} = \frac{|P \cap S|}{|P|} \quad (5)$$

$$\text{F-measure} = \frac{2*(precision*recall)}{precision+recall} \quad (6)$$

a) Comparison on the Precision

We have computed the precision of the results search. Table I gives the values of precision of 20 sets of results. In this experience, we can see that MWAND presents best values of precision. The average precision of: a) WAND is 0.56. b) MWAND is 0.72. The times when the precision is equal to 1 is explained by two facts. The fact one is about a non-open data, it is a closed data where there are new documents insertions, so it is possible to get such precision. The second fact is about the number of responses. When this one is so little it will be possible to get the unique solution. The times where the precision of MWAND is high mean that in the first round where the threshold is null, this algorithm inserts documents where everyone shares at least (Lq-1) terms with query q.

Fig. 7 shows the curve of the average of precision for 10 queries, while Fig. 8 presents the cumulated gain observed on Fig. 7. The accumulation can give more information. These figures illustrate how much MWAND returns better precision than WAND. This observation is in the continuation of what we have seen on Table I. In order to see about the precision if we increase the number of queries, we have executed 100 queries in the same conditions.

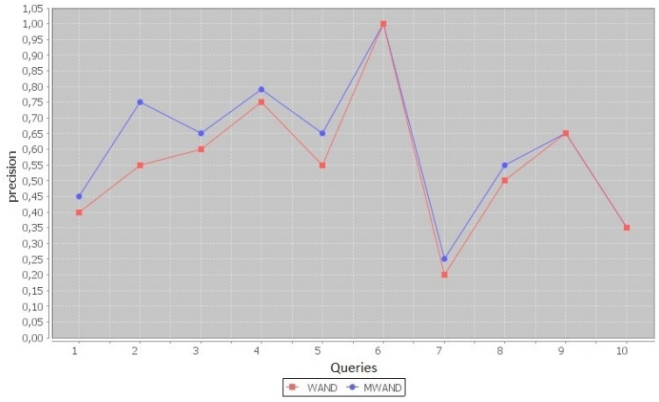


Fig. 7. Precision curve for 10 queries.

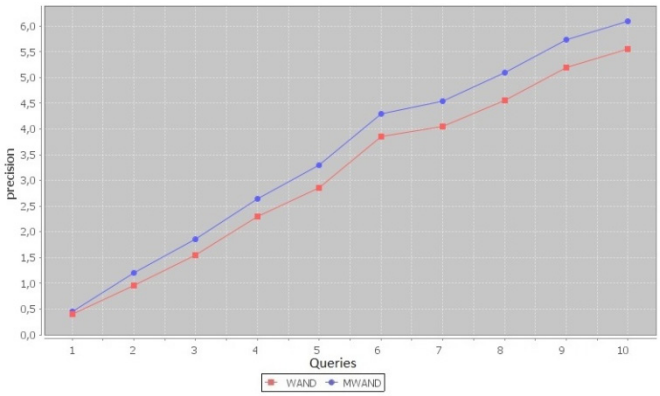


Fig. 8. Cumulated Precision curve for 10 queries.

Fig. 9 presents the curve of precision of these queries and Fig. 10 shows the cumulated precision gain. It is clear that responses in MWAND are more precise than those one returned by WAND. This result is due to the manner with what MWAND computes responses. In Fig. 11, we analyze the area of |P∩S| for results query of 5 terms. The initial value is 0.We test for each document (d) of top-k the existence of (d) in (p); when this condition is verified we increment the initial value. In this example, we see that all top-k documents returned by MWAND are relevant compared to WAND which loses some relevant documents.

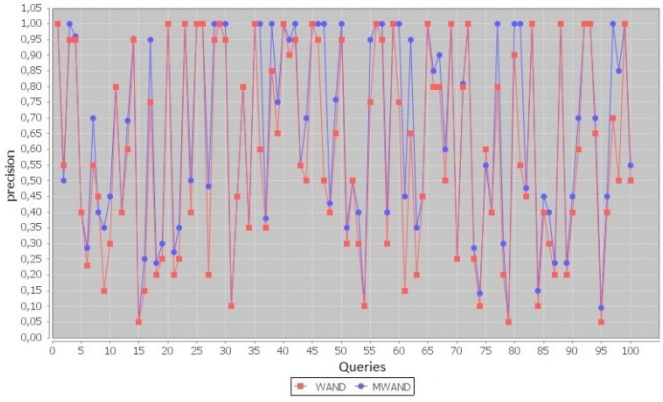


Fig. 9. Precision curve for 100 queries.

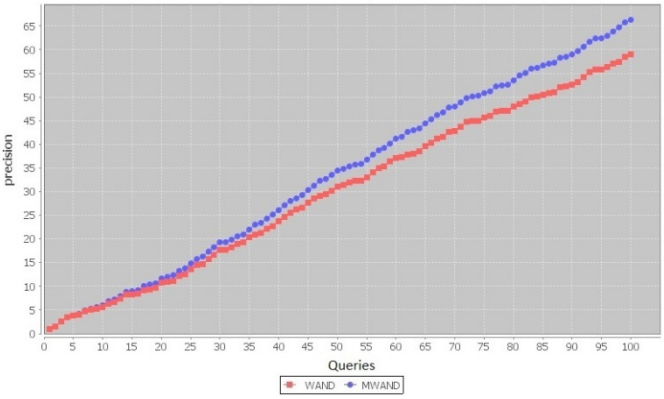


Fig. 10. Cumulated Precision curve for 100 queries.

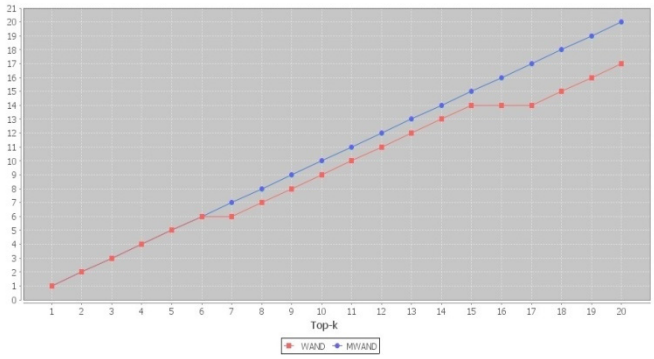


Fig. 11. |P∩S| curve for one query.

TABLE I. Comparison of WAND and MWAND Results in Terms of Precision

Queries	WAND	MWAND
1	0,7	1
2	0,3	0,3
3	0,35	1
4	0,45	1
5	0,4	0,45
6	1	1
7	0,35	0,38
8	0,45	0,8
9	0,8	1
10	0,45	0,476
11	0,01	0,15
12	1	1
13	0,5	0,5
14	0,1	0,15
15	1	1
16	0,55	0,65
17	0,85	1
18	0,05	0,5
19	1	1
20	1	1

b) Comparison of WAND and MWAND in Terms of Recall

In order to compare them on this measure, we have compared them according to the responses of the system. Thus, we have computed the recall for sets of results of 10 queries and 100 queries. Fig. 12 and Fig. 13 show clearly that MWAND has returned the best top-k documents, but WAND is not at the same level of quality of responses as in the precedent level. The missing of documents is the cause of the drop in the case of WAND.

c) Comparison According to the F-measure

In this experimentation, we compare these two methods according to the f-measure metric. As in the precedent experimentation, we have executed sets of 10 and 100 queries and we have calculated the f-measure of each top-k results. Results of this experimentation are depicted in Fig. 14 and Fig. 15; we found that MWAND has returned more relevant documents.

It is known that if the recall is high so the precision will be low. Here, if we plot the curves (precision-recall) of MWAND and WAND in the same plane, we will get MWANDS'curve higher than WAND's one.

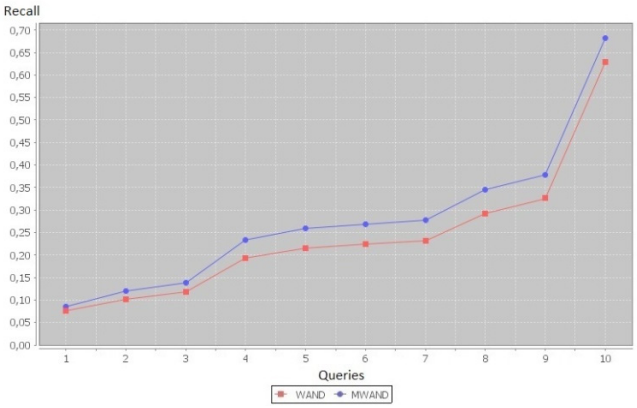


Fig. 12. Cumulated Recall curve for 10 queries.

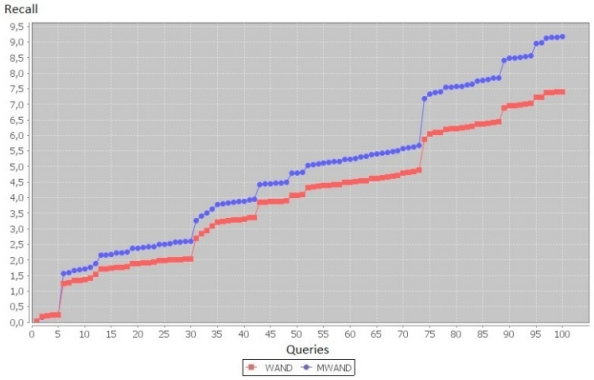


Fig. 13. Cumulated Recall curve for 100 queries.

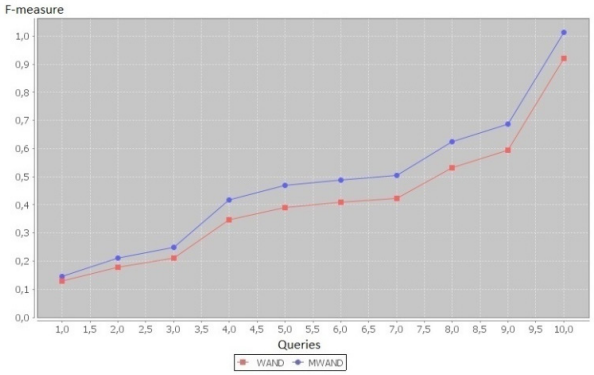


Fig. 14. Cumulated F-measure curve for 10 queries.

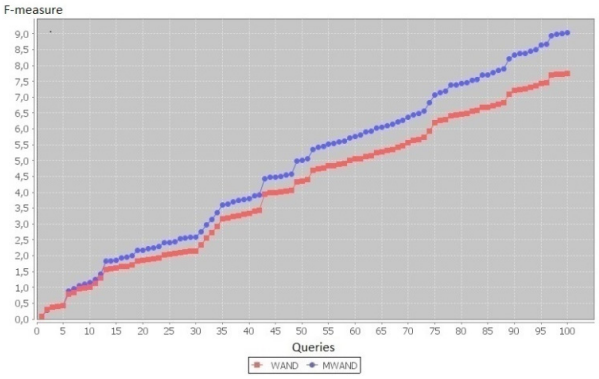


Fig. 15. Cumulated F-measure curve for 100 queries.

2. New Metrics

Before presenting our new metrics (for our knowledge, it is the first time that such definitions are given). We start with the definitions of relevance degree as Mechah explained in [73].

TABLE II. Different Definitions of Relevance For Document *d* and Query *Q*

Definition 1	Relevance of a Document	d ∩ q= ∅
Document d is relevant if it shares some terms with query q		
Definition 2	Total Relevance of a Document	d ∩ q=q
Document d is totally relevant if it contains all query terms		
Definition 3	Document Relevance Degree	d°(d) =|d ∩q |
Relevance degree is the number of terms shared between document d and query q		
Definition 4	Total Document Relevance Degree	td °(d) =|Q |
Relevance document degree is total if document contains all query terms		

The definitions cited in Table II give how much a document d is relevant for a query q. These definitions can give more information about the precision of the responses. Equations (4) and (5) given above are general, they cannot give a fine discrimination between algorithms. For this reason, we propose new metrics to find answers to the following questions:

- Are the returned documents so relevant to the queries?
- Are the total relevant documents positioned first?

In this context, we propose new definitions of results relevance degree and fidelity:

a) Exact Relevance Degree of Results

Definition 1. Exact Relevance Degree of Results: the exact relevance degree of result "r = {d_1, d_2…..d_l}" that contains a set of documents d, is noted Edr°(r). For a query q, Edr° (r) is the number of documents that contains |q| terms, divided by the total number of relevant documents P.

$$\text{For i in}\{1, 2\ldots\ldots l\}:$$

$$Xi = \begin{cases} 1 & \text{if } (td^\circ(d) = |q|) \\ 0 & \text{Otherwise} \end{cases}$$

$$|Edr^\circ(r)| = \sum_{i=1}^{l} X_i \quad (7)$$

$$Edr^\circ(r) = \frac{|Edr^\circ(r)|}{|p|} \quad (8)$$

A result is totally relevant if it contains K documents totally relevant.

b) Almost Total Relevance Degree of Results

Definition 2. Almost Total Relevance Degree of Results: the almost total relevance degree of result "r = {d1, d2…..dl}" that contains a set of documents d, is noted dr°(r). For a query q, dr°(r) is the number of documents that contains |q| terms or |q-1| terms, divided by the total number of relevant documents P.

$$\text{For i in } \{1, 2\ldots\ldots l\}:$$

$$Xi = \begin{cases} 1 & \text{if } (d^\circ(d) = |q| \text{ or } d^\circ(d) = |q-1|) \\ 0 & \text{Otherwise} \end{cases}$$

$$|dr^\circ(r)| = \sum_{i=1}^{l} X_i \quad (9)$$

$$dr^\circ(r) = \frac{|dr^\circ(r)|}{|p|} \quad (10)$$

A result is almost totally relevant if it contains k documents that share |q| or |q-1| terms.

Table III represents the study of comparison of exact relevance degree and almost total relevance degree of the first 20 (top-20) results.

Table III shows values of |Edr°(r)| and |dr°(r)| of 20 sets of results, every set contains 20 documents returned (top-20).We test results and we note that the results of MWAND in queries (1, 6, 9, 15, 17, 19, 20) are almost totally relevant, and results in queries (15, 19, 20) are totally relevant. Results of WAND in queries (15, 19) are almost totally relevant. We can see in Table IV the percentage of totally relevant results and almost totally relevant results of WAND compared to the MWAND.

TABLE III. COMPARISON OF SIMPLE AND EXACT RELEVANCE DEGREE

q	|Edr°(r)|		|dr°(r)|	
	WAND	MWAND	WAND	MWAND
1	1	7	4	20
2	0	0	0	0
3	0	1	0	17
4	1	1	1	15
5	1	1	2	3
6	1	1	5	20
7	1	2	1	2
8	1	1	2	2
9	0	0	4	20
10	1	2	1	2
11	1	1	1	2
12	0	0	0	3
13	4	4	4	4
14	2	2	2	3
15	9	20	20	20
16	0	1	0	3
17	4	6	5	20
18	1	1	1	1
19	9	20	20	20
20	8	20	16	20

TABLE IV. COMPARISON OF WAND AND MWAND IN TERMS OF RELEVANCE

	Percentage of total relevance		Percentage of almost total relevance	
WAND	0	0%	2/20	10%
MWAND	3/20	15%	7/20	35%

Note that, in all cases $|Edr^\circ(r)|_{MWAND} \geq |Edr^\circ(r)|_{WAND}$ and if |Edr°(r)|=K automatically |dr°(r)|=K; we can see this case in results queries evaluation of (15, 19, 20).There are two main observations: Firstly, in some cases relevant documents that contain all the query words are ignored by WAND. Secondly all the MWAND results are better than the WAND results.

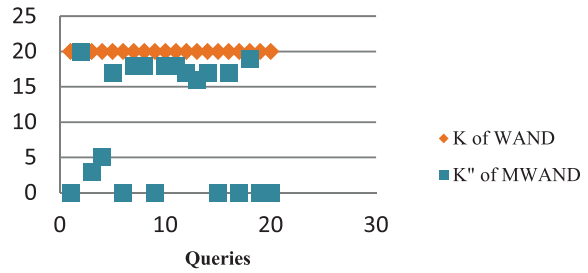


Fig. 16. K and K" variations.

As we explained in section V, K represents the number of documents to be returned, k' is the parameter returned by the first level and k"=k-k'. Fig. 16 shows the variation of K" during query processing. For WAND the value of k is static but for MWAND it varies between 0 and K (K=20). In MWAND we can find two cases: a) The best case is when k"=0, here MWAND stops processing in the first level and results are almost totally relevant or totally relevant. b) The worst case is when k"=20; in this case totally relevant documents do not appear in results set.

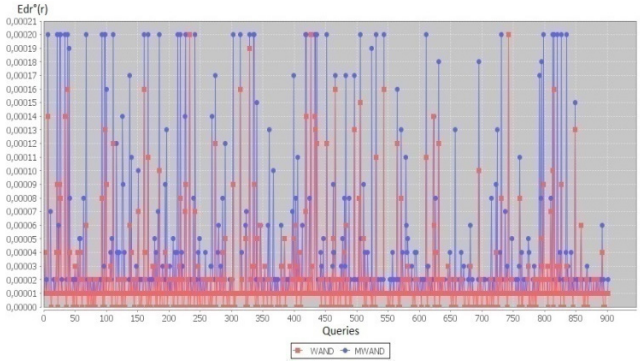


Fig. 17. Exact relevance degree of WAND and MWAND.

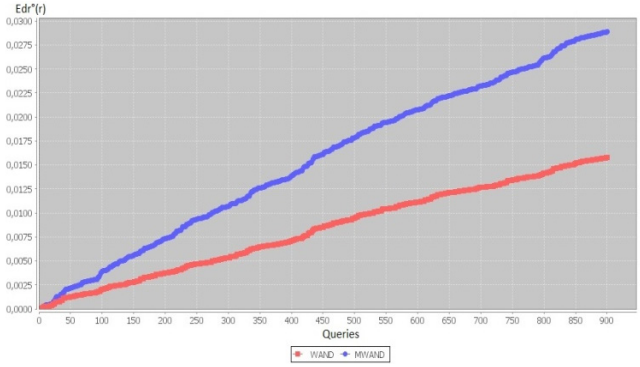


Fig. 18. Cumulated Exact relevance degree of WAND and MWAND.

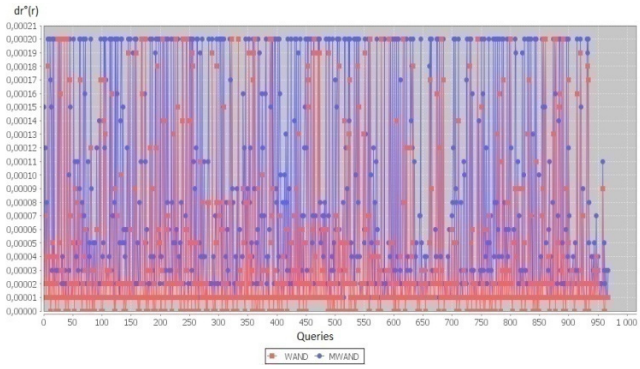


Fig. 19. Almost total relevance degree of WAND and MWAND.

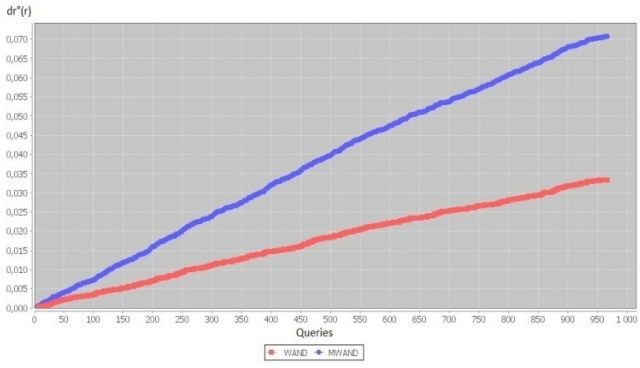


Fig. 20. Cumulated Almost total relevance degree of WAND and MWAND.

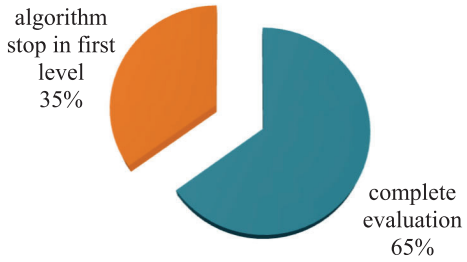


Fig. 21. Percentage of queries that stop processing in first level, of MWAND algorithm.

In Fig. 17 and Fig. 19 we represent exact relevance degree and almost total relevance degree of about 950 set of results and we see that MWAND results are more relevant than those of WAND. To clarify the results we show the Cumulated exact relevance degree and Cumulated almost total relevance degree in Fig. 18 and Fig. 20. We conclude that we can get more relevant documents using MWAND algorithm. We found that MWAND shares more documents with the real relevant documents. Fig. 21 shows that, in some cases MWAND can terminate earlier, where the value of k' returned in first level equals to K (see in Table III: queries (1, 6, 9, 15, 17, 19, 20) where d°(r)|=K=20)), the algorithm stops processing in first level and proceeds directly to the processing of the next query.

c) Exact Recall

The classical recall measures the ability of the system to retrieve all relevant documents responding to a query. We propose a new metric called Exact recall to highlight all set of results that contain total relevant documents.

Definition 3. Exact recall: This new measure noted ***Nrecall*** is defined as follows:

For query q of $|q|$ terms

$|q_i|$:The number of documents sharing i terms with the query q.

P: The number of real relevant documents.

For i{ 1, 2……|q| }:

$$\alpha_i = \frac{i}{|q|}$$

$$\text{Nrec} = \alpha_i \frac{|qi|}{|p|} + \alpha_{i-1} \frac{|qi-1|}{|p|} + \alpha_{i-2} \frac{|qi-2|}{|p|} \ldots\ldots + \alpha_1 \frac{|q1|}{|p|} \quad (11)$$

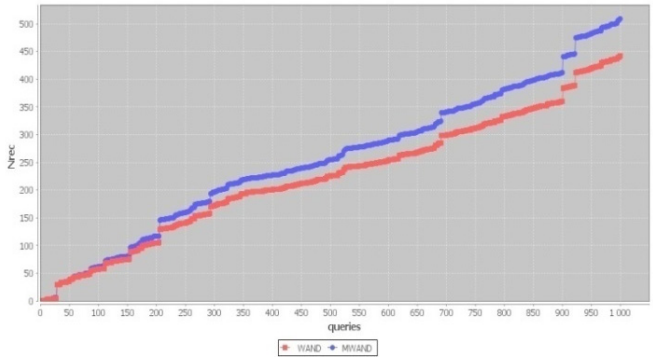


Fig. 22. Cumulated exact recall of 1000 queries.

In this stage of comparison, we executed a set of 1000 queries. So we tested their exact recall to evaluate the performance of algorithms, as shown on the Fig. 22. MWAND has presented better cumulated gain than WAND. This is due essentially to the manner of fitting responses in Top-k lists, where documents which share more terms with the query are inserted. For this we say that the strategy of MWAND is more efficient than that of WAND.

d) Fidelity

Definition 4. Fidelity: In this experience, we send a query q and we collect a set of documents, we also collect the positions of returned documents and we compared them with the positions of real relevant documents.

$$\text{Fid} = \frac{Doc_{fid}}{|S|} \quad (12)$$

Doc_{fid}: The number of documents returned by the system, having the same ranking position, as the relevant documents.

S: The number of documents returned by the system.

In this part we focus on ranking position. Our aim is to maximize fidelity value. We have executed sets of 500 and 6400 queries as shown in Fig. 23 and Fig. 24. The results of these experiences show the big performance of MWAND.

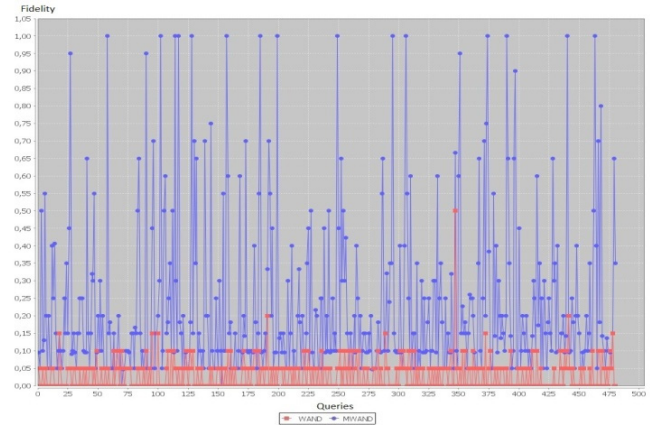


Fig. 23. Fidelity of 500 queries.

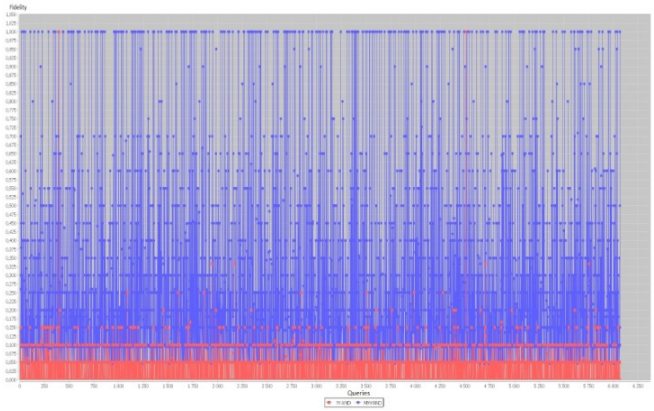


Fig. 24. Fidelity of 6400 queries.

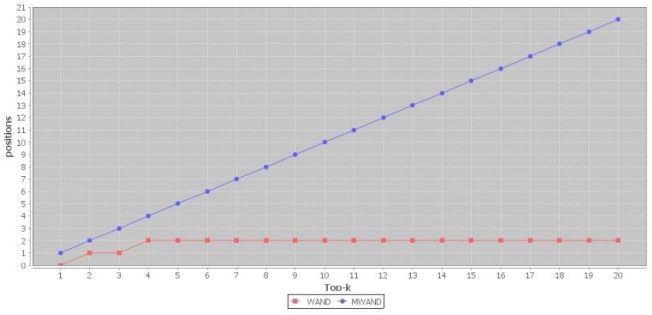


Fig. 25. Comparison of ranking Positions of WAND and MWAND.

We test if WAND and MWAND algorithms keep the same ranking position of the returned documents by comparing them with the ranking position of real relevant documents. In Fig. 25 we take an example and we find that MWAND keeps the same ranking for all documents and WAND keeps the same ranking for the second and fourth documents in top-k.

3. Comparison on Number of Operations

The goal of the first and second level in MWAND is to minimize the number of operations (NO) during query processing. Results of Fig. 26 show the variation of NO for a set of 20 queries. We observe that the MWAND processing is less expensive than WAND. We have also executed 1680 queries; we calculated the percentage of *NO* of WAND and MWAND. The results obtained show that MWAND percentage is 42.03% less than WAND percentage. That is because in MWAND, after executing the first and second level, the size of posting lists are smaller.

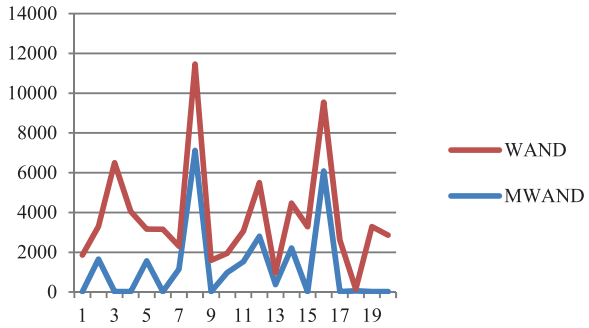


Fig. 26. Number of operations of WAND and MWAND.

4. Comparison on Time Processing

TABLE V. Comparison of Time Processing in (ms)

QUERIES	WAND	MWAND
1	234	209
10	561	499
20	1170	911
50	1966	1730
100	4070	4026
500	40131	19925
1000	54242	36438
6400	272551	271636
20052	759723	749879
39977	1775253	1768014

Table V shows the query processing time with different sets of queries. MWAND algorithm is faster than WAND algorithm. Especially, the performance of MWAND drops rapidly when: a) It stops in the first level. b) It selects a set of initial pivot terms in the second level, avoiding the approximate scoring.

VII. Conclusion and Future Works

Information retrieval is the science of searching information in a system; it calls a high number of techniques for satisfying users. This information can be a paper, a book, a piece of news, a photo or a state like in sentiment analysis or other thing, it depends on what we intend about the term “information”. The relevance is the greatest challenge, by which search engines are called to ensure for convincing users. In the context of retrieving documents, many techniques can be applied. The most important ones are term-at-a-time TAAT and document-at-a-time DAAT.

In this paper, we have presented techniques and structures that are used in retrieving documents in information systems. We have presented the strategies TAAT and DAAT, and we have described all what is known in this field. We have presented the structure of inverted lists and the major bottleneck of them. We have also presented their principal organizations in order to rapidly compute documents scores. The process of early termination is also presented. This process is applied for computing the exact solution without traversing the large inverted lists, as we have presented. In order to give new solution which is more effective and more efficient, we have presented our approach; we call it MWAND. It has the ability to avoid the unnecessary steps in WAND. We insert the best solution which shares the maximal number of terms with queries. By this way, we consume less time and we insert better documents. This one can give a better quality of responses.

We have compared WAND and MWAND according to general metrics, and in order to show how our work is efficient and effective; we have defined new metrics. For the best of our knowledge, it is the first time that such definitions are given. These metrics have the ability to analyze and show the hidden quality of responses. According to these metrics and by intensive experimentations, we have shown that MWAND can give better solutions in terms of quality and in execution time.

It is important to see that we have compared our work only to WAND as we do not use any other extra structure like BMW or others.

As future works, we want to do more comparisons with others works likes BMW. We will also insert the problem of intersection of ordered lists as we work intensively on it.

References

[1] J. Zobel, A. Moffat, "Inverted Files for Text Search Engines. ACM Computing Surveys," vol. 38, no. 2, article 6, July, 2006.

[2] J. Zobel, A. Moffat, R.S. Davis, "An Efficient Indexing Technique for Full-Text Database Systems," Proceedings of the 18th VLDB Conference Vancouver, British Columbia, Canada, 1992.

[3] C. Dimopoulos, S. Nepomnyachiy, T.Suel, "Optimizing Top-k Document Retrieval Strategies for Block-Max Indexes," WSDM'13, February 4–8, Rome, Italy, 2012.

[4] M. Persin, " Document filtering for fast ranking," In W. Croft and C. Van Rijsbergen, editors, Proc. ACM-SIGIR International Conference on Research and Development in Information Retrieval, Dublin, Ireland, 1994, pp 339-348.

[5] M. Persin, J. Zobel, R. Sacks-Davis, "Filtered document retrieval with frequency-sorted indexes," Journal of the American Society for Information Science, vol. 47, no. 10, pp 749-764, 1996.

[6] V. Anh, A. Moffat, "Index Compression using Fixed Binary Codewords," Proceedings of the Fifteenth Australasian Database Conference (ADC2004), Dunedin, New Zealand, 2004.

[7] A. Witten, A. Moffat, T. Bell, "Managing Gigabytes: Compressing and Indexing Documents and Images," Morgan Kaufmann Publishers, Los Altos, CA 94022, USA, second edition, 1999.

[8] S. Ding, T. Suel, "Faster Top-k Document Retrieval Using Block-Max Indexes,"SIGIR'11, 2011.

[9] H. Turtle, J. Flood, "Query evaluation: Strategies and optimizations," Information Processing and Management, vol. 31, no. 6, pp. 831–850, 1995.

[10] V. Anh, A. Moffat, "Pruned query evaluation using pre-computed impacts," In Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2006, pp. 372–379.

[11] T. Strohman, W. Bruce Croft, "Efficient Document Retrieval in Main Memory," SIGIR'07, July 23–27, Amsterdam, The Netherlands, 2007.

[12] H. Wu, H. Fang, "An Incremental Approach to Efficient Pseudo-Relevance Feedback," SIGIR'13, Dublin, Ireland, July 28–August 1, 2013.

[13] C. Buckley, A.F. Lewit, "Optimization of Inverted Searches," SIGIR '85 Proceedings of the 8th annual international ACM SIGIR conference on Research and development in information retrieval, 1985, pp. 97-110.

[14] R. Fagin, A. Lotem, M. Naor, "Optimal aggregation algorithms for middleware," In Symposium on Principles of Database Systems, 2001.

[15] O. Rojas, V. Gil-Costa, M. Marin, "Distributing efficiently the Block-Max WAND algorithm," In proceedings of the conference on computational Science, 2005.

[16] A. Moffat, J. Zobel, "Self-indexing inverted files for fast text retrieval," ACM Transactions on Information Systems, vol. 14, no. 4, October 1996, pp. 349 –379.

[17] R. Baeza-Yates, B. Ribeiro-neto, D. Mills, O. AmsterdamBonn, "Modern Information Retrieval," Addison Wesley, First edition, 1999.

[18] C.D. Manning, P. Raghavan, H. Schütze, "An Introduction to Information Retrieval," Cambridge University Press Cambridge, England, 2009.

[19] O. Rojas, V. Gil-Costa, M. Marin. "Efficient Parallel Block-Max WAND Algorithm," Euro-Par 2013 Parallel Processing - 19th International Conference, Aachen, Germany, 2013, pp. 26-30. Proceedings. Lecture Notes in Computer Science 8097, Springer 2013.

[20] A.Z. Border, M. Herscovici, A. Soffer, J. Zien, "Efficient query evaluation using a two-level retrieval process,"CKIM'3, 2003.

[21] E. Robertson, S. Walker, M. Beaulieu "Okapi at TREC–7: automatic ad hoc, filtering, VLC and filtering tracks," In Proceedings of the Seventh Text REtrieval Conference (TREC-7), NIST Special Publication 500-242, July 1999, pp. 253–264.

[22] V. Anh, A. Moffat, "Simplified Similarity Scoring Using Term Ranks," SIGIR'05, August 15–19, Salvador, Brazil, 2005.

[23] F. Scholer, H. Williams, J. Yiannis, J. Zobel," Compression of inverted indexes for fast query evaluation," In Procedure of the 25th Annual SIGIR, Tampere, Finland, pp. 222-229, 2002.

[24] P. Elias, "Universal codeword sets and representations of the integers," IEEE transactions on Information Theory, vol. 21, no. 2, pp. 194-203, 1975.

[25] S.W. Golomb, "Run-length encoding," IEEE Transactions on Information Theory, vol. 12, no. 3 ,pp. 399–401, 1966.

[26] H.E. Williams, J. Zobel, "Compressing integers for fast file access," Comput. J, vol. 42, no. 3, pp. 193–201, 1999.

[27] V.N. Anh, A. Moffat, "Inverted index compression using word-aligned binary codes," Inf. Retr, vol. 8, no. 1, pp. 151–166, 2005.

[28] A. Moffat, L. Stuiver, "Binary interpolative coding for effective index compression," Inf. Retr, vol. 3, no. 1, pp. 25–47, 2000.

[29] M. Zukowski, S. H´eman, N. Nes, P.A. Boncz, "Super-scalar ram-cpu cache compression," in: ICDE, pp. 59, 2006.

[30] H. Yan, S. Ding, T. Suel, "Inverted index compression and query processing with optimized document ordering," in: WWW, pp.401–410, 2009.

[31] H. Yan, S. Ding, T. Suel, "Compressing term positions in web indexes," in: SIGIR, pp. 147–154, 2009.

[32] A. Trotman,"Compressing Inverted Files. Information Retrieval," 6, 5–19. Kluwer Academic Publishers. Manufactured in The Netherlands, 2003.

[33] R. Fagin, "Combining fuzzy information: an overview," SIGMOD Record, 31, 2002.

[34] H. Bast, D. Majumdar, R. Schenkel, M. Theobald, G. Weikum. "IO-Top-K: Index-accessoptimized top-k query processing," In Proceedings of the 32th International Conference on Very Large Data Bases, 2006.

[35] P. Cao, Z. Wang, "Efficient Top-k Query Calculation in Distributed Networks.In Proceedings of the 23rd PODC. 2004.

[36] HW. Wong, D. Lee, "Implementations of partial document ranking using inverted files," Information Processing and Management, vol. 29, no. 5, pp. 647–669, 1993.

[37] R. Blanco, A. Barreiro, "Probabilistic static pruning of inverted files," ACM Trans. Inf. Syst, vol. 28, no. 1, 2010.

[38] A. Soffer, D. Carmel, D. Cohen, R. Fagin, E. Farchi, M. Herscovici, Y.S. Maarek. "Static index pruning for information retrieval systems," in: SIGIR, 2001, pp. 43–50.

[39] A. Ntoulas, J. Cho, "Pruning policies for two-tiered inverted index with correctness guarantee," in: SIGIR, pp. 191–198, 2007.

[40] C. Buckley, A.F. Lewit, " Optimization of inverted vector searche,"In Proceedings of the 8th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1985.

[41] T. Strohman, H. Turtle, W.B. Croft, " Optimization strategies for complex queries," In Proc. SIGIR, pp. 219–225, 2005.

[42] M. Fontoura, V. Josifovski, J. Liu, S. Venkatesan, X. Zhu, J. Zien, "Evaluation strategies for top-k queries over memory-resident inverted

indexes," Proc. VLDB Endowment, vol. 4, no. pp. 12:1213–1224,2011.
[43] M. Kaszkiel and J. Zobel. "Term-ordered query evaluation versus document-ordered query evaluation for large document databases," In Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Melbourne, Austrailia, August 1998, pp 343–344.
[44] F. Zhang, S. Shi, H. Yan, J. Wen, "Revisiting Globally Sorted Indexes for Efficient DocumentRetrieval," WSDM'10, February 4-6, New York City, New York, USA, 2010.
[45] J. Mackenzie, F. Scholer, J. Shane, "Early Termination Heuristics for Score-at-a-Time Index Traversal," ADCS 2017, December 7–8, 2017, Brisbane, QLD, Australia.
[46] J. Lin, A. Trotman, "Anytime Ranking for Impact-Ordered Indexes," ICTIR'15, September 27–30, Northampton, MA, USA.
[47] A. Elbagoury, M. Crane, J. Lin, "Rank-at-a-Time Query Processing," ICTIR '16, September 12 - 16, Newark, DE, USA.
[48] H. He, A.K. Singh, "Graphs-at-a-time: Query Language and Access Methods for Graph Databases," SIGMOD'08, June 9–12, Vancouver, BC, Canada.
[49] A. Anagnostopoulos, A.Z. Broder, D. Carmel, "Sampling Search-Engine Results," WWW 2005, May 10-14, 2005, Chiba, Japan.
[50] N. Tonellotto, C. Macdonald, I. Ounis, "Efficient and Effective Retrieval using Selective Pruning," WSDM'13, February 4–8, 2013, Rome, Italy.
[51] N. Tonellotto, C. Macdonald, I. Ounis, "Efficient Dynamic Pruning with Proximity Support," LSDS-IR 2010, July 23, 2010, Geneva, Switzerland.
[52] N. Tonellotto, C. Macdonald, I. Ounis, "Effect of Different Docid Orderings on Dynamic Pruning Retrieval Strategies," SIGIR'11, July 24–28, 2011, Beijing, China.
[53] N. Asadi, J. Lin, "Effectiveness/Efficiency Tradeoffs for Candidate Generation in Multi-Stage Retrieval Architectures,"IGIR'13, July 28–August 1, 2013, Dublin, Ireland.
[54] J.S. Culpepper, A. Moffat, "Efficient set intersection for inverted indexing," TOIS, vol. 29, no. 1, 2010.
[55] N. Asadi, J. Lin, "Fast Candidate Generation for Real-Time Tweet Search with Bloom Filter Chains," ACM Transactions on Information Systems, vol. 31, no. 3, article 13, July 2013.
[56] M. Petri, J.S.Culpepper, A.Moffat, "Exploring the Magic of WAND," ADCS '13, December 05 - 06 2013, Brisbane, QLD, Australia.
[57] Yubin Kim1, Jamie Callan, J.Shane Culpepper, Alistair Moffat, "Does Selective Search Benefit from WAND Optimization," ECIR 2016, LNCS 9626, pp. 145–158.
[58] A. Mallia, G.O. aviano, E. Porciani, N. Tonello, R. Venturini, "Faster BlockMax WAND with Variable-sized Blocks," SIGIR'17, Shinjuku, Tokyo, Japan, 2017.
[59] E. Bortnikov, D. Carmel, G. Golan-Gueta, "Top-k Query Processing with Conditional Skips," WWW 2017 Companion, April 3–7, 2017, Perth, Australia.
[60] C.M. Daoud, E. Silva de Moura, D. Fernandes, A. Soares da Silva, C. Rossi, A. Carvalho, "Waves: a fast multi-tier top-k query processing algorithm," Information Retrieval Journal, vol.3, n. 20, march 2017.
[61] D. Shan, S. Ding, J. He, H. Yan, X. Li. "Optimized Top-K Processing with Global Page Scores on Block-Max Indexes," WSDM'12, February 8-12, + Seattle, Washington, USA, 2012.
[62] M. Petri, J.S. Culpepper, A. Moffat, "Exploring the Magic of WAND," ADCS '13, December 05 - 06, 2013, Brisbane, QLD, Australia.
[63] A. Broder, L. Garcia-Pueyo, V. Josifovski, "Scalable K-Means by Ranked Retrieva,"WSDM'14, February 24–28, 2014, New York, New York, USA, 2012.
[64] A. Kane, F.W. Tompa, "Split-Lists and Initial Thresholds for WAND-based Search," SIGIR '18, July 8–12, 2018, Ann Arbor, MI, USA, 2018.
[65] M. Crane, J.S. Culpepper, J. Lin, J. Mackenzie, A. Trotman, "A Comparison of Document-at-a-Time and Score-at-a-Time Query Evaluation," WSDM 2017, February 06 - 10, 2017, Cambridge, United Kingdom.
[66] J. Mackenzie, J.S. Culpepper, R. Blanco, M. Crane, C.L.A. Clarke, J. Lin, "Query Driven Algorithm Selection in Early Stage Retrieval,"WSDM 2018, February 5–9, 2018, Marina Del Rey, CA, USA.
[67] L. Wang, J. Lin, D. Metzler, " A cascade ranking model for efficient ranked retrieval," In Proc. of SIGIR, 2011.
[68] D. Gusfield,"An increment-by-one approach to su x arrays and trees," Technical Report CSE-90-39 , UC Davis, Dept, Computer Science, 1990.
[69] R. Akbarinia, E. Pacitti, P. Valduriez. "Best position algorithms for top-k queries," In Proceedings of the 33rd international conference on Very large data bases, VLDB '07, 2007, pp. 495–506.
[70] H. Yu, H. Li, P. Wu, D. Agrawal, A. Abbadi, "Efficient Processing of Distributed Top-k Queries," In Proceedings of the 16th DEXA, 2005.
[71] S. Buttcher, Clarke, G.V. Cormack, "Information Retrieval: Implenting and Evaluating Search Engines," MIT Press, 2010, pp. 68.
[72] D. Lewis, Y. Yang , G. Rose, Li ,"RCVI/ A new benchmark collection for text categorization," Journal of Machine Research, 2004, pp.361-397.
[73] K. Mechach, L. Zekri, M.K. Abdi, "Collection and Selection Based Relevant Degrees Of Documents," In Journal of Digital Information Management (JDIM) , vol. 13, no. 2, pp. 110-119, 2015.

Zemani Imene Mansouria

Zemani Imene Mansouria is a PhD student at the department of computer sciences at Oran1 Ahmed Ben Bella University. She obtained her master in the field of operational research. Actually, she is interested in information retrieval, indexation and big data.

Zekri Lougmiri

Zekri Lougmiri is assistant professor at the department of computer sciences at Oran1 Ahmed Ben Bella University. He received his Phd at the same university. He works on multi-criteria optimization, information retrieval and Big data.

Senouci Mohamed

Senouci Mohamed is professor at the department of computer sciences at Oran1 Ahmed Ben Bella University. His current research area includes pattern recognition, deep learning and ad hoc networks.

Automatic Irony Detection using Feature Fusion and Ensemble Classifier

H. M. Keerthi Kumar[1], B. S. Harish[2] *

[1] JSSRF, JSS TI Campus, Mysuru, Karnataka (India)
[2] Department of Information Science and Engineering, JSS Science and Technology University, Mysuru, Karnataka (India)

Received 15 March 2019 | Accepted 12 June 2019 | Published 2 July 2019

Abstract

With the advent of micro-blogging sites, users are pioneer in expressing their sentiments and emotions on global issues through text. Automatic detection and classification of sentiments like sarcastic or ironic content in micro-blogging reviews is a challenging task. It requires a system that manages some kind of knowledge to interpret the sentiment expressed in text. The available approaches are quite limited in their capabilities and scope to detect ironic utterances present in the text. In this regards, the paper propose feature fusion to provide knowledge to the system by alternative sets of features obtained using linguistic and content based text features. The proposed work extracts five sets of linguistic features and fuses with features selected using two stages of a feature selection method. In order to demonstrate the effectiveness of the proposed method, we conduct extensive experimentation by selecting different feature subsets. The performances of the proposed method are evaluated using Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), Decision Tree (DT) and ensemble classifiers. The experimental result shows the proposed approach significantly out-performs the conventional methods.

Keywords

Classification, Feature Selection, Feature Fusion, Ensemble Method, Irony, K-means Clustering, Sentiment Analysis.

DOI: 10.9781/ijimai.2019.07.002

I. Introduction

IN the era of internet, micro-blogging sites like twitter, facebook, and review forums allow users to express their sentiment or opinion [56]. The sentiment or opinion in micro-blogs may relate to product, event or political debate in the form of text, image and video clips, where text plays an essential role in expressing opinions. Micro-blogging textual information is rich, progressively expanding in large volume of data, with a variety of information ranging from product to political events [1]. This textual information plays a vital role in determining sentiment of the population. An enormous amount of textual information provides valuable insight to governments, business organizations and individual decision makers [2]. The manual summarization of micro-blog textual information is time consuming. Hence, the automatic summarization of subjective information is very essential to determine polarity of the population [3]. The automatic text polarity identification process is known as Sentiment Analysis (SA) or Opinion Mining.

Sentiment Analysis (SA) aims to classify a given text into positive, negative or neutral polarity [4] [53]. There are many challenges related to SA which need to be addressed and resolve. Some of the challenges are: (a) language utterances, (b) punctuation marks to express sentiments, (c) shorten form of words (mainly in micro-blogs), (d) sarcasm/irony present in text snippet and many more. Sarcasm/irony detection in text is one of the major challenges in sentiment analysis.

* Corresponding author.
E-mail addresses: hmkeerthikumar@gmail.com (H. M. Keerthi Kumar), bsharish@jssstuniv.in (B. S. Harish).

The sarcasm has been studied by multidisciplinary endeavors such as sociology [5], psychologists [6], linguists [7] and computer scientists [8] for different types of text: twitter tweets, product reviews, internet dialogs, etc. [19]. Over the time, human have developed the ability to recognize sarcastic/ironic intent in utterances from childhood through social interaction [10]. Sarcasm is portrayed as ironic, intended to insult, mock or amuse. However, irony or sarcasm is a complicated mode of communication, which is informally connected with the expression of feelings, attitudes and emotions [11]. Sarcasm is closely related to irony [16]. Irony shifts the polarity of an apparently positive/negative utterances into its opposite [12]. Human intervention to recognize irony is extremely studious and time consuming. Due to this, researchers aim to develop an automatic system to recognize the ironic utterances present in the text.

Understanding ironic utterances from stance of both semantic and grammatical is another practice of Nature Language Processing [13]. Irony detection techniques are roughly categorized into machine learning and lexicon based approaches [14]. A lexicon based approach uses dictionary/corpus using statistical and semantic features to detect ironic utterances in a given text. On the other hand, a machine learning approach uses text features to classify ironic utterances using supervised/unsupervised techniques based on label or unlabeled text. Both approaches perform well in detecting ironic utterances present in sentences [47].

In irony detection, feature extraction and selection plays a vital role in determining the ironic utterances present in sentences. The features are extracted based on linguistic and content based approaches. A linguistic approach is an extremely broad phrase to extract textual features such as lexical, hyperbole and pragmatic features. The lexical

approach uses text properties such as unigram, bigram, n-grams, etc. for detecting irony in text [42]. In a lexical approach, the dictionary or corpus related to vocabularies of words are used to identify irony present in text snippet [15]. Similarly, hyperbole is another key feature often used in irony detection from textual data. A hyperbolic text contains interjection (wow, aha, etc.), punctuation marks (question marks and exclamation mark), quotes (‘ ’, “ ”) and intensifiers (noun, adverbs, adjectives) to detect irony in tweets [43]. The pragmatic feature includes symbolic or figurative texts such as emoticons of happy, sad, laughing, and crying etc., expressed in the sentences [15]. Researchers [17] [47] [51] [52] used various linguistic features to detect ironic utterances in short texts. However, identifying appropriate patterns to detect ironic utterances remains an open challenge.

In addition to the wide range of linguistic features, many researchers [9] [22] [46] [55] studied the content based approach i.e., presence or absence of term/features in reviews. In the content based approach, the number of features plays a vital role in accurate classification. The high dimensionality and sparsity is one of the major challenges faced during classification task. To curse the dimensionality, many researchers reported in [48] [49] [50] used feature selection methods to select discriminative features from a high dimensional feature space. The conventional feature selection methods such as Chi-square (χ^2), Information Gain (IG), Mutual Information (MI) are used to select discriminative features from high dimensional feature spaces. However, the selected feature subset may have features which convey similar information [26]. On this line, we propose a two stage feature subset selection using conventional feature selection methods and a clustering method to select the most discriminative features from a high dimensional feature space. On the other hand, linguistic features are extracted to detect sarcastic utterances in short text. The five groups of linguistic features that are extracted viz: Rating Feature, Word Feature, Acronym Feature, Symbol Feature and Emoticon Feature. Further, features are fused to capture various dimensions of characteristics of review. The fused features are classified using various classifiers such as Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF) and Decision Tree (DT). To enhance the performance of the classifiers, we ensemble Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF) and Decision Tree (DT) using weighted majority voting schema.

The main contribution of the paper is as follows:

- Five types of features are extracted using linguistic approach viz: Rating, Word, Acronym, Symbol and Emoticon Features.
- Two stage feature subset selection to select the most discriminative features.
- Features are fused (linguistic and content based feature subsets) to capture various dimensions of characteristics of review.
- Use of weighted majority voting schema to ensemble decision of each classifier.

The remainder of the paper is structured as follows: Section II depicts related works on irony detection. Section III presents the methodology of the proposed work. Experimentation and related results along with discussion are presented in Section IV. Finally, the work is concluded with future research directions in Section V.

II. Related Work

In recent decades, prominent research works are carried out by various researches [17]-[21] [27] for automatic detection of irony in various micro-blogs such as twitter, product reviews and movie reviews. A brief survey on automatic sarcasm detection by Joshi et al. [16] described various datasets, approaches, trends and issues in sarcasm detection. Some of the related works in literature are reviewed based on supervised, semi-unsupervised and rule based approaches. Similarly, in [13] Wicana et al., described sarcasm detection from the machine-leaning perspective. The research tried to explore supervised, unsupervised, rule based approaches and hybrid approaches to process data. Dave and Desai in [14], examine various lexicons based and machine leaning techniques for sarcasm detection on textual data. The comprehensive survey highlights the use of hybrid techniques, i.e. usage of both lexicons based and machine leaning techniques together for sarcasm classification.

Ravi and Ravi in [17], proposed a framework to automatically detect satire, sarcasm and irony found in news and customer reviews. The framework extracts features based on linguistic, semantic, psychological and unigram features. The various feature selection techniques are used to select the relevant feature subset from unigram features. The extracted and selected features subsets are fused and classified using Support Vector Machine (SVM) with various kernels, Logistic Regression (LR), Random Forest (RF), Naive Bayes (NB), Multilayer perceptron (MLP), etc. Similarly, Buschmeier et al. [18], described impact of features in a classification approach to detect irony in product review [20]. The method uses 29 special features such as positive/negative imbalance of reviews, hyperbole, positive/negative word with punctuation, quotes etc., along with bag-of-word features (21,773 features). The various features set comparison are drawn on different classifiers such as Linear SVM, LR, Decision Tree (DT), RF, and NB classifiers. Filatova in [19], identifies the sentiment shifts in sarcasm product review dataset [20]. The method demonstrated sentiment flow shifts (from negative to positive and likewise) using bi-gram feature along with 8 classification features (very negative-positive, very negative-very positive, negative-positive, negative-very positive and likewise). Justo et al., [21] proposed to detect sarcasm and nastiness in the social web. The various features such as mechanical turk, statistical cues, linguistic information, semantic information using Linguistic Inquiry Word Count (LIWC) and n-gram distribution of Part-of-Speech (POS) taggers are used to extract features. The feature subset is selected using the chi-square (χ^2) feature selection method. The binary classification was performed using rule-based and NB classifiers.

In literature, many researchers proposed various feature extraction techniques to classify ironic review. The existing methods uses NLP and machine learning approaches to extract various patterns to classify ironic content in reviews. However, some of the observations made from literature are as follows: (a) the number of features was too large, (b) feature extraction using POS tagger, and (c) searching each word in sentiment dictionary is clumsy. Hence, the proposed research developed a new approach to address these issues such as (a) Feature extraction (use of linguistic and content based features), (b) Applying feature selection methods to select discriminative features in content based approach and (c) Fusion of both features to provide useful insights of ironic contents. The content based feature subset is selected using a two stage feature selection method. In the first stage, the conventional feature selection methods such as Chi-square (χ^2), Information Gain (IG) and Mutual Information (MI) are used to select relevant feature subsets from a high dimensionality feature space. The selected feature subset may have features which convey similar information. Due to this reason, the features exhibiting similar information are grouped and features belonging to each group are selected. The second stage of feature selection is used to select the representative feature from each first stage feature subset. In this second stage, the features are grouped based on features exhibiting similar information. The clustering algorithm is used to cluster or group the similar information features subset. In this work, k-means clustering algorithm is used to cluster such that the members in each group are as similar (close) as possible to one another. The feature nearer to each cluster center is consider as

the representative feature among other features within the cluster. On the other hand, the linguistic features are extracted and categorized into five groups such as Rating, Word, Acronym, Symbol and Emoticon Features. Overall twenty special features are extracted using linguistic based feature extraction and categorized into these groups. The special features symbolize frequency of occurrences of the each feature in a review. Further, the special feature and content based feature subset are fused and evaluated using various classifiers such as Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF) and Decision Tree (DT) on product review dataset [20]. Furthermore, we construct an ensemble of four classifiers: SVM, LR, RF and DT based on combination rule to enhance the classification of the classifiers using weighted majority voting scheme.

III. Methodology

The proposed approach is a hybrid feature fusion method, which integrates linguistic features and content based text features. The proposed approach is used to classify product reviews into ironic or non-ironic content based on a feature fusion method. The general architecture of the proposed feature fusion approach is outlined in Fig. 1.

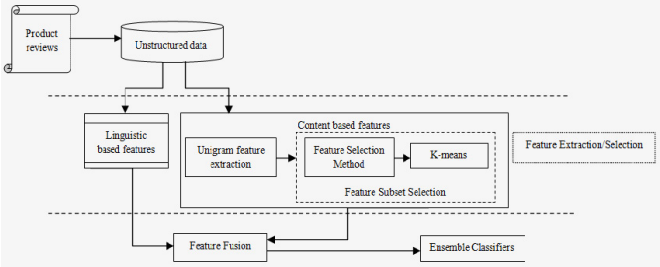


Fig. 1. Block diagram of the proposed feature fusion method.

As depicted in Fig. 1, the proposed approach comprises three phases: (i) feature extraction and selection, (ii) feature fusion and (iii) ensemble classification. The details of product review dataset are presented in section IV (A) and the rest of the above are described in this section. In order to identify ironic customer reviews, we developed feature fusion of linguistic and content based features. The fused features are classified using individual and ensemble classifiers such as Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DT) and Random Forest (RF) classifiers.

A. Feature Extraction and Selection for Irony Detection

In the classification of irony on product reviews, feature extraction and selection plays a vital role. Usually, ironic utterances expresses opposite meaning of the intended content. To extract ironic utterances present in reviews, linguistic based feature extraction is used. On the other hand, text features present in reviews yield promising results in the area of text classification [28] and sentiment analysis [25]. Hence, content based feature extraction and selection of features are used to detect ironic utterances in text.

B. Linguistic Feature Extraction

In this work, the features are extracted using a linguistic approach (hereafter, special features) to identify ironic utterances in text. To extract special features from sentences, we use syntactic information such as interjections, pragmatic, intensifier and many more. The special features are grouped into five sets of features. Table I provides an overview of group of features extracted using the linguistic approach.

In order to extract ironic feature, in [22] briefed that irony always express opposite of its actual content. The rating feature set groups the star rating (i.e 1* to 5* rating) for the reviews and the imbalance feature. The imbalance between star rating and over-all polarity of words in the review are considered. The work in [18], assumes there is imbalance when the star rating (i.e., 4* and 5*) is considered as positive review but polarity of reviews are negative. Similarly, imbalance exists when star-rating (1* and 2*) is considered as negative review but polarities of reviews are positive. This imbalance between the start rating and polarity of the text is considered as imbalance feature. The polarity of the review text is determined based on dictionary of [23], which consists of 6,800 words with positive and negative polarity words.

TABLE I. Feature Groups Used for Irony Detection in Product Review Datasets

No.	Groups	Features
1	Rating Features	Star rating, Imbalance feature
2	Word Feature	Hyperbole, Quotes, Ellipsis, Punctuation, Interjection
3	Acronym Feature	Acronym for laughter, Onomatopoeia for laugher, Acronym for Grin
4	Symbol Feature	Exclamation mark, Question mark, Ellipsis only, Ellipsis and Punctuation mark, Question and Exclamation mark.
5	Emoticon Feature	Sad, Happy, Laughing, Surprise, Winking

In general, user tends to exaggerate his/her sentiment through quoting certain words in sequences or between symbols. The word feature set consists of the features related to polarity of words present in the reviews. The feature hyperbole [24] implies the exaggeration present in sentences, which are extracted, based on appearance of three consecutive positive or negative words in a row. The feature quotes considers two consecutive intensifiers such as noun, adverbs and adjectives, which have positive or negative polarity in quotation marks [18]. In linguistic, ellipsis refers to the omission of words rather than repeating them unnecessarily and it is represented as three consecutive dots ("..."). In this work, feature ellipsis is considered as any positive or negative word end with an ellipsis. The feature punctuation considers positive or negative word with punctuation mark such as question marks and exclamation mark. The feature interjection indicates the occurrence of terms such as "wow", "ah", "aha" and many more in the sentences.

The use of language is constantly changing across space and across social group. The usage of acronymic word in micro blog text has grown enormously over the time. The feature such as acronym for laughter (lol, lawl, luls and many more) is used as a short form of laugh. The Onomatopoeia feature mimic the verbal conversation for laugh such as "haha", "mu-ba", "hehe", "hihi" and many more in the sentence. The feature acronym for Grin (*g*, *gg* and many more) depicts expression smiling broadly. These features are grouped into acronym feature set to describe emotions jargon present in reviews.

Usually, user tends to highlight his/her emotion by making more intense through exclamation mark, question mark, ellipsis and combination of these symbols. The symbol feature set groups these set of features to intensify the sentiment present in review text. The feature exclamatory mark ("!") symbolizes to express strong emotions in a review. Similarly, Question marks ("?") represent uncertainty about something in the sentence. Hence, exclamatory mark and question mark symbols are considered as set of features. The ellipsis only feature indicates the situation in which words are left out of a sentence but the sentence can still be understood by its context. The consecutive three dots ("...") are scrutinized as ellipsis feature. Further, the combination of ellipsis and punctuation (ellipsis followed by multiple exclamation or combination of exclamation and question mark), question mark and exclamation mark are considered as another set of features.

In social media text, emoticon plays a vital role in expressing a

person's feelings or mood by using punctuation marks, numbers, and letters. The emoticon feature set consists of emoticons such as Sad (":(", ")- :", "):" and ")o:"), Happy (":-)", ":0)", ":)", "(:" and many more), Laughing (":-D" and "xD"), Surprise (":-0", "0-:" and "0.0") and Winking (";-)", "; 0)" and ";)") as features. Overall, twenty linguistic features are extracted from five sets of features and hereafter linguistic features are referred as Special Feature (SF). However, in [18] the features are concentrated more on twitter related data. In this work, we generalize the linguistic features more on review text.

C. Content based Feature Extraction

In addition to special features, the content based features are extracted. To extract content based features, raw data are preprocessed by removing non-informative and trivial information. The review text consists of digits, punctuation, HTML tags and stop words, which occur more often and do not contribute to the analysis [57]. The preprocessed text are represented using unigram features with term frequency (*tf*) schema. The work in [32] suggests that unigram with term frequency (*tf*) performs well on sentiment analysis for micro-blogging data. Hence, we considered, unigram with term frequency (*tf*) schema to represent review text into machine understandable form. The extracted features are high in dimension feature space, which need to be reduced to low dimensionality feature space by applying various feature selection methods. The aim of feature selection methods is to select relevant and non-redundant features from high dimensionality feature space. In this work, the two stages of Feature Subset Selection (FSS) are used to select the most discriminative feature from high dimensionality feature space.

D. Feature Subset Selection (FSS)

The feature subset selection consists of two stages: In the first stage, the conventional feature selections method is used to select relevant feature subsets from a high dimensionality feature space. The selected feature subset may have features which convey similar information. Due to this reason, the features exhibiting similar information are grouped and feature belonging to each group are selected. The second stage feature selection is used to select the representative feature from each first stage FSS group.

In the first stage, the conventional feature selection methods such as CHI-square (χ^2), Information Gain (IG) and Mutual Information (MI) feature selection methods are used. These features selection methods are widely used to select relevant feature subset in text processing domain. The Chi-square (χ^2) is a statistical method used to test specific feature correlated with the class. The higher value of feature (χ^2) score indicates the likelihood of feature occurrence is highly dependent on the occurrence of the class. The IG is frequently used in the field of machine learning to determine the term of goodness criterion. IG measures the information that is gained by knowing the value of the attribute, which is the difference between the entropy of the distribution before the split and the entropy of the distribution after the split. The higher IG value indicates features contribute with more information for category prediction of the review, whereas lesser IG value indicates they do not add much information. Similarly, MI measures how much information a feature contains about a class. If the feature distribution is the same in intra-class and inter-class, the MI value reaches the minimum. Otherwise, MI value reaches the maximum when feature distribution is in intra-class only.

Let there be m the number of reviews and n the number of total features in the feature space. The preprocessed texts are represented in the form of a Document Term Matrix *DTM* ($m \times n$). The feature selection methods are applied to select relevant features among the features space. The conventional feature selection methods generate scores (S) corresponding to each feature. The feature scores are arranged in descending order to select top rated feature scores. The feature subsets are selected by fixing threshold value (l) empirically. The first stage feature subset selection using conventional feature selection methods has been shown in Algorithm 1.

Algorithm 1: First stage feature subset selection

Data: Document Term Matrix *DTM* ($m \times n$); m= number of reviews; n = total number of features; S = Feature score;

Result: l number of features

Initialize threshold value to l

//compute score for each feature using Feature Selection Method (FSM)

Step1: $S = FSM[DTM]$

Step 2: $S = [S_1, S_2, \ldots\ldots S_n]$ //n = total number of features

Step 3: Sort S in descending order // to select top ranked feature scores

Step 4: Select first l number of features from S, the selected feature subset l is represented using Document Term Matrix *DTM* ($l \times m$), where m represents the number of reviews and l indicates the number of feature subset selected from feature selection methods

($l << n$) .

The work presented in [26] suggested that features may convey similar information in the feature space. In conjecture to that, features evininge similar information are grouped to select the most representative features from each group. Due to this reason, the second stage feature selection is applied to select the most representative features from the feature subset obtained from the first stage.

In the second stage, the features are grouped based on features exhibiting similar information. The clustering algorithm is used to cluster or group the similar information feature subset. In this work, k-means clustering algorithm is used to group or cluster, such that the members in each group are as similar (close) as possible to one another. The k-means clustering algorithm works iteratively to assign features to one of the k clusters based on the similar information features. To determine the optimal number of clusters (k) as mentioned in [26] [29], number of cluster (k) is varied from $\sqrt{l}$ to $l/2$. The feature nearer to each cluster centers is considered as the representative feature among other features within the cluster. The cosine similarity measure is used to determine the similarity between the features and cluster center.

The *DTM* ($m \times l$) is obtained from the first stage of FSS and is transposed and represented as Term Document Matrix *TDM* ($m \times l$), where l indicates the number of features in the subset and m represents the number of reviews. The algorithm 2 presents the second stage of FSS to select the most discriminative features among the features exhibiting similar information.

Algorithm 2: Second stage of feature subset selection

Data: Term Document Matrix *TDM* ($m \times l$), l = number of features, m = number of reviews, k = clusters

Result: *TDM'* ($k \times m$), k = number of discriminative features ($k << l$)

Initialize k points at random as cluster centers, t=0, $F = \{\}$, $U^0 = \{u^0_1, \ldots\ldots u^0_k\}$

Repeat

Step 1: *for* $i \leftarrow 1$ *to* k *do*

Step 2: *for* $j \leftarrow 1$ *to* l *do*

//compute distance between cluster center U_i and feature F_j

$D_{ij} = dist(U^t_i, F_j)$

end

end

Step 3: Assign F_j to nearest cluster center u_i

// Update cluster center

Step 4: *for* $i \leftarrow 1$ *to* k *do*

$$u_i^{(t+1)} = \frac{\sum_{j \in i} F_j}{|u_i|}$$

end

Until $U^{(t-1)} - U^{(t)} < \varsigma$, t=t+1

//find representative feature for each cluster center

Step 5: *for* $i \leftarrow 1$ *to* k *do*

$F_j = \min_j \text{dist}_{j \in i}(u_t, F_{ji})$

$F = F \cup F_j$

end

Further, the feature subset F, which consists of k number of discriminative features is represented in TDM' ($k \times m$), where m is the number of reviews and k is the reduced number of features $k < l < n$.

E. Feature Fusion

In addition to linguistic features, content based features play a significant role in sentiment analysis [17]. In order to capture various dimensions of characteristics of review, linguistic features and subset of content based features are fused. Overall twenty special features are extracted using linguistic based feature extraction and categorized into five groups. The special features symbolize frequency of occurrences of each feature in a review. On the other hand, content based feature subsets selections are applied to select the most discriminative features from a high dimensionality feature space. The special feature and content based Feature Subset Selection (FSS) are fused and evaluated using various classifiers.

F. Ensemble Classifiers

In this work, the predictive classifiers such as Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DT) and Random Forest (RF) are used to evaluate the efficiency of the proposed method. The SVM classifier works by finding the hyper-plane, which maximizes the margin between the two classes. The vectors that define the hyper-plan are known as support vectors. The SVM is used for both classification and regression problems [34]. The LR is a statistical method used for binary classification problems (problems with two class values) [33]. The DT is a non-parametric approach used to construct a tree in top-down and recursive divide-and-conquer manner [17]. DT is mainly used for classification and regression problem. Similarly, RF is an automatic learning technique which combines the concepts of random subspaces and bagging [36]. Random Forest operates by constructing a multitude of decision trees at training time and outputting the class based on decision of individual trees [35]. These four classifiers are widely used in classification of ironic content in reviews [18] [30].

On other hand, work of [37] [38] emphasizes that individual prediction of various classifiers can be ensemble so that a more robust and accurate classification model can be built. The ensemble learning plays a vital role in recent research activity of pattern recognition and machine learning [38]. The main aim of the ensemble learning is to weigh several individual classifiers and combine the prediction of the multiple classifiers, which outperforms prediction of individual classifiers [31]. The majority voting and weighted majority voting are the most popular combination schemas, which are widely used in ensemble classification [39]. The simple majority voting schema selects one of many alternatives of the predicted classes with the most votes [40]. The weighted majority voting schema assigns weight for each prediction of the classifiers based on the performance of the classifier [41]. In this work, weighted majority voting schema is used to assign the weight for each decision of the classifiers. For weighted majority voting schema, first let there be (H_1, H_2,....., H_l) classifiers with accuracies (A_1, A_2,....., A_l), respectively. Then, $h_{i,j}$ be defined as the decision of i^{th} classifier which chooses j^{th} label from class C. In weighted majority voting schema, weights w_i are assigned to the individual decision of the classifiers h_i. The ensemble classifier decision using weighted majority voting schema $H(X)$ is as follows:

$$H(X) = \max_{j=1,..,C} \sum_{i=1}^{I} w_i h_{i,j} \quad (1)$$

In weighted majority voting schema, the optimal weight w_i for the classifiers are assigned based on the accuracy of the classifiers h_i, i.e., $w_i \alpha {}^{a_i}/_{(1-a_i)}$ where a_i is the accuracy of the independent classifiers h_i. In [54], the weight w_i equip each classifier output with different weights between 0 and 1 value ($0 \le w_i \le 1$) and sum of the weight w_i is equal to 1. Hence, weights w_i are assigned based on the performance of the individual classifiers as indicated in [54].

IV. Experimental Results and Discussion

In this section, the effectiveness of feature fusion is examined on a publically available product review dataset [20].

A. Dataset Description

The Amazon product review dataset consists of 1,254 reviews, consists of 437 ironic and 817 non-ironic or regular reviews created by [20]. The structure of the dataset contains * rating ranging from 1* to 5* star, along with labeled ironic and regular review content. Table II depicts the distribution of reviews by star-rating.

TABLE II. Distribution of Reviews by Star-rating

		Number of reviews				
		1*	2*	3*	4*	5*
Ironic	437	262	27	20	14	114
Regular	817	64	17	35	96	605

B. Experimental Setup

In this work, we conducted an experiment based on 80 training and 20 testing splits on product review dataset [20]. The experiments are conducted similarly to [18], [30] to give comparison with existing methods. The experiments are conducted based on two baselines: Feature Subset Selection (FSS) that accomplishes two stages feature selection and Feature Fusion baseline that exploits feature fusion of linguistic features and FSS. Initially, twenty Special Feature (SF) are extracted using the linguistic approach from each review. On the other hand, various preprocessing techniques are applied to extract content based features in review text. The 20,985 distinct features are extracted and represented using unigram term frequency schema. Further, extracted features are processed to select discriminative feature using FSS. The Feature Subset Selection (FSS) consists of two stages: In the first stage, conventional feature selection methods such as Chi-square, IG and MI are used to select feature subsets. The features are arranged in descending order based on the scores obtained from individual feature selection methods. The threshold value is used to select top scored features from the feature space. The threshold value is varied between 1,000 to 20,000 numbers of features by empirically. In the second stage, k-means clustering is applied to select the most discriminative features from first stage feature subsets. The optimal number of cluster is determined based on [26] and it is explained in previous section. In Feature Fusion baseline, the extracted linguistic and selected content based features are fused. The extracted linguistic and selected content based features are represented using frequency of occurrence of each feature. Hence, both the features are fused and evaluated using SVM, LR, DT and RF classifiers. Further, individual predictions of classifiers

are ensembled based on weighted majority voting schema as explained in section III (F). The efficiency of the proposed method is evaluated based on Precision, Recall and F-measure obtained from individual classifiers and ensemble classifier.

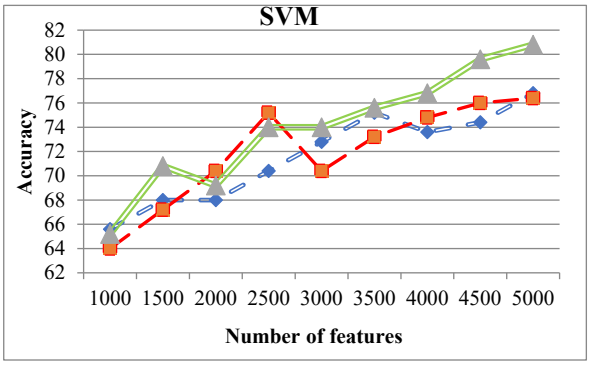
C. Experimental Results

To report the performance of the proposed method, the experiments are conducted in two baselines. In the first baseline, FSS consists of the two stage feature selection method: in the first stage threshold values are varied from 1,000 to 20,000 features and empirically is found that 10,000 features depict competitive results for various classifiers. Table III elucidates the comparisons of overall features with conventional feature selection method of the first stage of FSS using SVM, DT and RF classifiers. The Mutual Information (MI) with 10,000 features outperforms the other selection methods in accuracy, precision, recall and f-measure using SVM, DT and RF classifiers. The Information Gain (IG) with 10,000 feature subset achieves 0.658 recall compared to other feature selection methods using DT classifier. Overall from Table III, we can observe that feature selection method plays a vital role in reducing high dimensional feature space and maximizing the performance of the classifiers. Further, the second stage of FSS is applied on 10,000 features obtained from various feature selection methods.

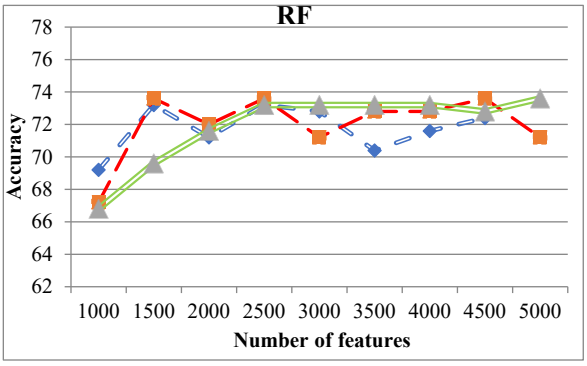
In the second stage of FSS method, k-means clustering algorithm is used by varying cluster numbers with difference of 500 from 100 to 5000 features. The k-means clustering with cosine distance yields promising results compared to Euclidean, City block, Correlation and Hamming Distance metrics [44] [45]. Hence, k-means clustering with cosine distance is used to select most discriminative features from the first stage FSS method. As explained in Algorithm 2, similar features are clustered and the features nearer to the cluster center are considered as the most discriminative feature subset. Fig. 2 depicts the classification accuracy of the proposed method with varying feature subsets using SVM, RF, LR and DT classifiers. Fig. 1 shows the feature subset ranging from 1000 to 5000 features with differences of 500 features. However, feature subset less than 1000 does not yield promising results. Hence, the feature subsets less than 1000 are excluded in Fig. 2. From Fig. 2 (a), we observe that the FSS of MI with k-means gradually increases and reaches maximum accuracy for 5000 features using SVM classifier. Similar variation is observed in Fig. 2 (b) and Fig. 2 (c), where MI with k-means clustering reaches maximum accuracy for 5000 features using RF and LR classifiers, respectively. In Fig. 2 (d), CHI-square with k-means clustering achieves maximum accuracy for 2500 features using DT classifiers. Overall, performance of MI with k-means feature subset increases with increasing number of feature subset using SVM, RF, LR and DT classifiers.

TABLE III. Performance Comparison of Overall Features with Reduced Set of Features; A-Accuracy, P-Precision, R-Recall, F-F-measure

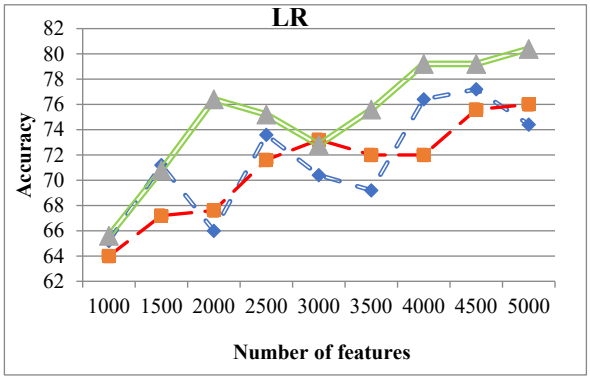
	No. of Features	SVM				DT				RF			
		A	P	R	F	A	P	R	F	A	P	R	F
Total features	20985	68.40	0.646	0.597	0.595	64.80	0.606	0.601	0.603	69.20	0.748	0.566	0.529
CHI	10000	73.60	0.771	0.637	0.638	68.40	0.645	0.626	0.631	69.60	0.755	0.571	0.539
IG	10000	73.60	0.756	0.642	0.646	70.80	0.675	0.658	0.663	70.00	0.761	0.577	0.548
MI	10000	75.60	0.791	0.666	0.675	72.00	0.693	0.657	0.664	71.60	0.821	0.595	0.572



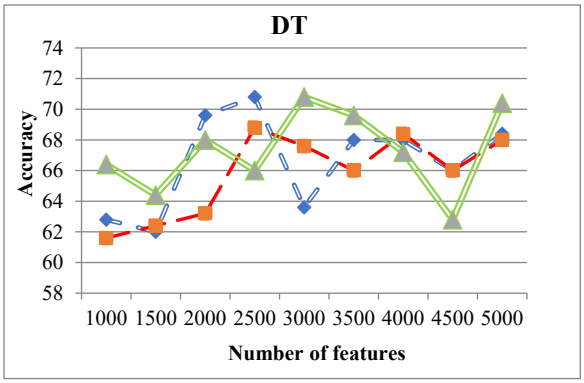
2 (a)



2 (b)



2 (c)



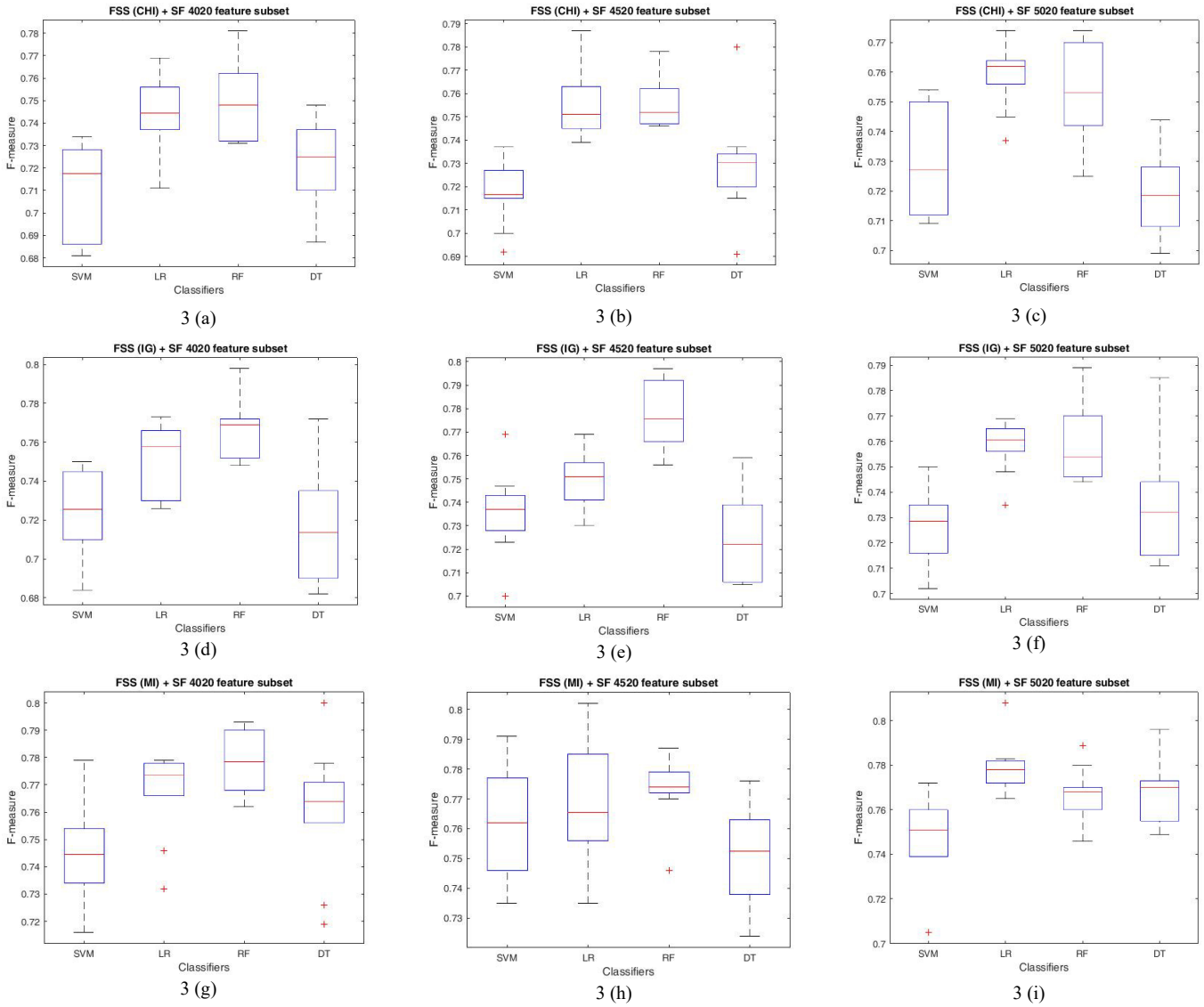
2 (d)

CHI-square + k-means Information Gain + k-means Mutual Information + k-means

Fig. 2. Classification accuracy of feature selection methods with various feature subsets.

The feature fusion baseline consists of Special Features (SF) and FSS features. Both features are in same weighting scheme i.e., frequency of occurrence of each feature. From Fig. 2, we observe that feature subset ranging from 4000 to 5000 yield promising result using various classifiers. Hence, feature fusion experiments are conducted by varying FSS from 4000 to 5000 features with 20 Special Features using 10 fold cross validation method. The number of features considered for feature fusion (FSS+SF) experiments are 4020, 4520 and 5020. In Fig. 3, the box plot representation is given to describe the variations observed in F-measure of feature fusion with various FSS using different classifiers. The Fig. 3 represents minimum, maximum and mean of 10 fold cross validation using SVM, LR, RF and DT classifiers. Fig. 3 (a), (b) and (c) present the result of feature fusion of FSS (Chi) + SF feature subsets using SVM, LR, RF and DT classifiers. In Fig. 3 (b), the feature fusion of FSS (Chi-square) achieves maximum F-measure for 4520 features using LR classifier. Fig. 3 (d), (e) and (f) present F-measure results of feature fusion of FSS (IG) + SF feature subsets. Fig. 3 (e) shows feature fusion of FSS (IG) + SF achieving maximum F-measure for 4520 feature subset using RF classifier. Fig. 3 (g), (h) and (i) present the results of feature fusion of FSS (MI) + SF using SVM, LR, RF and DT classifiers. From Fig. 3 (h), we observe that feature fusion of FSS (MI) + SF features achieve maximum F-measure for 4520 feature subset using the LR classifier. From Fig. 3, we can conclude that feature fusion (FSS (MI) + SF) for 4520 feature subset achieves maximum F-measure using SVM, LR, RF and DT classifiers. Hence, we consider feature fusion method (FSS (MI) + SF) with 4520 features for comparing with other existing methods.

Further, the classifiers are ensemble to seek collective opinions of the classifiers to enhance the performance of the models. Here, weighted majority voting schema is used to assign the weight for each decision of the classifiers. As explained in section III (F), weights are varied between 0 and 1, where the sum of the weights are equal to 1. The best combinations of weights are assigned empirically for each model. The Table IV presents the maximum accuracy obtained in 10 folds cross validation for the proposed feature fusion method (FSS (MI) + SF) with 4520 features using various classifiers. From Table IV, it can be inferred that SVM, DT and RF outperform LR classifier by a good margin in terms of accuracy, precision, recall and F-measure. The LR model performs least among other models due to the possibility of overfitting. Hence, the best possible weights assigned are 0.3 for SVM, 0.1 for LR, 0.3 for RF and 0.3 for DT classifiers, respectively. Table IV elucidates the performance comparison of the ensemble model with individual classifiers. The ensemble model outperforms other classifiers with a maximum margin in Accuracy, Precision, Recall and F-measure. Hence, the performance of the ensemble model



3 (a) 3 (b) 3 (c)

3 (d) 3 (e) 3 (f)

3 (g) 3 (h) 3 (i)

Fig. 3. F-measure of Feature Fusion with various Feature Subset Selection (FSS) using SVM, LR, RF and DT classifiers.

is considered for comparison with existing methods on product review dataset [20].

TABLE IV. Performance Comparison of Ensemble Model

Classifiers	Accuracy	Precision	Recall	F-Measure
SVM	82.00	0.814	0.779	0.791
LR	78.80	0.769	0.754	0.760
DT	80.00	0.781	0.772	0.776
RF	82.40	0.838	0.769	0.787
Ensemble	86.00	0.871	0.818	0.835

D. Comparison with Existing Methods

From literature, Buschmeier et al. [18] and Reganti et al., [30] describe irony/satire detection on product review dataset [20] using SVM, LR, DT and RF classifiers. The similar set of experiment is conducted and compared the results. The Table V elucidates the comparisons of the proposed method with [18] and [30] method. In [18], models are evaluated based on precision, recall and F-measure using linear SVM, LR, DT and RF classifiers. The features are extracted and fused using lexicon based (29 special features) and content based approach (21744 Bag of Words features). The total number of features considered for the model are 21773 features. On the other hand, Reganti et al. [30] performed feature fusion on 42 special features and baseline features. The baseline features such as character n-gram, word n-gram and word skip gram are used but the total number of features are not stated clearly. The fused features are classified using linear SVM, LR, DT and RF classifiers. Further, the decisions of individual classifiers are ensembled using weighted majority voting schema by varying weighted from 0 to 1 value. The performance of the model was evaluated using F-Measure. Hence, we conducted our experiment on same lines with 20 Special Features (SF) and 4500 text features using individual classifiers and ensemble classifiers using weighted majority voting schema. However, product review dataset [20] is used to describe the impact of sarcasm detection in sentiment shift [19]. Therefore, the proposed method is compared with the methods proposed in [18] and [30]. From Table V, it can be observed that the proposed feature fusion method outperforms the existing methods in terms of Precision, Recall and F-measure using individual classifiers and ensemble classifiers.

TABLE V. Performance Comparison with Existing Methods

Classifiers	Performance Measure	Methods		
		Buschmeier et al., [18]	Reganti et al., [30]	Proposed Method
SVM	Precision	0.680	--	0.814
	Recall	0.613	--	0.740
	F-Measure	0.643	0.740	0.791
LR	Precision	0.752	--	0.769
	Recall	0.622	--	0.754
	F-Measure	0.678	0.768	0.760
DT	Precision	0.598	--	0.781
	Recall	0.550	--	0.772
	F-Measure	0.569	0.681	0.776
RF	Precision	0.732	--	0.838
	Recall	0.241	--	0.769
	F-Measure	0.353	0.710	0.787
Ensemble	Precision	--	--	0.871
	Recall	--	--	0.818
	F-Measure	--	0.781	0.835

E. Discussion

Feature fusion is a process of integrating various characteristics of features to produce more consistent, accurate, and useful information. In irony detection, feature extraction and selection plays a vital role in determining the ironic utterances present in sentences. In this work, linguistic based and content based Feature Subset Selection (FSS) are extracted and features are fused to provide various characteristics of features of product review dataset [20]. The Table II elucidates distribution of product reviews by its star-rating. The 20 linguistic features are extracted and represented based on frequency of occurrence of features in each review. Table I presents 20 Special Feature (SF) extracted using linguistic approach. These features are used to detect ironic utterance present in reviews. On the other hand, content based features play a vital role in accurate classification of ironic reviews. The high dimensionality and sparsity is one of the major challenges faced during classification task. To curse the dimensionality, two stages of content based Feature Subset Selection (FSS) is proposed. In the first stage FSS, conventional Feature Selection Method (FSM) is applied to select relevant features set. In this experiment, conventional FSM such as Chi-square, IG and MI are used to select the relevant feature set from original features based on scores of individual FSM. It is noticeable from Table III, that first stage FSS using conventional FSM reduces high dimensionality feature space by selecting a subset of relevant features. The MI feature selection method yields maximum performance for 10,000 features compared to other FSM using SVM, LR, DT and RF classifiers, respectively. However, selected feature subset may have features which convey similar information. Due to this reason, second stage FSS is applied to select more discriminative features among feature subset. In the second stage, k-means clustering algorithm is used to group the features which convey similar information. Fig. 2 (a), (b), (c) and (d) present the classification accuracy of the proposed method with varying number of features (1000 to 5000) using SVM, RF, LR and DT classifiers. The MI with k-means consistently outrages other feature selection combination using various classifiers because MI compares the probability of observing features and class together (joint probability) instead of observing independently. MI reaches its maximum value, when the feature is a perfect indicator for class membership. On the other hand, k-means groups the similar information. The feature closer to cluster center is considered as the most representative feature within each cluster. The representative features are considered as discriminative features among other clusters. Hence, the combination of these methods exhibits maximum accuracy on all the classifiers. Further, features are fused to detect ironic utterances present in review using linguistic and content based FSS. A crucial observation is noted from Fig. 3 (a) – 3 (i), which depicts the F-measure of 10 folds cross validation of feature fusion using various classifiers. Fig. 2 (h) elucidates maximum F-measure on feature fusion (FSS (MI + k-means) +SF) of 4520 feature set using various classifiers. Further, the prediction of each classifier is ensemble to enhance the classification performance. The weighted majority voting scheme is used to ensemble different classifier based on the prediction of each classifier. Table IV elucidates the performance comparison of ensemble model with individual classifiers. The ensemble model outperforms other classifiers with a maximum margin of Accuracy, Precision, Recall and F-measure. To evaluate the effectiveness of the proposed method, the proposed method is compared with existing methods present in the literature. From Table V, it can be observed that the proposed method outperforms other existing methods in terms of minimal number of feature set and the maximum classification performance of individual and ensemble classifiers.

V. Conclusion and Future Work

In this paper, we employed feature fusion to capture various dimensions of characteristics of the reviews. The proposed approach extracts features based on linguistic and content based text features. The five types of features are extracted using linguistic approach, viz: Rating, Word, Acronym, Symbol and Emoticon Features. On the other hand, content based text feature consists of two stages of Feature Subset Selection (FSS) to select the most discriminative features. Both the features are fused and classified using SVM, LR, RF and DT classifiers. With the series of experimentation, we demonstrated the proposed approach has an ability to capture ironic utterances present in the reviews. To enhance the performance of the classifiers, we make use of weighted majority voting schema to create an ensemble from the decision of each classifier. The results show that the proposed feature fusion out-performs the existing methods on benchmark dataset.

In future, the proposed approach can be extended to (i) extract more number of linguistic features to identify ironic utterances present in text, (ii) use of more sophisticated feature selection methods and (iii) employing various clustering algorithms along with ensemble methods which enhances the classification accuracies. Further, the proposed approach can be also extended to various fields such as sentiment classification, spam detection and many more.

Acknowledgment

H M Keerthi Kumar has been financially supported by UGC under Rajiv Gandhi National Fellowship (RGNF) Letter no: F1-17.1/2016-17/RGNF-2015-17-SC-KAR-6370/(SA-III Website), JSSRF (University of Mysore), Karnataka, India.

References

[1] M. P. S. Bhatia, and A. K. Khalid, Information retrieval and machine learning: Supporting technologies for web mining research and practice. Webology, 2008, vol. 5, pp. 2–19.

[2] G. Wang, J. Sun, J. Ma, K. Xu, and J. Gu, "Sentiment classification: The contribution of ensemble learning". Decision Support Systems, 2014, vol. 57, pp. 77–93.

[3] E. Fersini, E. Messina, and F. A. Pozzi, "Sentiment analysis: Bayesian ensemble learning". Decision Support Systems, 2014, vol. 68, pp. 26–38.

[4] A. Onan, S. Korukoğlu, and H. Bulut, "A hybrid ensemble pruning approach based on consensus clustering and multi-objective evolutionary algorithm for sentiment classification". Information Processing & Management, 2017, vol. 53(4), pp. 814-833.

[5] V. David. "Towards a sociology of attitudes", Sociological Research Online, 2011, vo. 19.1, pp. 12.

[6] R. J. Davidson, K. R. Sherer, and H. H. Goldsmith, "Handbook of affective sciences", Oxford University Press, 2009.

[7] J. R. Martin, and R. Peter White, "The language of evaluation". 2003, vol. 2. Basingstoke: Palgrave Macmillan.

[8] B. Pang, and L. Lee, "Opinion mining and sentiment analysis". Foundations and Trends in Information Retrieval, 2008 vol. 2(1–2), pp. 1-135.

[9] C. C Florian, F. A. Kunneman, and A. P. J. Van den Bosch, "The perfect solution for detecting sarcasm in tweets #not", In Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, WASSA 2013.

[10] I. Hernández-Farías, J. M. Benedí, and P. Rosso, "Applying basic features from sentiment analysis for automatic irony detection", In Iberian Conference on Pattern Recognition and Image Analysis, 2015, pp. 337-344. Springer, Cham.

[11] S. Attardo, "Irony markers and functions: Towards a goal-oriented theory of irony and its processing", Rask, 2000, vol. 12(1), pp. 3-20.

[12] R. González-Ibánez, S. Muresan, and N. Wacholder, "Identifying sarcasm in Twitter: a closer look", In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers-Volume 2, pp. 581-586, Association for Computational Linguistics. 2011

[13] S. G. Wicana, T. Y. İbisoglu, and U. Yavanoglu, "A Review on Sarcasm Detection from Machine-Learning Perspective", In Semantic Computing (ICSC), 2017 IEEE 11th International Conference on, pp. 469-476. IEEE.

[14] A. D. Dave, and N. P. Desai, "A comprehensive study of classification techniques for sarcasm detection on textual data", In Electrical, Electronics, and Optimization Techniques (ICEEOT), International Conference on, pp. 1985-1991. IEEE, 2016.

[15] S. K. Bharti, B. Vachha, R. K. Pradhan, K. S. Babu, and S. K. Jena, "Sarcastic sentiment detection in tweets streamed in real time: A big data approach", Digital Communications and Networks, 2016, vol. 2(3), pp. 108-121.

[16] A. Joshi, P. Bhattacharyya, and M. J. Carman, "Automatic Sarcasm Detection: A Survey", ACM Computing Surveys (CSUR), 2017, vol. 50(5), pp. 73.

[17] K. Ravi, and V. Ravi. "A novel automatic satire and irony detection using ensembled feature selection and data mining", Knowledge-Based Systems, 2017, vol. 120, pp. 15-33.

[18] K. Buschmeier, P. Cimiano, and R. Klinger. "An Impact Analysis of Features in a Classification Approach to Irony Detection in Product Reviews", In WASSA@ ACL, pp. 42-49. 2014.

[19] E. Filatova, "Sarcasm Detection Using Sentiment Flow Shifts", Florida Artificial Intelligence Research Society Conference, 2017, pp 264-269.

[20] E. Filatova, "Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing", In LREC, 2012, pp. 392-398.

[21] R. Justo, T. Corcoran, S. M. Lukin, M. Walker, and M. I. Torres, "Extracting relevant knowledge for the detection of sarcasm and nastiness in the social web". Knowledge-Based Systems, 2014, 69, pp. 124-133.

[22] D. Davidov, O. Tsur, and A. Rappoport, "Semi-supervised recognition of sarcastic sentences in twitter and amazon", In Proceedings of the fourteenth conference on computational natural language learning, 2013, pp. 107-116. Association for Computational Linguistics.

[23] M. Hu, and B. Liu, "Mining and summarizing customer reviews", Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2004.

[24] R. W. Gibbs, "Irony in talk among friends", Metaphor and symbol, 2000, vol. 15(1-2), pp. 5-27.

[25] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques", In Proceedings of the ACL-02 conference on Empirical methods in natural language processing, 2002, vol. 10, pp. 79-86. Association for Computational Linguistics.

[26] S. Dey Sarkar, S. Goswami, A. Agarwal, and J. Aktar, "A Novel Feature Selection Technique for Text Classification Using Naïve Bayes", International Scholarly Research Notices, 2014.

[27] A. Reyes, and P. Rosso, "Mining subjective knowledge from customer reviews: A specific case of irony detection", In Proceedings of the 2nd workshop on computational approaches to subjectivity and sentiment analysis, 2011, pp. 118-124. Association for Computational Linguistics.

[28] J. Fürnkranz, "A study using n-gram features for text categorization", Austrian Research Institute for Artificial Intelligence, 1998, vol. 3, pp. 1-10.

[29] G. Li, X. Hu, X. Shen, X. Chen, and Z. Li. "A novel unsupervised feature selection method for bioinformatics data sets through feature clustering", In Granular Computing, 2008, IEEE International Conference on, pp. 41-47. IEEE,

[30] A. N. Reganti, T. Maheshwari, U. Kumar, A. Das, and R. Bajpai, "Modeling Satire in English Text for Automatic Detection", In Data Mining Workshops (ICDMW), 2016 IEEE 16th International Conference on (pp. 970-977). IEEE.

[31] L. Rokach, "Ensemble-based classifiers" Artificial Intelligence Review 33, no. 1-2 (2010), pp. 1-39.

[32] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment analysis of twitter data". In Proceedings of the workshop on languages in social media, 2011, pp. 30-38. Association for Computational Linguistics.

[33] A. B. Musa, "Comparative study on classification performance between support vector machine and logistic regression", International Journal of Machine Learning and Cybernetics, 2013, Vol. 4(1), pp. 13-24.

[34] I. H. Witten, E. Frank, M. A. Hall, C. J. Pal, DataMining: Practical machine learning tools and techniques. Morgan Kaufmann, 2016.

[35] A. Bouaziz, C. Dartigues-Pallez, C. da Costa Pereira, F. Precioso and P. Lloret, "Short text classification using semantic random forest", in: International Conference on Data Warehousing and Knowledge Discovery, Springer, 2014, pp. 288–299.
[36] Y. Al Amrani, M. Lazaar, and K. E. El Kadiri, "Random Forest and Support Vector Machine based Hybrid Approach to Sentiment Analysis", Procedia Computer Science, 2018, vol. 127, vol. 511-520.
[37] T. G. Dietterich, "Ensemble methods in machine learning", In International workshop on multiple classifier systems, 2000, pp. 1-15, Springer, Berlin, Heidelberg.
[38] A. Onan, S. Korukoğlu, and H. Bulut, "A hybrid ensemble pruning approach based on consensus clustering and multi-objective evolutionary algorithm for sentiment classification", Information Processing & Management, 2017, vol.53(4), pp. 814-833.
[39] Y. Zhang, H. Zhang, J. Cai, and B. Yang, "A weighted voting classifier based on differential evolution". In Abstract and Applied Analysis, 2014 vol. 2014. Hindawi.
[40] L. I. Kuncheva, and J. J Rodríguez, "A weighted voting framework for classifiers ensembles". Knowledge and Information Systems, 2014, vol. 38(2), pp. 259-275.
[41] A. Ekbal, and S. Saha, "Weighted vote-based classifier ensemble for named entity recognition: A genetic algorithm-based approach". ACM Transactions on Asian Language Information Processing (TALIP), 2011, vol. 10(2), pp. 9.
[42] R. J. Kreuz, and R. M. Roberts, "Two cues for verbal irony: Hyperbole and the ironic tone of voice". Metaphor and symbol, 1985, vol. 10(1), pp. 21-31.
[43] A. Utsumi, "Verbal irony as implicit display of ironic environment: Distinguishing ironic utterances from nonirony", Journal of Pragmatics, 2000, vol. 32(12), pp. 1777-1806.
[44] A. Strehl, J. Ghosh, and R. Mooney, "Impact of similarity measures on web-page clustering", In Workshop on Artificial Intelligence for Web Search (AAAI 2000), pp. 58-64.
[45] L. Sahu, and B. R. Mohan, "An improved K-means algorithm using modified cosine distance measure for document clustering using Mahout with Hadoop", In Industrial and Information Systems (ICIIS), 2014 9th International Conference on, pp. 1-5. IEEE.
[46] K. Chakraborty, S. Bhattacharyya, R. Bag, and A. E. Hassanien, "Comparative Sentiment Analysis on a Set of Movie Reviews Using Deep Learning Approach", In International Conference on Advanced Machine Learning Technologies and Applications, 2018, pp. 311-318.
[47] O. Tsur, D. Davidov, and A. Rappoport. "ICWSM-A Great Catchy Name: Semi-Supervised Recognition of Sarcastic Sentences in Online Product Reviews", In ICWSM, 2010, pp. 162-169.
[48] B. S. Harish, M. B. Revanasiddappa, "A comprehensive survey on various feature selection methods to categorize text documents". International Journal of Computer Applications, 2017, vol. 164, pp. 1–7.
[49] A. Yousefpour, R. Ibrahim, and H. N. A. Hamed, "Ordinal-based and frequency-based integration of feature selection methods for sentiment analysis". Expert Systems with Applications, 2017, vol. 75, pp. 80-93.
[50] A. Tripathy, A. Agrawal, S. K. Rath, "Classification of sentiment reviews using n-gram machine learning approach". Expert Systems with Applications, 2016, vol. 57, pp. 117–126.
[51] A. Joshi, V. Sharma, P. Bhattacharyya, "Harnessing context incongruity for sarcasm detection", Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics Int. Joint Conf. Natural Lang. Process. (ACL-IJCNLP), 2015, vol. 2, pp. 757-762.
[52] M. Bouazizi, and T. O. Ohtsuki, "A pattern-based approach for sarcasm detection on Twitter", IEEE Access, 2016, vol. 4, pp. 5477-5488.
[53] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey". Ain Shams Engineering Journal, 2014, vol. 5(4), pp. 1093-1113.
[54] H. Toman, L. Kovacs, A. Jonas, L. Hajdu, and A. Hajdu, "Generalized weighted majority voting with an application to algorithms having spatial output", In International Conference on Hybrid Artificial Intelligence Systems, 2011, pp. 56-67.
[55] U. Yavanoglu, T. Y. Ibisoglu, and S. G. Wıcana, "Technical Review: Sarcasm Detection Algorithms". International Journal of Semantic Computing, 2018, vol. 12(03), pp. 457-478.
[56] H. M. Kumar, B. S. Harish, and H. K. Darshan. "Sentiment Analysis on IMDb Movie Reviews Using Hybrid Feature Extraction Method". International Journal of Interactive Multimedia and Artificial Intelligence, 2019, vol. 5(5), pp. 109-114.
[57] H. M. Kumar and B. S. Harish. "Classification of Short Text Using Various Preprocessing Techniques: An Empirical Evaluation." In Recent Findings in Intelligent Computing Techniques pp. 19-30. Springer, Singapore, 2018.

H. M. Keerthi Kumar

He received B.E in Information Science and Engineering and M.Tech in Software Engineering from Visvesvaraya Technological University, India. He is currently pursuing Ph.D degree in Computer Science from University of Mysore, India. His area of research includes Data Mining, Pattern Recognition and Machine Learning.

B. S. Harish

He obtained his B.E in Electronics and Communication (2002), M.Tech in Networking and Internet Engineering (2004) from Visvesvaraya Technological University, Belagavi, Karnataka, India. He completed his Ph.D. in Computer Science (2011); thesis entitled "Classification of Large Text Data" from University of Mysore. He is presently working as an Associate Professor in the Department of Information Science & Engineering, JSS Science & Technology University, Mysuru. He was invited as a Visiting Researcher to DIBRIS - Department of Informatics, Bio Engineering, Robotics and System Engineering, University of Genova, Italy from May-July 2016. He delivered various technical talks in National and International Conferences. He has invited as a resource person to deliver various technical talks on Data Mining, Image Processing, Pattern Recognition, and Soft Computing. He is also serving and served as a reviewer for National, International Conferences and Journals. He has published more than 50 International reputed peer reviewed journals and conferences proceedings. He successfully executed AICTE-RPS project which was sanctioned by AICTE, Government of India. He also served as a secretary, CSI Mysore chapter. He is a Member of IEEE (93068688), Life Member of CSI (09872), Life Member of Institute of Engineers and Life Member of ISTE. His area of interest includes Machine Learning, Text Mining and Computational Intelligence.

"Hello, is There Anybody Who Reads Me?" Radio Programs and Popular Facebook Posts

Tal Laor*

Ariel University (Israel)

Received 3 March 2019 | Accepted 2 October 2019 | Published 30 October 2019

Abstract

Radio stations are increasingly active on social networks, as radio continues to adjust and adapt to online spaces. This research is intended to conceptualize and characterize the success of radio programs beyond their native FM environment, focusing on their attempts at achieving popularity on social networks. Success on social networks is measured by user involvement and interaction with posted content and comments. This study looked at the activity of leading Israeli radio programs on Facebook pages and user engagement, evaluating highly involved posts by coding. It was found that radio program activity on social networks expands the reach of radio stations and promotes higher levels of interaction with listeners beyond broadcast schedule. In addition, integration of various media forms such as videos or images increases the likelihood of a post becoming popular. This research presents the convergence of radio programs in accordance with the theoretical framework of technological determinism.

Keywords

Convergence, Facebook, Radio, New Media, Social Networks.

DOI: 10.9781/ijimai.2019.10.003

I. Introduction

Radio is a traditional media form in the process of coping with recent new media developments. New media has thus emerged as increasingly powerful, even dominant over other forms of media such as radio. One of the most significant new challenges for radio is the rise of social media. Radio now seeks new ways to stay current and relevant. A growing number of radio stations manage a social network presence. They do this to promote the station and reinforce the connection and involvement of listeners. This research tested the activity of Israeli radio program stations on Facebook, focusing on analysis of popular measures of engagement. We thus looked at the following questions in relation to Israeli radio program Facebook pages: How do radio programs exploit the visual aspects of Facebook? Does this exploitation make the content popular on Facebook? What constitutes the most popular posts and is there a correlation between typical subjects that generate engagement (such as celebrities) and those that do the same on Facebook? How do Facebook pages influence viewer/listener behaviors? We will examine the relevance of the theory of technological determinism on the activity of radio programs on Facebook and the convergence of radio stations and the Internet.

II. Literature Review

Unique features characterize radio as a means of communication. As a medium, radio can distribute copious information. But it has also developed vital connections with its listeners. As a result, it provides them with answers to personal cognitive needs. These can be distilled into four fundamental features: intimacy, credibility, authoritativeness, and accessibility. First, conversations broadcasted on the radio tend to create for the listener a feeling of intimate and personal interaction. McLuhan referred to this as "a private experience" [1]. The technology of broadcasting enables this sense of intimacy. The distance between broadcaster and microphone is heard by the ear of the listener as the distance between one person and another in close conversation. Therefore, radio programs have adopted a culture of speech and manner of language to correlate with their potential target audiences. Secondly, radio tends to be conceived by listeners as reliable and reflecting real social conditions [2]. The feeling of trust translates into a sense of authoritativeness that listeners imbue into radio. Radio capitalizes on this notion of itself as a source of credibility by fostering a variety of opinions, which are intended as for broadening the conversation. As such, persistent recruitment of commentators, specialists, and professionals as well as the audience itself encourages participation in many programs. However, radio is primarily a tool for providing information accessibility to the general population. Listening to the radio does not require special skills needed to navigate other media forms such as written journalism. Finally, traditional radio is usefully described as a secondary medium for accompanying other initial actions. Since radio concentrates entirely on the sense of hearing, simultaneous performance of parallel tasks is facilitated. Radio is thus amenable to being used as a kind of "background" for other activities [1][3][4].

As noted, the language of conversation constitutes a unique quality of radio. Because of its exclusive auditory focus, it is called "a blind medium" [5][6]. As such, it does not require the same level of attention and commitment as visual media. As the listener is free to engage in parallel tasks while listening, the broadcaster is not required to ascribe any importance to physical appearance, much less any other visual element [5].

The traditional features of radio were ultimately confronted by the Internet Era [7][8]. However, radio as an industry is slowly coming to the realization that it must leverage the Internet and multimedia

* Corresponding author.
E-mail address: tallaor10@gmail.com

in order to survive in a market dominated by digital competition. Traditional radio is now animated by a new purpose to give listeners fresh possibilities. Radio is adapting to digital broadcasting to meet listener needs and attract new audiences. To accomplish this, radio has realized that it must utilize new media tactics. As such, radio has colonized different digital platforms, and is now available today on any computer, smartphone, and mobile device [9].

The Internet offers new options such as streaming established stations while simultaneously broadcasting on traditional radio wavelengths. In addition, users can listen to program fragments, accessing these segments or entire programs on demand – even long since archived content [7]

Traditional radio stations maximize the internet to widen their target audience, reach new listeners all over the world, and improve interaction quality and rates. Radio content is now creatively combined with varied forms of media – text, audio, and video [9]-[13]. It has been claimed that this fusion of audio, text, and video constitutes an innovative adaption of radio to new media, perhaps even generating a new model of radio [7] [9]. Even so, this convergence of media in radio represents more than supplementing auditory content (the foundation of traditional radio) with new features [9]. Radio stations add interactive content to their internet sites, including online archives, podcasts, and more [14]. A staple of online presence is a dedicated social media account – sharing platforms such as Facebook and Twitter are practically obligatory. Thus, along with traditional radio transmissions, new patterns and unprecedented modes of listener participation and interaction are developing. Cordeiro refers to this new radio model as r@dio – traditional radio that contains internet features and exploits a variety of online platforms [9].

Online radio challenges the received perceptions of its traditional counterpart, which were accepted for almost a hundred years. As noted, Cordeiro sees internet radio as a traditional medium that uses new platforms. However, other researchers question if it can still be referred to as "radio" [15]. Tacchi maintains that the transition from traditional to online transmission problematizes the definition of online radio. However, he notes that throughout history radio was transmitted by different technological means. Thus, the significance of radio is not derived from its transmission platform, but from techniques of disseminating content. Therefore, as long as radio maintains its features and consumers are using it in a consistent way, it can be perceived as the same medium – even if the mechanisms of transmission are different [16]. In contrast, Huong claims that online radio provides an updated definition of radio as a medium and technology. Both online and traditional radio stations share many activities and features. They provide a fusion of interactive and informative content. However, the former is more effective in generating user interaction to develop connections with listeners by means of unique online functions [17].

In sum, there is some scholarly consensus that internet radio does not necessarily constitute the creation of a new medium. Furthermore, radio stations are exploiting different Internet platforms to transmit traditional programs. That is, the Internet serves radio as a content distribution too. It provides related products and added value to the station and its listeners, sustaining a connection between the two that spans traditional medium and online space.

However, the internet today does more than complete the mission and function of traditional radio broadcasts. New media offers new unprecedented content options to radio listeners, precipitating a profound change in the structure of traditional radio and listening habits. As digital trends continue to accelerate, the radio industry is forced to continually reinvent itself to maintain its relevance in a multimedia web-based world [7]-[9][18]-[21]. About 40% of radio users now listen to their favorite programs on a digital platform such as an iPhone [22].

New media offers several significant benefits to radio stations. First, new media extend radio's listening range, eliminating the constraints of limited frequency reception ranges [23][24]. In this respect, the Internet complements and even replaces FM transmissions [9][21]. For example, in younger radio listener rankings, the preference of digital radio listening over terrestrial signal listening is significant: Digital radio listening is the most preferred method by younger radio listeners, "iPod use" is ranked second, and over-the air radio signals third [25].

This is in line with younger generations, who tend to be more active in digital media and on social networks, to the point of addiction, with a constant need to connect to technology [26].

In addition, the Internet maximizes the fact that radio is a mass media tool whose function is to provide a platform for public exchange of views. Habermas referred to the public sphere as a space where citizens meet to exchange opinions and ideas. In this free market of opinions, people engage in rational discussion on topics of political and social importance. For Habermas, the media is useful for producing the foundation of the public sphere [26]. The growth of commercial mass communication has, however, diminished the media's function as a tool of public debate and increased its role as a consumer product. As a result, media messaging no longer undergoes a potentially challenging process of deliberative discussion, and its messages are conveyed as matters of fact.

Digital radio can use new media technology to maximize its potential as a multi-platform, interactive broadcaster that offers its audience multiple means to react, respond, and provide content and critique [27]. Digital radio also enhances the accessibility of radio hosts, since listeners are more likely to encounter their personalities on its social media platforms. As a result, social network presence has become a crucial part of audience interaction strategies in radio stations [28]. Web 2.0 interactivity gives an immediate voice to an audience, one of the most important tasks of radio [28] [29].

Social networks thus enable development of a community of listeners related to radio activity [31], characterized by: feelings of membership (belonging to, identifying with); feelings of influence; integration and fulfillment of needs; feelings of being supported by others in the community while also supporting them; and shared emotional connection: relationships, shared history, and a "spirit of community".

Furthermore, Internet radio also adds a visual dimension to what has traditionally been a "blind" medium for listeners who could only listen to the broadcasters. Digital technologies have transformed radio into a medium both heard and seen, as social media visuals give audiences a real-time glimpse into broadcast operations and talent [32].

Even so, traditional radio stations are only making inroads on the Internet within their comfort zone. They use it as a platform for simultaneously broadening and sending their content. However, they generally do so without producing dedicated content for the new medium. As such, Internet content is practically speaking merely a direct reflection of traditional content, a sort of display window of station transmissions [33][34].

In Israel, however, most radio stations have websites that listeners can access in order to listen to station broadcasts and avail themselves of various content services [7]. One-third of Israelis use such digital radio services [20]. Slightly over 50% of all traditional radio stations have a mobile app [7], and 18% of Israelis consume radio contents on a station app [35].

The ascent of social networks as the main activity of many Internet users has also influenced mass media. Now it is understood that every form of media ought to have a presence on social networks. Interestingly, research conducted in recent years has pointed out that despite the great popularity of social media, the popularity level

of traditional communications channels is unrelated to number of followers or level of engagement between users and these channels [36][37].

Only a few studies have focused on the activity of traditional communications media on social networks. Nevertheless, social media offers many advantages such as free advertisement, global sharing of content, and more [37]. For example, Facebook has been shown to be one of the most popular means for people to access news sites [39][40], particularly among younger users [40]. Social networks especially dominate in the culture and entertainment sector, where most content is still consumed via traditional media. Many news sites have quickly assimilated the best qualities of social networks on their websites. In the near future, home pages of news websites may well no longer be relevant. Users will simply be exposed to curated and customized news and information through social networks and other technologies [40].

III. Radio on Social Media

As noted, television and radio use social media primarily to increase the popularity of their transmitted programs and promote them among listeners and viewers [8][38][41]. In studies on the activity of television stations on the social networks in the United States, it was found that Twitter was particularly used to raise and promote the content popularity. Public stations tended to engage in this publicity more heavily, also promoting their web site. However, commercial stations tended to favor tweets focusing on news [36][37][41]. In addition, television stations used Twitter more as a platform for calls to action (for example, eliciting user opinion or calling on users to watch a particular show). They engaged in this activity on Twitter more than sending out social tweets that convey feelings [41]. Relatedly, research on US television station social media activity on Pinterest showed a marked avoidance for sharing station content and promotional posts. Instead, TV stations increased sharing of images related to lifestyle such as food, design, clothing, and more [36].

As mentioned, the number of studies that have focused on the activity of radio stations on social networks is limited. In Spain, half of radio stations maintain an active Facebook page. Even so, they do not yet optimize the full advantages that social networks offer. A more detailed study on radio station activity on Facebook mostly looked at promotion of station products and isolated three types of major posts: 1) posts designed to create involvement, 2) posts designed to promote the station and listening to it, and 3) posts designed to promote the station broadcasters. The authors noticed that often questions and general comments were designed to produce a conversation with listeners (for example, “Who plans to watch the Opera show today?”). Yet these did not achieve the desired effect and generate meaningful feedback. On the other hand, content dealing with relevant artists gained higher number of “likes” and comments. These posts tended to more involve users, with the content unique from one station to another and from one program to another [38].

Therefore, a major conclusion is that there is as of yet no organized plan of action or grand strategy for radio stations on Facebook [41].

In Israel, it was found that the radio stations use their Facebook pages mostly to promote the products of the station. An analysis of activity on Israeli radio stations with active Facebook pages revealed that these stations experience difficulties in generating and sustaining high levels of listener engagement in the comments section, “likes,” and sharing [8].

Different kinds of posts on Twitter were found to be connected to the particular identity of the station. For example, music radio stations tweeted more self-promotional content, while non-music talk stations tweeted more news-related tweets. In general, radio stations with a focus on news tended to tweet more than their music counterparts [37].

Even though social networks are powerful communications tools, Facebook was found to not solidify stronger connections between listeners and radio broadcasters. Greater connectivity results from when broadcasters demonstrate clarity, resonate with, and are appreciated by listeners. Exposure on Facebook may actually cause a decrease in this sense of connection and erode the credibility of the broadcaster in the eyes of listeners [42].

As noted, social media managers of commercial pages on Facebook are tasked with driving up user interest in the product, brand, or company. As such, the question of which features make a post go viral is urgent and consequential [43]. Virality causes spikes in people sharing popular content to the extent that it even crosses social networks. This represents the focus of our research – testing the features of the most popular content on Israeli radio program Facebook pages. Some work has already analyzed the ingredients of a viral post. It was found that a significant predictor for initiating virality is the measure of positivity of the content [43]-[45]. Emotion thus constitutes an important factor in determining content popularity on social networks. In a study on viral tweets, it was found that a range of emotions – positive or negative – influences level of virality [41][46].

A connection was also drawn between physiological situations and virality: content that causes physiological arousal, characterized by stimulation of the autonomic nervous system and leading to bodily state of readiness for action such as anger, anxiety, and revulsion, tends to increase the likelihood of content going viral – even if negative. That is, content that provokes negative emotions that are not physiologically stimulating such as sadness will be less viral [43]-[45].

In the same way, we find that certain conversational elements of Israeli radio station Facebook pages lead to higher engagement. For example, video segments and images were shown to produce more comments and “likes” as compared to exclusively textual posts. However, posts with some text garner more comments than other kinds of posts [8].

Radio stations understand that to stay relevant, they must leverage and optimize social networks. As noted, this understanding is only partial for stations continue to use these platforms mainly for promoting programs and broadcasters. Yet, as mentioned, this promotion can degrade the credibility bond between broadcaster and audience. Social media emphasis should be on encouraging user engagement with the station and its online contents. This research shows that the radio industry lacks a guiding social media strategy. Instead, it operates through trial and error, although tactics do exist. Advertising positive contents and content generating physiologically stimulating emotions should be key parts of its social media strategy.

IV. Technologic Determinism

The theory of technologic determinism sees technological developments as determining the conduct of society. Accordingly, technology is not only a tool for transmitting messages, but, as McLuhan famously claimed, is the message itself. Technologies are the ones that influence the social and historical processes of the world [1]. In his words: “We are not paying attention to the fact that the medium is really the influencing factor, and not the content […] ‘content’ of every medium is another medium. Talking is the content of writing, as much as written word is the content of printing and printing is the content of the telegraph” [47].

Technologies create new situations and thus change society and its modes of thinking. Any medium functions in such a way as to widen one of the senses. Each use of another medium affects the senses and its inter-dynamics. Thus, a person who uses the medium will have her

perception changed by the particular medium. For McLuhan, every historical period must be approached through its dominant medium and its various effects on the sensory system [1]. For example, in a literate society the book is usually the dominant medium. However, in a colloquial society, it is often radio that is the most important. A colloquial society will also favor the Internet as a dominant medium. Technologic determinism emphasizes that there is no substantial meaning to the essence of content transmitted on the Internet – only the fact that it was transmitted through the Internet. Therefore, if a certain content is transmitted both on FM radio and the Internet – such that the content is the same, but the medium is different – theoretical implications follow from a technologic deterministic standpoint.

V. Research Hypotheses

As noted, this research focuses on social media patterns of traditional Israeli radio stations, with emphasis on Facebook as the dominant platform. This study looks at popular content dedicated to the creation of interaction and user engagement. These activities can be assessed by closely tracking engagement measures such as comments, "likes," and sharing.

The research hypotheses are based on how radio programs influence consumers to be involved in Facebook post content. Thus, consistent with the notion that viral posts contain informative-rich elements of visibility, informality, and emotionality, resulting in popularity and high user engagement, the following can be hypothesized:

Hypothesis 1: As combinations of media (video and images photos) add to the listener's experience and lead to higher engagement, it can be posited that posts which include various media elements will be more popular than other posts that do not contain a combination of media elements.

Hypothesis 2: Since user visits to Facebook are particularly motivated by entertainment and information, it can be posited that informative posts and/or posts that combine celebrity news and images will garner higher popularity than content lacking such elements.

Hypothesis 3: As radio programs try to engage listeners in activities and transmit content in different ways, interactivity is a central component in the strategy of creating the perception of a directly experienced relationship. Therefore, we claim that posts that include a call for on/offline action and/or posts that contain greetings will be more popular.

Hypothesis 4: Based on research demonstrating that virality is connected to user emotions, we posit the posts that integrate emotional and informal language will gain a higher measure of popularity and user engagement.

VI. Methodology

This current study examines Facebook pages of Israeli radio programs. Data was collected between January and May of 2015. At the time of compiling the data, 145 Israeli radio programs had an active Facebook page. During the study's first stage, the two most popular Facebook pages from each station's latest posts were scraped (two programs with the highest number of "likes"). Then, 1,000 posts from each program's Facebook page were scraped. Data was collected via 'Netvizz', an application which allows scholars to extract posts from official Facebook pages along with relevant data, e.g. number of 'Likes', comments and shares per-post, author-identity (i.e. user or station), post format (i.e. image, video, link, or textual status update) and the post-text.[1]

For the next stage, and in order to offer an in-depth analysis of the most popular posts on the programs' Facebook pages, the leading posts for each week were sampled out of each program's page according to the following four indexes – most-'Liked' post; most commented-on post; most-shared post; and lastly, the post whose user-comments received the highest number of 'Likes' for that week. And so, per each station page, up to four weekly posts were sampled (in some instances where a certain post was at the top of several indexes for that week, it was only sampled once). All in all, a total of 876 posts were sampled and manually coded.

Coding was done according to an elaborate coding book which included the following categories:

Automatic categories – information received via Netvizz from Facebook's API and saved automatically by the tool:

- Post format: the format of the post. Could be audio, video, image-based, a status update or a hyperlink.
- Post-author identity: Is the author a user or the page administrator, representing the station?
- Number of 'Likes' the post received.
- Number of user-comments on the post.
- Number of times the post was shared.

The following categories relate to the post's text (in the cases of image/video-based posts, the text accompanying the post was analyzed) and were manually coded.

- **Emotionality**: Did the post contain strictly rational or emotional language or did elements of both recur? Emotional elements could come in the form of interjections such as "so fun!" or "must share!", as well as subjective descriptions and figures-of-speech betraying the author's sentiments. Rational elements offer an objective account of facts or certain messages in a manner that neither attests to the author's emotional state nor reflects on the post's theme.
- **Formality**: Is the text's tone formal or informal? A text may be deemed 'formal' if phrased officially and grammatically, e.g. "Do write to us and you may enter the draw." Informal text, on the other hand, will use colloquial language or "web-speak", as well as humorous elements or syntax more typical of spoken language, e.g. "We wanna pamper ya'll" or "Cold day innit?"
- **Information**: Does the text feature information of any sort on certain events and news items, as well as station-related information such as a program's timeslot? A binary yes/no category. An example of an informative text can be, "Did you know that today is the birthday of the first transgender singer?" but also, "On today's show, so-and-so will tell us exactly what puts a spring in his step every morning".
- **Call-to-action**: Online – Is the text attempting to elicit some online action on Facebook, e.g., 'Liking', "Take pictures and send to us", etc. A binary category.
- **Call-to-action**: Offline – Is the text attempting to elicit some offline action off Facebook, e.g., "Listen on frequency X", "Call the number Y", etc. A binary category.
- **Object of text**: A multiple-choice category. Is the text about a celebrity, a non-public person, an in-house station event or an event unrelated to the station?

Content coding was carried out by two coders, who serve as research assistants at the School of Communications and have taken courses related to radio and its role and on social and new media networks, as well as received training in quantitative and qualitative research methods.

The coding process began only after completing their training

1 The author would like to thank Digital Methods Initiative and app-developer Bernhard Rieder for their permission to use the application.

during which all coding categories were thoroughly explained and illustrated to each coder. All the posts taken from one of the sampled radio programs, which make up 10 percent of the sum total of posts in the study, were analyzed by the two coders as part of a reliability test.

VII. Research Findings

Every popular post garnered on average 232 "Likes" and engagement of 377. It should be noted that for this data, high standard deviations were found (823 and 1632, respectively).

A. Type of Media and Engagement

The first research hypothesis claimed differences in "Likes" and engagement between different varieties of posts: media (image and video), status (wording), link, survey, and audio. Therefore, two one-way analyses of variance were conducted.

TABLE I. Means and Standard Deviations of "Likes" Quantity and Engagement

Media Type	Engagement		Likes Quantity	
	Standard Deviation	Mean	Standard Deviation	Mean
Video	3229.54	625.35	1528.48	390.67
Picture	923.90	416.55	600.38	285.83
Link	149.13	138.50	90.20	84.463
Status	232.41	165.17	39.15	41.14
Audio	12.88	12.00	11.09	10.50

Table I shows the sample means and standard deviations of each category, and we can see there are differences between the different categories. Posts that included video received the highest average quantity of "Likes" and Engagements, followed by post that included a picture. Next, but separated by a major gap, follow posts that included links (for "Likes") and statuses (for Engagement), and last were posts that included audio, which received the lowest average quantity of "Likes" and Engagements.

In order to check whether these differences are significant, we used a one-way variance analysis (anova), since this kind of statistical analyses test the hypothesis that the true means of the categories are different (when the dependent variable is continuous, of course).

The first one-way anova was carried out with a dependent variable of measure of "Likes" and the independent variable of Media type. As $F(4,870)=5.67$, $p=0.0001$, which is less than the standard thresholds for the F test- 0.05, we can conclude that a there is a significant difference in "Likes" between different media types of posts.

Then, a Post hoc analysis using Scheffe's test was conducted. Post hoc analysis of that kind is used in order to find the source of the significant difference that was revealed in the one-way anova analysis (i.e. to find the sub-group of categories that their true means are different from each other). The post hoc analysis found that posts that included a video or an image received significantly more "likes" with p.value=0.001 and p.value=0.015 accordingly, both below the standard thresholds for this test- 0.05). Thus, we can conclude that posts that included a video or an image are significantly more popular than statuses. All other Scheffe's test differences produced p.values that were higher than 0.05, thus we concluded that all other differences between different post types are not significant.

In addition, a one-way anova was conducted with the dependent variable of engagement (and the same independent variable as before). It has revealed a significant difference between the different categories. Thus, this research hypothesis was reinforced ($F(4,870)=2.46$, $p=0.04$, which is less than the standard thresholds for the F test- 0.05).

The post hoc analysis found that posts that included a video received more engagement, and as p.value=0.097, which is higher than the standard thresholds for this test (0.05), we can conclude that posts that included a video are borderline significantly more popular than statuses. All other Scheffe's test differences produced p.values that were higher than 0.05, thus we concluded that all other differences between different post types are not significant.

In conclusion, the results shown during this section, in table I and in particular in the significant results from the statistical test that were carried, support the claim that the media type of a post affects its user "Likes" and engagement. Specifically, posts that include a video generate higher user "Likes" and engagement.

B. Information and Engagement

The second research hypothesis posited differences in "Likes," engagement, and sharing of posts based on presence of information and celebrity media in posts.

TABLE II. Means and Standard Deviations of "Likes" Quantity and Engagement

Information and Celebrity Media	Engagement		Likes Quantity	
	Mean	Standard Deviation	Mean	Standard Deviation
Famous figure in the picture	675.10	1268.29	439.82	728.89
No picture	344.17	2067.91	187.80	985.08
No famous figure in the picture	281.32	639.91	202.65	498.312

First, we tested for differences in "Likes" quantity and engagement between posts with and without images, and between images with and without celebrity content. Table II shows the sample means and standard deviations of each category, and we can see there are differences between the different categories. Posts that included a famous figure in a picture received the highest average quantity of "Likes" and Engagements. Surprisingly, posts that included no picture at all received the second high average quantity of Engagements, higher than posts that included an image, but without a famous figure. Nevertheless, posts that included an image, but without a famous figure received a higher average quantity of "Likes" compared to posts that included no picture at all.

In order to check whether these differences are significant, we used a one-way variance analysis (anova). The first one-way anova was carried out with a dependent variable of measure of "Likes" and the independent variable of information and celebrity image category. As $F(2,873)=5.34$, p.value=0.005, which is less than the standard threshold for this test- 0.05, we can deduce that there is a significant difference in "Likes" between the different categories.

A Post hoc analysis, using Sheffe's test, confirmed that the average quantity of "Likes" was significantly higher when posts contained celebrity media – more than posts without images and posts with images that contained no celebrity content (p.value=0.006 and p.value=0.021 accordingly, both below the standard thresholds for this test).

In addition, a second one-way anova was conducted with the dependent variable of engagement (and the same independent variable as before). It was found that there is a borderline significant difference in engagement among posts containing celebrity images ($F(4,873)=2.94$, p.value=0.02, which is again lower than the standard thresholds for this test).

A Post hoc analysis, using Scheffe's test, confirmed that the

average quantity of Engagement was borderline significantly higher when posts that contained celebrity media, compared to posts that did not contain celebrity content (p.value=0.064. which is very close to the standard thresholds for this test- 0.05). However, the difference between posts that contained celebrity media and posts that did not include a picture was not significant (p.value=.107, which is above the standard thresholds for this test).

TABLE III. Means and Standard Deviations of "Likes" Quantity and Engagement

Post Type	Engagement		Likes Quantity	
	Standard Deviation	Mean	Standard Deviation	Mean
Famous figure in a video	6236.45	2019.21	2880.62	1231.06
No video	737.84	310.45	477.54	188.85
No famous figure in a video	222.01	155.03	184.48	114.43

Second, we tested for differences in "Likes" quantity and engagement between posts with and without celebrity video content. Table III shows the sample means and standard deviations of each category, and we can see there are differences between the different categories. Posts that included a famous figure in a video received the highest average number of "Likes" and Engagements, followed, interestingly, by posts with no video at all, and last were posts that included a video, but without a famous figure.

In order to check whether these differences are significant, a one-way variance analysis (anova) was carried out with a dependent variable of measure of "Likes" and the independent variable of information and celebrity video category. It has revealed a significant difference in "Likes" between the different categories, as F(2,873)=40.29, p.value=1*(10^-16), which is way less than the standard thresholds for this test. Thus, the research hypothesis was reinforced

A Post hoc analysis, using Scheffe's test, produced a very low p.value (less than 0.0001), and thus confirmed that the average quantity of "Likes" was significantly higher for posts the contained celebrity videos – more than posts without videos, or with video, but no celebrity content. In addition, a second one-way anova was conducted with the dependent variable of engagement (and the same independent variable as before). As F(2,873)=27.50, p.value=1*(10^-12), which is way less than the standard thresholds for this test, we can deduce that there is a significant difference in engagement among posts containing celebrity images.

A Post hoc analysis, using Scheffe's test, produces a very low p.value (again, less than 0.0001), and thus confirmed that the average quantity of Engagement was significantly higher for posts the contained celebrity videos – more than posts without videos, or with video, but no celebrity content.

In conclusion, the results shown during this section, in tables II-III and in particular in the significant results from the statistical test that were carried, support the claim that the inclusion of a famous figure within the media (video or picture) of a post affects its user "Likes" and engagement. Specifically, posts with media (video or picture) that includes a celebrity generates higher user "Likes" and engagement.

C. Calls for Action and Integration of Greetings

The third research hypothesis claims that posts that include a call for online and/or offline action or greetings will be more popular.

TABLE IV. Means and Standard Deviations of Quantity of "Likes" and Engagement in Posts that Integrate Greetings

Post Type	Engagement		Likes Quantity	
	Standard Deviation	Mean	Standard Deviation	Mean
Greetings	2514.99	580.77	1193.20	337.26
No greetings	903.55	277.87	554.48	181.88

We tested for differences in "Likes" quantity and Engagement between posts with and without greetings content. Table VI shows that there is a difference in "Likes" and Engagement between the two categories, posts that included greetings received, in average, more "Likes" and engagement than posts without greetings.

In order to check whether this difference is significant, a t-test for independent samples analysis, which is the equivalent of the one-way anova for 2 categories, was carried. The dependent variable was the measure of "Likes" and the independent variable was the type of the post, with/without greetings (which are 2 independent groups). The t-test analysis produced a test statistic of t(347.71)=2.1, and p.value=0.02, which is less than the standard threshold for t-test- 0.05. Thus, we concluded that there is a significant difference in average quantity of "Likes" between the two categories. Thus, "Likes" quantity of greeting posts was shown to be significantly higher.

In addition, another t-test for independent samples analysis was conducted with the dependent variable quantity of engagement (and the same independent variable as before).Again, the t-test analysis produced a test statistic of t(322.48)=1.98, and p.value=0.02, which is less than the standard threshold for this test. Thus, we concluded that there is a significant difference in average quantity of engagement between greeting and non-greeting posts. Thus, this part of the research hypothesis was confirmed.

In conclusion, the results shown during this section, in table IV and in particular in the significant results from the statistical test that were carried, support the claim that integrating greeting in a post affects its user "Likes" and engagement. Specifically, it was found that radio stations posts that integrate greetings manage to engage more listeners and produce more "Likes".

D. Formal and Rational Language

The fourth research hypothesis maintains that differences will be found in "Likes" and engagement between formal and informal posts. In addition, our claim is that differences in posts will be measurable based on use of emotional or rational language. To test this hypothesis, a one-way analysis of variance was conducted.

TABLE V. Means and Standard Deviations of "Likes" Quantity and Engagement.

Formal/Informal Posts	Engagement		Likes quantity	
	Standard Deviation	Mean	Standard Deviation	Mean
Formality	563.54	249.45	474.91	175.02
Informal and rational	642.80	300.22	425.55	190.75
Informal and emotional	3385.10	822.82	1567.12	433.21

First, we tested for differences in "Likes" quantity and engagement between formal and informal posts. Table V shows that there are differences between the different categories. Informal and emotional posts received the highest average number of "Likes" and Engagements, followed by informal and rational posts, and last were formal posts.

In order to check whether these differences are significant, we used a one-way anova. The first one-way anova was carried out with a dependent variable of measure of "Likes" and the independent variable of post formality. The test produces a test statistic of F(3,868)=4.47, and p.value=0.004, which is less than the standard threshold for this test. Thus, we concluded that there is a significant difference in "Likes" between the different categories.

A Post hoc analysis using Scheffe's test revealed that the average quantity of "Likes" for informal emotional posts was significantly higher compared to the average quantity of "Likes" for non-emotional formal and informal posts(p.value=0.005 and p.value=0.06, which are below and very close to the standard threshold for this test, accordingly).

In addition, a second one-way anova was conducted with the dependent variable of engagement (and the same independent variable as before). The test produces a test statistic of F(3,868)=5.69, and p.value=0.001, which is less than the standard threshold for this test. Thus, we concluded that there is is a significant difference in engagement between the different categories. A Post hoc analysis, using Scheffe's test, confirmed that the average quantity of Engagement for informal emotional posts was significantly higher compared to the average quantity of Engagement for non-emotional formal and informal posts (p.value=0.01 and p.value=0.03, accordingly, which are below the standard threshold for this test). In conclusion, the results shown during this section, in table V and in particular in the significant results from the statistical tests that were carried, support the claim that the formulation of a post (formally/informally and emotionally/unemotionally) affects its user "Likes" and engagement. Specifically, it was found that a post formulated informally and emotionally receives more "Likes" and engagement compared to other kind of formulations.

VIII. Discussion and Conclusions

Radio stations are active on social media and still adjusting to the realities of the Internet [8]. This research looks at the success of radio programs on Facebook as primarily measured through user engagement. We show how radio programs use technological tools to expand their capabilities, producing more developed interaction with listeners and transmission of information in a variety of modes. Moreover, radio programs increase social networking. That is, the medium (i.e. the platform) is critical in order to realize their goal of maximizing popularity with consumers.

In this way, social networks enable radio to maintain an online community with all its characteristics [30, 48], once again a powerful platform for development and exchange of opinions and public involvement in public discourse [27].

It was found that integration of media (video or image) will generate more popular – even viral – content. In addition, audio posts proved less popular, even though 30 % of analyzed programs were music oriented. These findings correlate with research on radio stations on Facebook, which found visual posts are the most popular form of content [8]. As such, a blurring of media boundaries can be seen, and this enhances possibilities of user engagement.

As noted, radio is characterized as a "blind" medium due to its exclusive auditory focus [5][6]. However, radio has increasingly leveraged Facebook, which exploits visual resources such as text, image, and video.

It was found that radio station posts integrating celebrities in image or video garner higher rates of popularity. These findings match the notion that celebrity posts represent a salient factor in driving up content popularity, as found in previous research [8]. Even programs based on raising awareness and engagement can improve their message dissemination by making some reference to celebrity, especially when linked in some direct or indirect way to the program. The audience appetite for this type of content only gains in popularity through such forms of association.

The finding support previous research, which found that Radio Facebook pages with formal posts tend to exploit calls for action [8][41], and these are successful in generating substantial listener reactions, as was found in the current study [8].

Radio has always been perceived as a medium that personally appeals to its listeners [1]. For instance, most programs contain personalized broadcaster greetings and exhortations such as "good morning," "drive carefully," "enjoy the songs," "thanks to the listeners," and "have a good week!" This convention has proved successful in the past for radio, and has been translated onto Facebook posts. Even more, in the FM-medium, no capability exists for easy audience responses. But Facebook is based on instant messaging, and so radio is suddenly turned into a two-way channel of communication.

As noted, previous research has pointed out that emotional messaging leads to higher engagement and virality of posts [8][41][46]. This study found that posts integrating emotional, informal, and supportive appeals achieved an extremely high quantity of "Likes" and user engagement. Facebook in general is more characterized by a culture of informal activity such that radio programs are intentionally channeling this cultural preference into their messaging.

Radio programs are thus adjusting to the digital world, in an attempt to access the full potential of new platforms and tools. They use social networks to raise levels of popularity, online as well as offline. Using Facebook, traditional radio has broadened its reach and appeal in music and talk-based programs. This is also comprises an opportunity to reconnect with a younger audience that has long been active on social networks [26]

By developing content channels on Facebook, radio programs are using novel strategies for promoting and integrating audience reaction and listener interaction. This activity is continuous and ongoing – 24 hours, seven days a week. This messaging is not constrained in time like programs and live content on FM. The latter are restricted to specific scheduling hours or the time of live transmission. More importantly, radio programs are now utilizing visual elements to promote content and popularize the station/program – a relatively unprecedented development for the medium. Radio is thus merging with the Internet, especially with social networks such as Facebook. Therefore, radio professionals who are also media specialists need to recognize this convergence. They then need to adapt new media strategies in more fluent, consistent, and professional ways. As noted, critical elements in any coherent strategy should include: 1) integration of media types in all social media; 2) maximum reference to celebrity culture in linked content; and 3) supportive messaging characterized by personal, informal, and emotional wording.

However, as a dominant player in social media, Facebook may sometimes appears to be on the verge of coopting and assimilating traditional radio. As such, radio's unique character may end up being threatened by such a powerful iteration of new media. Users may start forgetting radio's special qualities: 1) exclusive auditory appeal (a single sense) and 2) usefulness in secondary tasks [3][4][6].

In conclusion, this research continues to confirm McLuhan's powerful medium-is-the-message claim [1]. Radio has successfully adjusted to new media, even if it has yet to maximize the latter's full potential. Radio should be seen as creatively adapting itself to suit its activity for Internet ecosystems, especially for social networks.

References

[1] M. McLuhan. "Understanding media: The Extension of man". New York: McGraw Hill, 1964.

[2] E. Pease and E. E. Dennis, “Radio, The Forgotten Medium. New York: Transaction Publishers”, 1993.
[3] O. Soffer, “Mass communication in Israel”. Raanana: Open University. [in Hebrew], 2011
[4] D. McQuail, “Mass communication theory: An introduction”. Sage Publications, Inc, 2005.
[5] A. Crisell. “Understanding Radio”. London: Routledge. (2nd edition), 1994.
[6] M. McLuhan, M. “The medium is the message”. In: d. Caspi (ed.), Mass Communication, Reader (pp. 26-35). Tel Aviv: The Open University McLeish, R. (1978). The Technique of Radio Production. London: Focal Press, 1995.
[7] T. Laor, Y. Galily, and I. Tamir, “Radio presence in online platforms in Israel”. Israel Affairs vol. 23, no.5, pp. 951-969, 2017.
[8] T. Laor, and N. Steinfeld “From FM to FB: radio stations on Facebook”. Israel Affairs, vol. 24, no. 2, pp. 265-284, 2018.
[9] P. Cordeiro, “From Radio to R@dio: broadcasting in the 21st century In Radio Evolution”: In M. Oliveria., P. Portela. & L. A. Santos (Ed.), Radio Evolution Conference Proceedings (pp. 155-166). Braga: University of Minho, 2012.
[10] H. A. Chiumbu and D. Ligaga, “Communities of strangerhoods?: Internet, mobile phones and the changing nature of radio cultures in South Africa”. Telematics and Information, vol. 30, pp. 242-251, 2013.
[11] C. Greer and T. Phipps, “Noncommercial religious radio stations and the Web”. Journal of Radio Studies, vol. 10, no. 1, pp. 17-32, 2003.
[12] R. A. Lind and N. J. Medoff, “Radio stations and the world wide web”. Journal of Radio Studies, vol. 6, no. 2, pp. 203-221,1999.
[13] J. M. Te, S. M. Asbir, R. L. De la Cruz, N. M. Jabel, K. Refugido and D. Marcial. «Developing E-Radio: An Online Audio Streaming Application”. Parallel and Distributed Processing with Applications Workshops (ISPAW), 2011 Ninth IEEE International Symposium on, pp. 322-327, 2011.
[14] T. Samuel-Azran, T. Laor and D. Tal. “Who listens to podcasts, and why?: the Israeli case”. Online Information Review, 2019.
[15] D. T. T. Huong, “Radio and its listenership in the internet age: case studies of the Voice of Vietnam (VOV) and VOVNews”. United Kingdom: Bournemouth University, 2008.
[16] J. Tacchi, “The need for radio theory in the digital age”. International Journal Of Cultural Studies, vol. 3 no. 2, pp. 289-298, 2010.
[17] W. Ren and S. M. Chan-Olmsted, “Radio content on the World Wide Web: Comparing streaming radio stations in the United States”. Journal of Radio Studies, vol. 11 no. 1, pp. 6-25, 2004.
[18] J. Farrell, “A review of multimedia formats and social media use for traditional radio broadcasting in Ireland”, 2006. Retrieved 29 April, 2018, from: https://scss.tcd.ie/publications/theses/diss/2016/TCD-SCSS-DISSERTATION-2016-029.pdf.
[19] T. Laor, Y. Galily and S. Lissitsa, “Online digital Radion apps usages in Israel: Consumers, consumption and meaning”. Technology in Society, 2019.
[20] M. Moshe, T. Laor and S. Friedkin, “‘Digital soap opera’ online radio listening patterns and the digital divide”. Israel Affairs, vol. 23 no. 2, pp. 361-384, 2017.
[21] G. Starkey, “Radio: The resilient medium in today’s increasingly diverse multiplatform media environment”, Convergence, vol. 23, no. 6, pp. 660-670, 2017.
[22] I. Punnett, “Digital is the future—and the now: EmPosium on digital platform listening trends”, 2016.
[23] N. Ignatiew, “Music radio stations from the “On Air” to the Online: Identifying media logics in the content and formats of Radio FIP on its digital platforms”. M. A. thesis, Uppsala University, 2017.
[24] S. McClung, “College Radio Station Web Sites: Perceptions of Value and Use”. Journalism and Mass Communication Educator, vol. 56 no. 1, pp. 62-73, 2011.
[25] M. J. Saffran, “Effects of local-market radio ownership concentration on radio localism, the public interest, and listener opinions and use of local radio”, Journal of Radio & Audio Media, vol. 18, pp. 281–294, 2011.
[26] J. Cabral, “Is generation Y addicted to social media”. Future of children, vol. 18, no. 25, 2008.
[27] J. Habermas, “The structural transformation of the public sphere”. Cambridge: Polity Press, pp. 114-120, 1989.
[28] M. Sinton, “No Longer One-to-Many: How Web 2.0 Interactivity is Changing Public Service Radio’s Relationship with its Audience”. Journal of Radio & Audio Media, vol. 25 no. 1, pp. 62-76, 2018.
[29] S. N. Mohammed and A. Thombre, “An Investigation of User Comments on Facebook Pages of Trinidad and Tobago’s Indian Music Format Radio Stations”. Journal of Radio & Audio Media, vol. 24 no. 1, pp. 111-129, 2017.
[30] F. G. F. E. Silva, J. Colussi and P. M. Rocha, “ WhatsApp as a Tool for Participation on Spanish Radio: A Preliminary Study of the Program Las Mañanas on RNE”. Journal of Radio & Audio Media, vol. 25 no. 1, pp. 77-91, 2018.
[31] K. Ouzts, “Sense of community in online courses”. Quarterly Review of Distance Education, vol. 7 no. 3, 2006.
[32] D. A. Ferguson. and C. F. Greer, “Visualizing a non-visual medium through social media: The semiotics of radio station posts on Instagram”. Journal of Radio & Audio Media, vol. 25 no. 1, pp. 126-141, 2018.
[33] A. Moody, J. Greer and T. Linn, “Public radio station Web sites and their users”. Journal of Radio Studies, vol 10 no. 2, pp. 255-261, 2003.
[34] M. Seelig, “Interactivity on traditional media web sites”. Journal of Radio & Audio Media, vol. 15, no. 2, pp. 231-248, 2008.
[35] Bezeq Survey, “Life at digital age: Bezeq Internet report 2017”. Retrieved 22 December 2018 from: https://www.bezeq.co.il/media/PDF/internetreport_2017.pdf
[36] D. A. Ferguson and, C. F. Greer, “Local radio and microblogging: How radio stations in the US are using Twitter”. Journal of Radio & Audio Media, vol. 18, no. 1, pp. 33-46, 2011.
[37] D. A. Ferguson and C. F. Greer, “Pinning and promotion: How local television stations are using Pinterest for branding and audience connectivity”, 2011. Retrieved from: http://www.academia.edu/2018214
[38] B. C. Freeman, J. Klapczynski, and E. Wood, “Radio and Facebook: The relationship between broadcast and social media software in the US, Germany, and Singapore”. First Monday, vol. 17, no. 4, 2012. Retrieved from: http://www.firstmonday.dk/ojs/index.php/fm/article/view/3768/3194
[39] Pew research center, “Social, Search & Direct: Pathway to digital news”, 2014.
[40] J. Wihbey, “The Challenges of Democratizing News and Information: Examining Data on Social Media, Viral Patterns and Digital Influence”. Viral Patterns and Digital Influence (June 9, 2014). Retrieved from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2466058
[41] J. S. Lin and J. Peña, “Are you following me? A content analysis of TV networks’ brand communication on Twitter”. Journal of Interactive Advertising, vol. 12, no. 1, pp. 7-29, 2011.
[42] M. E. Savage and P. R. Spence, “Will you listen? An examination of parasocial interaction and credibility in radio”. Journal of Radio & Audio Media, vol. 21 no. 1, pp. 3-19, 2014.
[43] J. Berger, Contagious: Why things catch on. Simon and Schuster, 2016.
[44] J. Berger, K. L. Milkman,” What makes online content viral?”. Journal of Marketing Research, vol. 49, no. 2, pp. 192-205, 2012.
[45] R. E. Guadagno, D. M. Rempala, S. Murphy and B. M. Okdie, “What makes a video go viral? An analysis of emotional contagion and Internet memes”. Computers in Human Behavior, vol. 29 no. 6, pp. 2312-2319, 2013.
[46] S. Stieglitz and L. Dang-Xuan, “Emotions and Information Diffusion in Social Media-Sentiment of Microblogs and Sharing Behavior”. Journal of Management Information Systems, vol. 29 no. 4, pp. 217-248, 2013.
[47] M. McLuhan. “Understanding media: The Extension of man”. New York: McGraw Hill, 1964, pp. 26.
[48] F. De Rada, A. Mochón Sáez, and Y. Sáez Achaerandio. “TV Series and Social Media: Powerful Engagement Factors in Mobile Video Games.” International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5. no. 3, pp. 46-55, 2018.

Dr. Tal Laor, PhD.

Lecturer at Ariel University, Israel
Researcher in the media field
Head of the broadcasting teaching program

Design and Validation of a Framework for the Creation of User Experience Questionnaires

Martin Schrepp[1], Jörg Thomaschewski[2]*

[1] SAP AG (Germany)
[2] University of Applied Sciences Emden/Leer (Germany)

Received 10 May 2019 | Accepted 26 June 2019 | Published 28 June 2019

Abstract

Existing user experience questionnaires have a fixed number of scales. Each of these scales measures a distinct aspect of user experience. These questionnaires can be used with little effort and provide a number of useful support materials that make the application of such a questionnaire quite easy. However, in practical evaluation scenarios it can happen that none of the existing questionnaires contains all scales necessary to answer the research question. It is of course possible to combine several UX questionnaires in such cases, but due to the variations of item formats this is also not an optimal solution. In this paper, we describe the development and first validation studies of a modular framework that allows the creation of user experience questionnaires that fit perfectly to a given research question. The framework contains several scales that measure different UX aspects. These scales can be combined to cover the relevant research questions.

Keywords

User Experience, Questionnaires, Evaluation.

DOI: 10.9781/ijimai.2019.06.006

I. Introduction

There are several methods available that make it possible to evaluate the user experience (UX) of a product. Well-known approaches are usability tests [1], expert reviews (for example heuristic evaluations [2]-[4] or cognitive walkthrough techniques [5] [6]) and UX questionnaires [7].

Usability testing and expert reviews create mainly qualitative data, while questionnaires create mainly quantitative data. In addition, due to the required effort involved in usability testing and expert reviews, these allow only to collect data from small target groups of users (typical usability tests are in the range of 10 to 15 participants) or experts (here we have rarely more than 3-5 experts involved in one evaluation).

Questionnaires on the other hand are a simple and inexpensive method (especially if they are used as online questionnaires) that allow the collection of data from larger groups of users of a product. This is especially important since the impression of a person concerning the UX of a product is highly subjective. Depending on the subject's personality or previous experiences with the product or even similar products, the opinion of two users concerning the UX of a product can be quite different.

Assume, for example, that a new product is introduced in a company to unify certain business processes over different locations. Users in one location may have a quite difficult time to learn the new product, since concepts and handling are very different to the tools they used before the change. Users of a second location may have in contrast the impression that the new product is easy to learn or even intuitive to use, since the tools they used before were similar to the new product concerning work flows and general user interface concepts.

Concerning hedonic [8] UX aspects, for example the aesthetic impression, the users' personal tastes or preferences also play a role. One user may find the visual design of a product appealing and interesting, while another finds it boring and unimpressive. In such situations it is quite important to collect data from a large number of individuals to get a clear and reliable view on how well the design of the product fits to the needs of the target group.

Questionnaires allow to collect data from larger samples of users and transfer the subjective impressions of those users into a numerical scale value that describes how the corresponding UX quality of the product is perceived inside the target group.

This ability to measure the user experience of a product quantitatively is quite important for several typical questions in product evaluations [9]. First, it allows a direct comparison of different products or different design variants of a single product concerning their UX. Second, it can be used to continuously monitor the UX quality of a product over time. Third, it allows setting objective goals concerning UX by defining a threshold for the mean values of the scales of the questionnaire, which should be reached over time. Fourth, the comparison of the evaluation results of a product with a benchmark allows deciding if the UX quality of the product fulfils general user expectations [10].

User experience is a complex product characteristic [11] that results from the perception of many distinct quality aspects. These are classical task-related UX qualities, for example, efficiency of use, ease of learning, controllability, error tolerance [12], intuitive use [13], visual complexity [14], usefulness [15], or non-task related UX aspects like, fun of use [16], identity [17], aesthetics of the visual design [18], novelty of the product concept [19] or content quality [20].

However, not all of these UX aspects are of relevance for every single product [21] [22]. The importance of such UX aspects can vary widely between products supporting different tasks and use cases.

* Corresponding author.
E-mail address: joerg.thomaschewski@hs-emden-leer.de

For a simple self-service application, e.g. creating a leave request or an application to change personal data of an employee, it is crucial that it can be used intuitively, i.e. without asking for help of another person or reading a lengthy manual. Such applications are used quite infrequently, and we cannot expect that the user will remember how to use the application between two usage points. Because of the rare usage frequency, efficiency does not play a role here. An unnecessary click does not hurt much if an application is used only once in a month or even less frequently.

For a business application, for example an application to create sales orders or service requests, that is used repeatedly during a typical work day, things are completely different. Intuitive use is nice to have, but not crucial. Typically, a learning period is required for such applications to understand the use case and the mapping of real-world processes and tasks to the elements and flows of the application. Therefore, some learning is acceptable and expected by users. In addition, due to the heavy usage during a typical work day, efficiency is key for these types of applications, i.e. an unnecessary click really hurts, when it needs to be repeated 50 times a day.

The huge number of existing UX aspects and the different levels of importance for different types of products explains the high number of different UX questionnaires that are available, for example SUS [23], SUMI [24], UEQ [19], VISAWI [25], meCUE [26] or ISOMETRICS [27], just to name a few. Each of these questionnaires realizes by its scales a different set of measured UX aspects. For example, SUS only measures overall usability and the items in this questionnaire address mainly *Learnability* and *Efficiency* of use. VISAWI measures the visual appeal of a product by 4 subscales (*Simplicity*, *Diversity*, *Colorfulness*, *Craftsmanship*). The UEQ measures 6 distinct UX aspects (*Attractiveness*, *Efficiency*, *Perspicuity*, *Dependability*, *Stimulation*, *Novelty*). The ISOMETRICS contains the quality aspects described in the ISO 9241 – 210 as scales. Thus, what is actually measured differs heavily between different UX questionnaires.

Of course, none of these questionnaires contains all UX aspects discussed in research literature, since this would increase the length of the questionnaire above any reasonable limit.

For a UX researcher evaluating a concrete product, this can cause some problems. If he or she has narrowed down which UX aspects are important for the users of the concrete product and should be thus measured in the evaluation, it can easily happen that no one UX questionnaire exists that contains exactly those UX aspects as scales. Sometimes, it is possible to combine several UX questionnaires to cover all relevant aspects, but usually this is also not optimal, since different questionnaires often have different item and answer formats. This makes it difficult for participants to fill out the questionnaires and makes it harder to compare the scale means obtained from different questionnaires.

In this paper, we try to address this dilemma by introducing a modular framework that allows the researcher to select the relevant UX aspects out of a larger catalogue of UX scales. All UX scales have a common item and answer format and can thus easily be combined to create a UX questionnaire fitting to the research question behind a product evaluation.

II. Previous Work in the Field

The UEQ+ framework is based on some earlier work which we describe here shortly to make the connection transparent.

In [21] [22] it is investigated how important different UX aspects (for example, *Efficiency*, *Stimulation*, *Trust*, *Aesthetics*) are for certain types of products (for example, social networks, word processing, programming tools, web sites, messengers). The study uses 16 UX aspects extracted from research papers and from an extensive study of the scales used in existing UX questionnaires. Participants of the studies rated the importance of these UX aspects for 15 product categories. Both studies found some clear dependencies between the different product types and the importance ratings for the UX aspects.

Based on the results, it is possible to provide a recommendation on which UX aspects are important for a product category and should therefore be measured in UX evaluations of product of this type [7]. The UX aspects investigated in these studies are good candidates for a framework that should be able to help synthesize UX questionnaires.

Follow-up research [28] shows that quite similar importance ratings are obtained in the context of another culture (Indonesia). The importance of an UX aspect for a type of product thus seems to be mainly a result of the characteristics of the product and not so much by cultural aspects.

The User Experience Questionnaire (UEQ) is an established and widely used UX questionnaire. It already contains the 6 UX scales *Attractiveness*, *Efficiency*, *Perspicuity*, *Dependability*, *Stimulation* and *Novelty* [19]. Scales are represented by 4 items (except *Attractiveness* which contains 6 items) that represent two terms with opposite meanings, for example:

difficult to learn o o o o o o o easy to learn

boring o o o o o o o exciting

inefficient o o o o o o o efficient

slow o o o o o o o fast

Thus, the UEQ is a semantic differential with a 7-point Likert-scale for the answers. The simple item format seems to be quite suitable to define additional scales.

This was already used by some authors to define extension scales for some special product types. In [29], a scale to measure *Trust* was defined. This UX aspect is, for example, highly relevant for online banking applications or web shops.

For household appliances there are also quite specialized UX requirements that strongly influence the overall impression of a product. In [30], two scales for the sounds caused by the operation of a household appliance and for the haptic feeling were developed.

Due to the item format and the fact that a number of scales in a common format already exist, it was decided to base the framework on the UEQ. To make this connection evident, the name UEQ+ was chosen for the framework.

III. Changes in the Item Format

Due to the requirement that it should be possible to combine scales depending on the examined product type, some changes concerning the item format are necessary. We assume that the researcher can freely decide which combination of the available scales he or she wants to use. In addition, the order in which the scales appear in the final questionnaire is up to the researcher.

In the UEQ, the order of the items is randomized. This is also true for the polarization of the items, where half of the items show the positive term in the left position (*fast o o o o o o o slow*) and the other half in the right position (*boring o o o o o o o exciting*).

Some studies (currently unpublished) showed that the polarization of the items does not influence the UEQ scale means (see also [30]), so we decided to use a common scheme with the negative term left and the positive term right for the UEQ+ scales.

Since it should be possible to combine scales in an arbitrary order, and some of the terms are quite similar or even identical in the different scales, it was necessary to group all items of a scale together and set

some context for the correct interpretation of the terms. This is done by introducing a short sentence that is shown on top of the items of a scale and that somehow set a context for the common interpretation of the items.

Thus, a scale in the UEQ+ has the following format (as an example we present the scale *Efficiency*):

To achieve my goals, I consider the product as								
slow	o	o	o	o	o	o	o	fast
inefficient	o	o	o	o	o	o	o	efficient
impractical	o	o	o	o	o	o	o	practical
cluttered	o	o	o	o	o	o	o	organized
I consider the product property described by these terms as								
Completely irrelevant	o	o	o	o	o	o	o	Very important

Thus, we have the statement that connects the items of the scale, then the 4 items with the negative term on the left and the positive term on the right and a final rating concerning the importance of the scale for the overall UX impression of the product. We describe the role of this importance rating at a later point in detail.

IV. Creation of Additional Scales

The UEQ already contains 6 suitable scales that were simply adopted into the UEQ+ (for the scale *Attractiveness,* two of the 6 items were removed to have 4 items for all scales). The same is true for the already available extensions for *Trust*, *Haptics* and *Acoustics*.

The list of UX aspects from [21] [22] was reviewed and the following UX aspects were selected for scale creation: *Aesthetics*, *Adaptability*, *Usefulness*, *Intuitive Use*, *Value*, and *Content Quality*.

Two experts then constructed for every UX aspect a set of items in the UEQ format which describe the aspect semantically. Item suggestions were jointly discussed and consolidated.

In an empirical study, 192 subjects (students that participated on a voluntary basis) rated several products with the corresponding lists of candidate items on a 7-point Likert scale. The average age of the participants (119 male, 73 female) was 30.42 years.

The resulting data were then analysed by principal component analysis. The analysis was done by the function *principal* of the R package *psych* [32]. It was first checked if a one-dimensional solution fits well to the data (which should be the case due to the fact that all items in a candidate set describe the same UX aspect).

We show as an example the candidate set and analysis for the UX aspect *Beauty*. A description for the data analysis for all scales can be found in [33].

The set of candidate items was given as: *ugly / beautiful, lacking style / stylish, unappealing / appealing, ugly in colour / beautiful in colour, inharmoniously / harmoniously, unpleasant / pleasant, not artistically / artistically, thoughtlessly / thought out*.

The original German candidate items used in the study are: *hässlich / schön, stillos / stilvoll, nicht ansprechend / ansprechend, farblich unschön / farblich schön, unharmonisch / harmonisch, unästhetisch / ästhetisch, nicht kunstvoll / kunstvoll, unüberlegt / durchdacht*.

The scree plot of the principal component analysis (see Fig. 1) clearly shows that a one-dimensional solution fits the data well. Proportion of variance explained is 0.64. The fit based upon off diagonal values is 0.99 (values > 0.95 indicate a good fit).

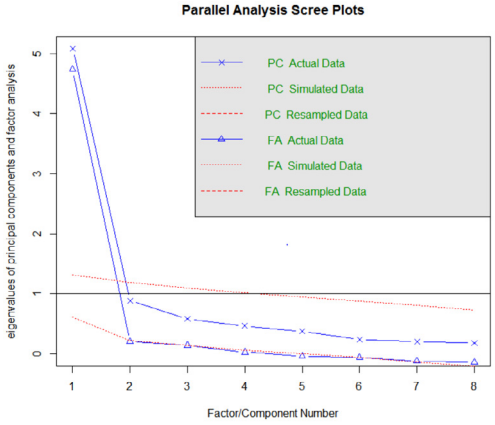


Fig. 1. Scree plot resulting from the principal component analysis of the items for the UX aspect Beauty.

The corresponding loadings of the items on the factor are shown in Table I.

TABLE I. Loadings of the Items on the Factor

Item	Loading
ugly / beautiful	**0.89**
lacking style / stylish	**0.86**
unappealing / appealing	**0.88**
ugly in colour / beautiful in colour	0.79
not harmoniously / harmoniously	0.84
unpleasant / pleasant	**0.88**
not artistically / artistically	0.63
thoughtlessly / thought out	0.51

Thus, the 4 items with the highest loadings (highlighted in bold in Table I) were selected to form the new scale *Beauty*.

If the one-dimensional solution fits the data well, we choose as in this example the 4 items with the highest loading on the factor as representatives for the new UEQ+ scale. This was the case for all UX aspects with the exception of *Content Quality* (see [33]).

For *Content Quality,* a two-dimensional solution fits the data better (see Fig. 2), i.e. there are two different dimensions detected in exploratory principal component analysis. Since the two detected factors could be interpreted, we decided to split this UX aspect into two scales *Trustworthiness of Content* and *Content Quality*.

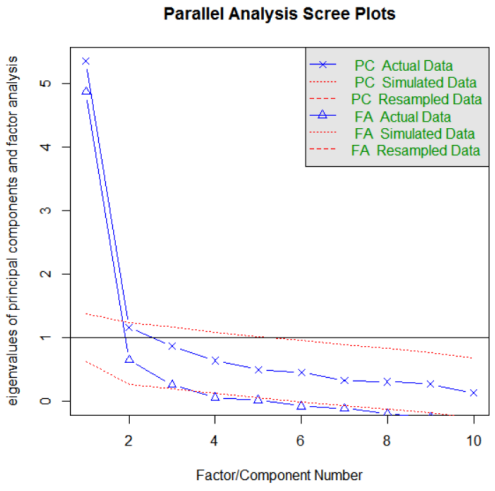


Fig. 2. Scree plot resulting from the principal component analysis of the items for the UX aspect Content Quality.

Table II shows the loadings of the candidate items for *Content Quality* on the two factors extracted by principal component analysis.

TABLE II. FACTOR LOADINGS FOR THE ITEMS OF CONTENT QUALITY

Item	Factor 1	Factor 2
obsolete / up-to-date	0.32	**0.64**
not informative / informative	0.55	0.56
not interesting / interesting	0.21	**0.68**
poorly prepared / well prepared	0.30	**0.77**
incomprehensible / comprehensible	0.10	**0.78**
inferior / valuable	0.58	0.58
useless / useful	**0.68**	0.33
implausible / plausible	**0.90**	0.18
untrustworthy / trustworthy	**0.89**	0.20
inaccurate / accurate	**0.77**	0.28

The items loading high on the first factor express trust in the correctness of the provided information. Items loading on the second factor cover semantically the actuality and quality of the information. Thus, we named the two factors *Trustworthiness of Content* and *Content Quality*.

V. Scales Included in the UEQ+ Framework

The UEQ+ framework currently offers the following UX scales; we show here only the scale names and a short description of the semantic meaning of the scale. The items per scale are listed in Appendix 1.

- *Attractiveness*: Overall impression from the product. Do users like or dislike the product?
- *Efficiency*: Users have the impression that they can complete their tasks without unnecessary effort.
- *Perspicuity*: Subjective impression that it is easy to get familiar with the product. It is easy to learn how to use the product.
- *Dependability*: The user has the impression that he or she controls the interaction.
- *Stimulation*: Feeling that it is exciting and motivating to use the product.
- *Novelty*: Feeling that the product is innovative and creative. The product catches the interest of the user.
- *Trust*: Subjective impression that the data entered into the product are in safe hands and are not used to the detriment of the user.
- *Aesthetics*: Impression that the product looks nice and appealing.
- *Adaptability*: Subjective impression that the product can be easily adapted to personal preferences or personal working styles.
- *Usefulness*: Subjective impression that using the product brings advantages, saves time or improves personal productivity.
- *Intuitive Use*: Subjective impression that the product can be used immediately without any training, instructions or help from other persons.
- *Value*: Subjective impression that the product is of high quality and professionally designed.
- *Trustworthiness of Content*: Subjective impression that the information provided by the product is reliable and accurate.
- *Quality of Content*: Subjective impression if the information provided by the product is up to date, well-prepared and interesting.
- *Haptics*: Subjective feelings resulting from touching the product.
- *Acoustics*: Subjective impression concerning the sound or operating noise of the product.

VI. Importance Rating and KPI

In some use cases it is beneficial to measure not only the means for the different scales, but to provide also a single number (a key performance indicator, or KPI) that summarizes the single scales and can be interpreted as a measure for the overall impression concerning UX.

An extension to calculate such a KPI for the UEQ is described in [34]. The same principle is used for the calculation of a KPI for the UEQ+. The basic idea is to calculate per participant the weighted sum of the scale means with the relative importance ratings. The KPI is then the average of these values of all participants. For the exact formula of the calculation please refer to [34].

VII. First Validation Studies

To evaluate the scale quality, the three product categories *Web Shops*, *Video Platforms* and *Programming Environments* were selected. Two products popular in Germany were chosen per product category (Web Shops: *Otto.de*, *Zalando.de*; Video Platforms: *Netflix*, *Amazon Prime*; Programming Environments: *Eclipse*, *Visual Studio*).

For each product category, a specialized UX questionnaire containing the scales that seemed to be most important for products of this category (see [22] for details) was constructed.

For web shops, these are the scales *Attractiveness*, *Dependability*, *Intuitive Use*, *Visual Aesthetics*, *Quality of Content*, *Trustworthiness of Content*, *Trust* and *Value*.

For video platforms, the scales *Attractiveness*, *Perspicuity*, *Intuitive Use*, *Visual Aesthetics*, *Quality of Content*, *Trustworthiness of Content* and *Trust* were used.

For programming environments, the questionnaire consists of the scales *Attractiveness*, *Dependability*, *Perspicuity*, *Efficiency*, *Usefulness*, *Personalization* and *Value*.

Participants were recruited per e-mail campaigns and by links posted to web sites. Each participant had the choice to rate one product that he or she used regularly from one of the product categories, thus we have different numbers of ratings for the different products (see Table III).

TABLE III. OVERVIEW OVER THE 6 PRODUCT EVALUATIONS

Product	**N**	**Age**	**Sex**	**Time [ms]**	**Clicks**
otto.de	42	34	16 m, 25 f, 1 NA	202,899	54
zalando.de	46	31	20 m, 24 f, 2 NA	187,803	53
Netflix	73	31	42 m, 27 f, 4 NA	211,112	48
Amazon Prime	57	32	36 m, 21 f	259,491	47
Eclipse	14	36	7 m, 4 f, 3 NA	368,552	42
Visual Studio	29	32	25 m, 1 f, 3 NA	225,006	50

Please note that 4 items and the importance of the scale must be rated for every scale. Thus, for 8 scales this requires 40 clicks. In addition, the overall satisfaction must be rated, and two clicks are required to state age and gender.

Thus, filling out the corresponding questionnaires seems to not require much effort from the participants. They spend around 4 minutes (= 240,000 milliseconds) in answering the questions and in addition selected answers seem to not have been changed too often afterwards. This indicates that the used terms are not problematic or difficult to understand.

Tables IV, V and VI show for each product category and evaluated product the scale mean, standard deviation and the Cronbach Alpha coefficient.

TABLE IV. Scale Means, Standard Deviations and Cronbachs Alpha for the Examined Web Shops

Scale	otto.de [N = 42]			zalando.de [N=46]		
	M	STD	Alpha	M	STD	Alpha
Attractiveness	1.30	1.19	0.93	1.68	1.09	0.92
Dependability	1.58	1.08	0.82	2.02	0.89	0.85
Intuitive Use	1.57	1.09	0.94	2.13	0.87	0.90
Vis. Aesthetics	0.89	1.41	0.95	1.47	1.29	0.95
Q. Content	1.35	1.13	0.89	1.91	0.96	0.78
Trustw. Content	1.33	1.15	0.86	1.73	1.02	0.81
Trust	1.28	1.20	0.90	1.26	1.19	0.93
Value	0.93	1.24	0.93	1.58	1.16	0.88

The scale means (see Table IV) are, with the exception of *Trust,* lower for otto.de than for zalando.de. That there is no difference for *Trust* is quite natural, since both shops are quite established shops with a longer history. The scales obviously allow to differentiate between different products.

TABLE V. Scale means, Standard Deviations and Cronbachs Alpha for the Examined Video Platforms

Scale	Netflix [N=73]			Amazon Prime [N=57]		
	M	STD	Alpha	M	STD	Alpha
Attractiveness	2.13	1.06	0.95	1.61	1.14	0.90
Perspicuity	2.04	1.19	0.80	1.62	1.41	0.91
Intuitive Use	1.86	1.16	0.90	1.36	1.38	0.94
Vis. Aesthetics	1.58	1.17	0.89	1.01	1.28	0.94
Q. Content	1.83	1.23	0.84	1.49	1.27	0.82
Trustw. Content	1.48	1.12	0.87	1.46	1.21	0.87
Trust	1.03	1.40	0.90	0.71	1.73	0.96

Except for *Trustworthiness of Content* (see Table V), the ratings are higher for Netflix than for Amazon Prime, which is not unexpected since the source of content of both tools is quite similar concerning trustworthiness. Again, the other scales differentiate between the two products.

TABLE VI. Scale Means, Standard Deviations and Cronbachs Alpha for the Evaluated Programming Environments

Scale	Eclipse [N=14]			Visual Studio [N=29]		
	M	STD	Alpha	M	STD	Alpha
Attractiveness	0.48	1.71	0.93	1.67	0.83	0.76
Dependability	0.84	1.80	0.97	1.77	0.82	0.83
Perspicuity	0.11	1.68	0.93	0.93	1.16	0.86
Efficiency	0.71	1.53	0.90	1.44	1.02	0.80
Usefulness	1.21	1.74	0.98	2.00	0.96	0.82
Personalization	1.25	1.56	0.98	1.78	0.91	0.80
Value	0.32	1.64	0.93	1.66	1.08	0.79

Visual Studio ratings (see Table VI) are for all scales much higher than the ratings for Eclipse. It must, however, be noted that we had only a small number of participants for programming environments, thus these results need to be interpreted with care.

In general (see Table IV, V, VI), the Cronbach Alpha values for all evaluated products are extremely high. Cronbach Alpha is an indicator for scale consistency based on the correlations of the items within a scale. Values above 0.7 are usually interpreted as reasonably high scale quality. The observed values for the scales of the UEQ+ are in all cases above this threshold.

The observed ratings for the importance of the scales confirm that the selected scales were considered as important for the evaluated products by the participants. Detailed values of the importance ratings and some additional information concerning the scale means can be found in [33].

As described above, it is possible to calculate a KPI using the scale means and the importance ratings of the scales. This KPI is interpreted as an indicator for the overall satisfaction concerning the UX of the product. To verify this assumption, each online questionnaire contains as one last point the item:

Overall, concerning the user friendliness of <Product> I am

Very dissatisfied o o o o o o o Very satisfied

Table VII shows the correlation of the ratings of this item to the calculated KPI.

TABLE VII. Satisfaction Ratings and KPI for All Studies

Product	Satisfaction		KPI		Corr
	M	STD	M	STD	
otto.de	5.48	1.24	1.27	0.90	0.71
zalando.de	5.65	0.91	1,70	0.69	0.66
Netflix	6.06	0.99	1.73	0.74	0.77
Amazon Prime	5.30	1.08	1.35	0.87	0.78
Eclipse	4.21	1.74	0.40	1.37	0.83
Visual Studio	5.55	0.97	1.59	0.57	0.71

Correlations between the satisfaction ratings and the calculated KPI are quite high. Thus, our interpretation of the KPI seems to be valid. In addition, since the correlation seems to be quite stable over different products and combinations of scales (each product category was evaluated with different sets of UEQ+ scales), it may be possible to develop a benchmark for the KPI that can be used independently of the selected scales for an evaluation.

VIII. Advantages and Disadvantages of a Modular Construction of UX Questionnaires

The big advantage of the UEQ+ is that it allows researchers to create UX questionnaires perfectly adapted to the research question, i.e. such a questionnaire contains exactly the scales that need to be measured. In addition, all scales follow a uniform item format, which makes it easy for the participants to answer the items.

But such a modular approach is not without its shortcomings. There are some disadvantages compared to using a standard questionnaire like the UEQ out of the box.

Obviously, the effort to set up the questionnaire is higher. An application of the UEQ+ requires that the researcher have a clear picture concerning the UX aspects that are relevant for the product and should therefore be measured. There are some recommendations available that show how important different UX aspects are for different types of products (see [22] and [7]). In addition, the UEQ+ handbook (can be downloaded from www.ueqplus.ueq-research.org) contains some detailed suggestions concerning the most relevant UEQ+ scales for several typical product categories. But of course, not all products will fall into one of the described product categories in these papers and it must be checked if there are maybe exceptions for a specific product.

Most standard questionnaires offer some tools for data analysis. Thus, it is sufficient to collect the data, drop it to the tool and not all, but many important analyses are done automatically. We also offer a data analysis tool (can be downloaded free of charge from www.ueqplus.ueq-research.org), but since the scale structure of the resulting questionnaire is not fixed, this tool only provides limited support.

In addition, interpretation of the results is a bit harder in the UEQ+ than in standard questionnaires. What does a scale value of 1.3 for a scale

mean? Is this a good, medium or bad value compared to other products?

Standard questionnaires, for example the UEQ itself [19] or the SUS [23] or VISAWI [25], offer large benchmark data sets that are based on evaluation results for larger sets of different products. Thus, a simple comparison of the result obtained in an evaluation to the results in the benchmark data set offers some insights concerning the question of how good or bad the impression of users towards the product is compared to other available products.

For the scales from the UEQ, such a benchmark is available, for the newly added scales this is at the moment not the case. For some frequently used scales this situation may change, but some of the scales are obviously only relevant for special types of products, so it may take a long time until a benchmark in the quality of the UEQ benchmark will be available for all scales of the UEQ+.

IX. When to use UEQ+?

Given the remarks concerning the advantages and disadvantages of a modular questionnaire, it is possible to give some recommendations.

If you are setting out to evaluate a single product and your main research question is to get an idea about the UX quality of this product, you should use the UEQ. Even if some of the scales do not perfectly match your product or if some scales that you think are important are missing, the availability of the UEQ benchmark and the ease of use of the available material, like the data analysis tool, would clearly speak for using the UEQ.

If you are planning to evaluate the same product multiple times, for example to get an insight if the product improves over time, and if the UEQ scales do not capture most of the UX aspects you consider relevant, then opting for your own special questionnaire built with the UEQ+ is the better choice. In this scenario, the lack of a benchmark is not a big issue, since you are mainly interested in comparing multiple measurements of the same product over time. Thus, capturing the UX quality in an optimized form is more important here.

If you want to set up an UX measurement as part of your quality process for a larger suite of similar products (in the sense that the same UX aspects apply to all of them) and if the scales of the original UEQ do not fit well to your needs, then it is also recommended to set up your own questionnaire using the UEQ+. In this case the additional effort required is minor, since you do this only once and reuse it in a large number of concrete evaluations. In addition, the lack of a benchmark is not so important, since over time you will generate your own data set of evaluations that will help to interpret then the results obtained for a single product, i.e. in such a scenario you will quickly generate enough data yourself.

X. Conclusions and Further Work

We described the development of a modular framework for the creation of UX questionnaires. This framework allows the researcher to select the UX aspects that are relevant for a certain product from a list of existing UX scales. Thus, a customized questionnaire containing exactly those UX scales that are important for the users of the product can be created.

Currently, the UEQ+ framework contains 16 scales. Of course, they do not cover the entire concept of UX. Other scales may be required for some products and new use cases and product types entering the market in the future will create the need for different, not-yet-considered UX scales. Thus, a framework like the UEQ+ is always a work in progress and at no point in time will it be truly finished. We will try to provide some additional scales in the near future and hope that other researchers will (as they did already by constructing some of the extension scales for the UEQ) help to provide new scales, which we can then integrate into the UEQ+ framework.

Another important area of future work is the improvement of the existing benchmarks. This simply requires time to collect sufficiently large sets of data.

Six of the UEQ+ scales are concerning their items identical to the original UEQ scales. However, the item format is slightly changed. Items of a scale are grouped in the UEQ+ and the positive term is always right (in the original UEQ items appear in random order and polarity). In addition, a statement has been added that described the common meaning of all items in a scale. It is currently not fully investigated if these changes have an impact on the results, i.e. if the scale means obtained from the UEQ+ scales are fully comparable to the scale means of the corresponding UEQ scales. We expect only minor deviations, but this must be of course evaluated in further studies.

Currently, the items for the extension scales of the UEQ are available only in German and English. The six scales taken over directly from the UEQ are available in more than 20 languages (see, for example, [35] [36] for the description of the Spanish and Portuguese language versions). Of course, we hope to provide some more translations in the future.

Appendix I

In the following we present the complete list of scales and items available in the UEQ+ framework.

Attractiveness

In my opinion, the product is generally:

- annoying / enjoyable
- bad / good
- unpleasant / pleasant
- unfriendly / friendly

Efficiency

To achieve my goals, I consider the product as:

- slow / fast
- inefficient / efficient
- impractical / practical
- organized / cluttered

Perspicuity

In my opinion, handling and using the product are:

- not understandable / understandable
- difficult to learn / easy to learn
- complicated / easy
- clear / confusing

Dependability

In my opinion, the reactions of the product to my input and command are:

- unpredictable / predictable
- obstructive / supportive
- not secure / secure
- does not meet expectations / meets expectations

Stimulation

In my opinion, handling and working with the product are:

- inferior / valuable
- boring / exciting
- not interesting / interesting
- demotivating / motivating

Novelty

In my opinion, the idea behind the product and its design are:

- dull / creative
- conventional / inventive
- common / cutting edge
- conservative / innovative

Trust
Regarding the use of my personal information and data, the product is:

- insecure / secure
- untrustworthy / trustworthy
- unreliable / reliable
- non-transparent / transparent

Aesthetics
In my opinion, the visual design of the product is:

- ugly / beautiful
- lacking style / stylish
- unappealing / appealing
- unpleasant / pleasant

Adaptability
Regarding my personal requirements and preferences, the product is:

- not adjustable / adjustable
- not changeable / changeable
- inflexible / flexible
- not extendable / extendable

Usefulness
I consider the possibility of using the product as:

- useless / useful
- not helpful / helpful
- not beneficial / beneficial
- not rewarding / rewarding

Intuitive Use
In my opinion, using the product is:

- difficult / easy
- illogical / logical
- not plausible / plausible
- inconclusive / conclusive

Value
I generally consider the design of the product as:

- inferior / valuable
- not presentable / presentable
- tasteless / tasteful
- not elegant / elegant

Trustworthiness of Content
In my opinion, the information and data provided by the product are:

- useless / useful
- implausible / plausible
- untrustworthy / trustworthy
- inaccurate / accurate

Quality of Content
In my opinion, the information and data provided by the product are:

- obsolete / up-to-date
- not interesting / interesting
- poorly prepared / well prepared
- incomprehensible / comprehensible

Acoustics
The noise during use of the product is:

- loud / quiet
- dissonant / melodic
- booming / dampened
- piercing / soft

Haptics
In my opinion, the surface of the product is:

- unstable / stable
- unpleasant to the touch / pleasant to the touch
- rough / smooth
- slippery / slip-resistant

References

[1] Dumas, J. S. & Redish, J. (1999). A practical guide to usability testing. Intellect books.

[2] Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. In Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 249-256. ACM.

[3] Nielsen, J. (1994). Enhancing the explanatory power of usability heuristics. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pp. 152-158. ACM.

[4] Bader, F., Schön, E.-M., Thomaschewski, J. (2017). Heuristics Considering UX and Quality Criteria for Heuristics. International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 4, No. 6, pp. 48-53. DOI: 10.9781/ijimai.2017.05.001.

[5] Wharton, C., Rieman, J., Lewis, C. & Polson, P. (1994). The cognitive walkthrough method. A practioner's guide. In: Jakob Nielsen, Robert L. Mack (Ed.): Usability Inspection Methods, pp. 105–140. John Wiley & Sons, New York NY.

[6] Lewis, C., & Wharton, C. (1997). Cognitive walkthroughs. In Handbook of human-computer interaction, pp. 717-732. North-Holland.

[7] Schrepp, M. (2018). User Experience mit Fragebögen messen [Measure user experience with questionnaires]. Amazon Kindle Direct Publishing, ISBN: 9781986843768.

[8] Hassenzahl, M. (2001). The effect of perceived hedonic quality on product appealingness. International Journal of Human-Computer Interaction 13(4), pp. 481–499.

[9] Schrepp, M., Hinderks, A., & Thomaschewski, J. (2014). Applying the user experience questionnaire (UEQ) in different evaluation scenarios. In International Conference of Design, User Experience, and Usability, pp. 383-392. Springer, Cham.

[10] Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017). Construction of a Benchmark for the User Experience Questionnaire (UEQ). International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 4, No. 4, pp. 39- 45. DOI: 10.9781/ijimai.2017.445.

[11] Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S. & Carey, T. (1994). Human-Computer Interaction. Workinghal, Addison-Wesley.

[12] ISO 9241-210: Ergonomics of Human-System interaction – Part 210. Human centred design for interactive systems. International Organization for Standardization.

[13] Naumann, A., Hurtienne, J., Israel, J. H., Mohs, C., Kindsmüller, M. C., Meyer, H. A., & Hußlein, S. (2007). Intuitive use of user interfaces: defining a vague concept. In: International Conference on Engineering Psychology and Cognitive Ergonomics, pp. 128-136. Springer, Berlin, Heidelberg.

[14] Comber, T., & Maltby, J. R. (1997). Layout complexity: does it measure usability?. In: Human-Computer Interaction INTERACT'97, pp. 623-626. Springer, Boston, MA.

[15] Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: a comparison of two theoretical models. Management science, 35(8), pp. 982-1003.

[16] Hatscher, M. (2001). Joy of use-Determinanten der Freude bei der Software-Nutzung [Determinants of enjoyment during use of software]. In: Mensch & Computer [Humans & computers] 2001: 1. Fachübergreifende Konferenz [Interdisciplinary conference]. BG Teubner.

[17] Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität [AttrakDiff: A questionnaire for the measurement

of perceived hedonic and pragmatic quality]. In: Ziegler, J. & Szwillus, G. (Ed.), Mensch & Computer [Humans & computers] 2003. Interaktion in Bewegung [Interaction in movement], S. 187-196, Stuttgart, Leipzig: B.G. Teubner.

[18] Moshagen, M., & Thielsch, M. T. (2010). Facets of visual aesthetics. International journal of human-computer studies, 68(10), pp. 689-709.

[19] Laugwitz, B.; Schrepp, M. & Held, T. (2006). Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten [Construction of a questionnaire for the measurement of user experience of software products]. In: A.M. Heinecke & H. Paul (Eds.): Mensch & Computer [Humans & computers] 2006 - Mensch und Computer im Strukturwandel [Humans and computers in times of change in structure]. Oldenbourg Verlag, pp. 125 – 134.

[20] Thielsch, M. T., Blotenberg, I., & Jaron, R. (2014). User evaluation of websites: From first impression to recommendation. Interacting with Computers, 26(1), pp. 89-102.

[21] Winter, D., Schrepp, M., & Thomaschewski, J. (2015). Faktoren der User Experience-Systematische Übersicht über produktrelevante UX-Qualitätsaspekte [Factors of user experience - A systematic overview of product-related UX quality aspects]. In: Mensch und Computer [Humans and computers] 2015–Usability Professionals.

[22] Winter, D., Hinderks, A., Schrepp, M., & Thomaschewski, J., (2017). Welche UX Faktoren sind für mein Produkt wichtig? [Which UX factors are important for my product?] In: Hess, S. & Fischer, H. (Ed.), Mensch und Computer [Humans and computers] 2017 - Usability Professionals. Regensburg: Gesellschaft für Informatik e.V., pp. 191 – 200.

[23] Brooke, J. (1996). SUS-A quick and dirty usability scale. Usability evaluation in industry, 189(194), pp. 4-7.

[24] Kirakowski, J., & Corbett, M. (1993). SUMI: The software usability measurement inventory. British journal of educational technology, 24(3), pp. 210-212.

[25] Thielsch, M. T., & Moshagen, M. (2011). Erfassung visueller Ästhetik mit dem VisAWI [Capture of visual aesthetics with VisAWI]. Tagungsband [Conference transcript] UP11.

[26] Minge, M. & Riedel, L. (2013). meCUE – Ein modularer Fragebogen zur Erfassung des Nutzungserlebens [meCue – A modular questionnaire for capturing the user experience]. In: S. Boll, S. Maaß & R. Malaka (Ed.): Mensch und Computer [Humans and computers] 2013: Interaktive Vielfalt [Interactive diversity], pp. 89-98. München, Oldenbourg Verlag.

[27] Gediga, G., Hamborg, K. C., & Düntsch, I. (1999). The IsoMetrics usability inventory: an operationalization of ISO 9241-10 supporting summative and formative evaluation of software systems. Behaviour & Information Technology, 18(3), pp. 151-164.

[28] Schrepp, M., & Santoso, H. (2018). Has Culture an Impact on the Importance of UX Aspects? In: Mensch und Computer [Humans and computers] 2018-Workshopband [Workshop transcript].

[29] Hinderks, A. (2016). Modifikation des User Experience Questionnaire (UEQ) zur Verbesserung der Reliabilität und Validität [Modification of the user experience (UEQ) for the improvement of reliability and validity]. Unpublished Master Thesis, University of Applied Sciences Emden/Leer.

[30] Boos, B. & Brau, H. (2017). Erweiterung des UEQ um die Dimensionen Akustik und Haptik [Extension of the UEQ by the dimensions acoustics and haptics]. In: Hess, S. & Fischer, H. (Eds.), Mensch und Computer [Humans and computers] 2017 – Usability Professionals, Regensburg: Gesellschaft für Informatik e.V., pp. 321-327.

[31] Sauro, J., & Lewis, J. R. (2011). When designing usability questionnaires, does it hurt to be positive?. In: CHI Proceedings, pp. 2215-2224.

[32] Revelle, W. (2018) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psych Version = 1.8.12.

[33] Schrepp, M. & Thomaschewski, J. (2019). Construction and first Validation of Extension Scales for the User Experience Questionnaire (UEQ). Research Report University of Applied Sciences Emden/Leer. DOI: 10.13140/RG.2.2.19260.08325.

[34] Hinderks, A., Schrepp, M., Domínguez Mayo, F.J., Escalona, M.J., Thomaschewski, J. (2019). Developing a UX KPI based on the User Experience Questionnaire. Computer Standards & Interfaces. DOI: 10.1016/j.csi.2019.01.007.

[35] Rauschenberger, M., Schrepp, M., Cota, M.P., Olschner, S. & Thomaschewski, J. (2013). Efficient measurement of the user experience of interactive products - How to use the User Experience Questionnaire (UEQ). Example: Spanish Language Version. In: International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 2, No. 1, pp. 39-45. DOI: 10.9781/ijimai.2013.215.

[36] Cota, M.P.; Thomaschewski, J.; Schrepp, M. & Goncalves, R. (2013). Efficient Measurement of the User Experience- A Portuguese Version of the User Experience Questionnaire. In: Procedia Computer Science, Vol. 27, 5th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, DSAI 2013, pp. 491 – 498. DOI: 10.1016/j.procs.2014.02.053.

Martin Schrepp

Martin Schrepp has been working as a user interface designer for SAP AG since 1994. He finished his diploma in mathematics in 1990 at the University of Heidelberg (Germany). In 1993 he received a PhD in Psychology (also from the University of Heidelberg). His research interests are the application of psychological theories to improve the design of software interfaces, the application of Design for All principles to increase accessibility of business software, measurement of usability and user experience, and the development of general data analysis methods. He has published several papers concerning these research fields.

Jörg Thomaschewski

Jörg Thomaschewski received a PhD in physics from the University of Bremen (Germany) in 1996. He became Full Professor at the University of Applied Sciences Emden/Leer (Germany) in September 2000. His research interests are human-computer interaction, e-learning, and software engineering. Dr. Thomaschewski is the head of the research group “Agile Software Development and User Experience”.

SOLAM: A Novel Approach of Spatial Aggregation in SOLAP Systems

Farah Amina Zemri*, Djamila Hamdadou

Computer Science Laboratory of Oran (LIO), University of Oran 1, Oran (Algeria)

Received 10 January 2019 | Accepted 2 August 2019 | Published 30 October 2019

Abstract

In the context of a data driven approach aimed to detect the real and responsible factors of the transmission of diseases and explaining its emergence or re-emergence, we suggest SOLAM (Spatial on Line Analytical Mining) system, an extension of Spatial On Line Analytical Processing (SOLAP) with Spatial Data Mining (SDM) techniques. Our approach consists of integrating EPISOLAP system, tailored for epidemiological surveillance, with spatial generalization method allowing the predictive evaluation of health risk in the presence of hazards and awareness of the vulnerability of the exposed population. The proposed architecture is a single integrated decision-making platform of knowledge discovery from spatial databases. Spatial generalization methods allow exploring the data at different semantic and spatial scales while reducing the unnecessary dimensions. The principle of the method is selecting and deleting attributes of low importance in data characterization, thus produces zones of homogeneous characteristics that will be merged.

Keywords

Spatial Data Mining, Spatial OLAP, Spatial Generalization, Attribute Oriented Induction, Characterization.

DOI: 10.9781/ijimai.2019.10.002

I. Introduction

THE prevention of epidemics is a great issue of public health and problematical phenomena. Geographical location of housing zones subject to contamination contributes in circumscribing and controlling the spread of epidemic through a prevention strategy and thoughtful management. In this context, the role of the decisional information system is essential in the access to information and decision-making support by health institutions and local authorities in charge of this prevention.

To identify a good strategy for preventing epidemics and to guarantee a careful management of the spread of epidemics, it is important to design a quality epidemiological surveillance system. Traditionally, decision-making information systems rely heavily on data warehouses offering tools for multidimensional exploration of data and indicators through online analysis or online analytical processing (OLAP).

With regard to phenomena such as the epidemic, the spatial (and temporal) dimensions are eminently important in the analysis and qualification of the propagation of the phenomenon in the neighborhood, as well as its emergence or re-emergence.

The last years have witnessed the development of research works on Spatial OLAP or SOLAP [1] [2] integrating spatial data in OLAP and linking exploration and visualization mapping. The natural approach is therefore to apply and experiment these approaches and the resulting system EPISOLAP [3] was proposed and dedicated to epidemiological analysis.

SOLAP method has proven its efficiency in a few years only from its creation. At the beginning of our study, we have chosen to use this new approach in the monitoring of epidemics. We assume that public health is a perfect field to show the advantages of SOLAP technology and its benefits in processing complex data (spatial data) with a decisional manner and not transactional as the one used in traditional geographic information systems(GISs). SOLAP did not cancel completely the usefulness of GISs but their use becomes limited to the interrogation of geo-spatial data, recording the results of analysis and executing simple queries not exceeding the standards of a relational database such as Insert or Select operations. The big inconvenient of transactional process is that more the query is complicated; the answer to this query is slower because of the tabular structure of a relational database. Contrariwise, in geospatial data warehouse, data is represented as cubes or hyper-cubes. Multidimensional structure facilitates access to information, speeds up the processing of complex queries and the response is given in few seconds only (time of reflection of a human being). Moreover, the reporting service of SOLAP process offers a quality representation by structuring the results of analysis in understandable forms facilitating the process of decision making.

Thereafter, limits have been detected. The objective of the current work is to combine the principles of SOLAP and spatial data mining methods in the same spirit of OLAM (On-Line Analytical Mining) or Multidimensional Mining, proposed in [4], [5]. It is a matter of enriching SOLAP functions not only to be restricted on exploration and visualization of spatial data but also to enrich the aggregation function which is not necessarily a simple average or a sum but which may be based on spatial regression, classification, association or characterization. The integration between SOLAP and SDM leads to the development of a new EPISOLAP-MINING decision-making system based on SOLAM engine (SOLAP Mining) that we are trying to design through this study. The main two objectives of this article are: (1) analyze epidemiological risk factors and rank them in order of relevance, allowing authorities to intervene and eliminate the most influential in priority and (2) apply aggregation function in SOLAP

* Corresponding author.
E-mail address: zemri_farah@yahoo.fr

systems using spatial generalization according to spatial characteristics of administrative zones instead of spatial generalization depending only on the administrative division hierarchy.

To achieve these two objectives, we introduce two key contributions which are respectively: (1) calculate information gain of each attribute in the Attribute Oriented Induction (AOI) algorithm, classify the attributes according to the values of the gains and drop first the attribute corresponding to the lowest gain, then proceed to generalize identical neighboring tuples, (2) Use comparison method which is a type of characterization method often used to distinguish one target class from one or more contrasting classes. Therefore, zones can be aggregated according to their identical characteristics.

In section II we present the related work found in the literature in relation to spatial generalization. Section III presents our proposed approach and SOLAM engine. The description of spatial generalization process is given in section IV. In section V, the generalization process is applied on epidemiological surveillance with some illustrations. At the end, we conclude our discussion in section VI by giving some perspectives.

II. Related Work

Since the introduction of the SOLAP system by Professor Yvan Bedard, several studies have emerged in order to improve and develop new proposals to make the SOLAP applications more complete. For example, in the study of [6], several SOLAP applications were presented (in different fields: environmental health, transportation, etc.), and gaps were raised and then remedied.

On the other hand, and in the last decade, several efforts have been made to enrich OLAP systems to support complex space objects. We quote here the work of [7] who proposed C^3 model presented as a new modeling strategy to support complex spatial objects in OLAP cubes. This model is the extension of the "BigCube" model proposed in [8]. The C^3 model includes three constructors: "categorization construct", "containment construct" and "combination or cubing construct" allowing all three to structure complex and multi-structured regions and build the multidimensional data cube. The authors also proposed another geographical constructor called "geo-construct" to manipulate and query the spatial hierarchies of geographical objects. These objects are subdivided into two types: Structured Object (SO) and Base Object (BO). On the basis of these two types of geographical objects they proposed the hierarchy of regions called "regH: region hierarchy" organized like this:

$$<region : SO> :=<regionLabel : BO><face : SO>$$

In [9] researchers proposed "MuSD" (multigranular spatial data warehouse) which is based on spatial fact, spatial dimension and multilevel spatial measure. "MuSD" manages and supports spatial measures in multiple levels of geometric granularity. In addition to the representation constructs, the model includes a set of SOLAP operators to navigate not only across dimensional levels, but also through the different levels of spatial measure. Recently, the work of [10] focused on the management of qualitative measures in addition to quantitative measures in OLAP systems as well as the proposal of new drilling operators to navigate in the different levels of the multidimensional cube. The objective of the research was the adaptation of drilling operations by introducing the notion of fuzzy to calculate probabilistic measures by aggregating the qualitative dimensions.

Another category of studies have tackled the problematic of non-strict spatial hierarchies in spatial data warehouses and proposed new logical models in [11], by introducing logic schemes with "Bridge Table". More recently in [12] the authors have proposed another solution for implementation of this special type of hierarchy that exists in some spatial applications by introducing the bitmap index (Index-NN).

As part of the enhancement of spatial data processing capability, researchers in [13] introduced the notion of auxiliary spatial dimensions and adapted SOLAP operations to handle this kind of spatial dimensions. Their role is to facilitate the processing and execution of complex spatial queries and to ensure a better response time. Another study in [14] focused on managing complex levels of details of hierarchy in OLAP systems. They were the first to propose a "multi-model" that uses semantics in data aggregation to avoid the introduction of incorrect queries and nonsense responses. The authors also introduced in this same work, a high level language for multidimensional databases called "SumQL" which was later extended to the "SumQL ++" language to support distributed object databases. In the same vein, recent work in [15] has been the subject of a new approach that defines the constraints of integrity in SOLAP systems in order to improve the accuracy, coherence and completeness of spatial databases. Then, a classification of SOLAP user constraints was applied in constellation schema composed of three fact tables sharing spatial, temporal and thematic dimensions. The approach was applied in the field of agro forestry in the region of Mostaganem in Algeria.

Some open research questions that are currently being investigated are listed as new SOLAP trends, in [16]. These trends include: Continuous SOLAP, Real-Time SOLAP, SOLAP for Emergencies and SOLAP Web Applications.

In the field of web mapping, in [17] researchers have set up the platform "UMapIT" (Unrestricted Mapping Interactive Tool) for an efficient management of spatially referenced data on the web. This web platform was introduced as a new paradigm that integrates the multidimensional approach (datacube) and the occurrence-driven approach to support the demands of users on the web. This approach combines geometry, semantics, and graphical representation as dimension tables in a star schema and is connected to the spatial fact table defined as "Vuel" (Element View). This approach allows a better management of the levels of details in spatial hierarchies on the web. In [18] GeWOLAP, another SOLAP Web-based prototype is presented with some new SOLAP operators adapted to web applications.

Our approach, in the current study, is inspired from the concept of OLAM that was first proposed in [19] where it has been suggested as an OLAP tool coupled with data mining. DBMiner prototype has been developed and generalized for multiple data mining functions like Characterization (Characterizer of DBMiner), Comparison, Classification (Classifier of DBMiner), Association (Associator of DBMiner), Prediction (Predictor of DBMiner) and Clustering. Several manners to perform this integration were proposed: cubing then mining, mining then cubing, cubing while mining, backtracking and comparative mining.

The generalization process is based on attribute oriented induction method (AOI) proposed in [20] and a basic attribute oriented induction algorithm that has been proposed in [21]. More recently, in [22], author has summarized spatial generalization method used in generating characteristic rules and using Attribute Oriented Induction method in eight sequential steps which are ordered as follows: data focusing; attribute removal; attribute generalization; count propagation; attribute generalization control; relation generalization control; rule transformation and finally, handling overlapping tuples. This last step is specified to generate classification/discriminant rules.

Afterward, GeoMiner has been designed in [23] as the extension of DBMiner to spatial data and proposed three modules: Geo-Characterizer, Geo-Comparator and Geo-associator. They are based on spatial generalization defined in [24] as the process of data aggregation according to conceptual hierarchies. The discovery of general relationships between spatial and non-spatial data can be performed

by attribute oriented induction in two ways: (1) spatial dominant generalization (SDG) and (2) non-spatial dominant generalization (NSDG). The former is spatial hierarchy directed induction which performs generalization on spatial data first; whereas the later is non-spatial attribute oriented induction which performs the generalization on non-spatial data first [25].

Other kinds of spatial generalization were also indicated as novel alternatives of generalization like interleaved generalization between spatial and non- spatial data and generalization on multiple thematic maps. Several algorithms were proposed to perform spatial data generalization like GenDis algorithm in [26].

According to [24], the client tools currently used to run conventional data warehouses are not suitable for geospatial data warehouses because they do not exploit the geometric structure of the data. In this optic, SOLAP technologies were developed and implemented [27].

III. Proposed Approach

The approach that we propose here includes an extension of the SOLAP systems with predictive spatial data mining (SDM) techniques to highlight the risk factors in the field of epidemiological surveillance. The purpose is describing and explaining the factors of epidemic phenomenon observed and recorded within "EPISOLAP" data warehouse, in order to better understand them and even avoid them. More precisely, SDM will help in the discovery of the correlations between spatial phenomena and will give a precise description of the epidemiological scenarios which should improve understanding risk factors of epidemic and orient the actions of prevention. In the context of this application, coupling SOLAP and SDM will fulfill the following functions:

- SOLAP tool represents the spatial aspect of the epidemiological study thanks to its ability of detection and location of epidemic outbreaks.
- SDM highlights certain socio-environmental aspects and socio-economic transformations that influence the spatial dynamic of epidemic phenomena and which contribute to various scenarios of the transmission.

However, in the literature, the two approaches have been proposed separately, and were not enough integrated. On the one hand, it is regrettable that SDM is not performed in an online way and only deals with data represented in tabular form (individual-variable table) and on the other hand, conventional SOLAP operators allow only aggregating, visualizing and exploring spatial data without any interpretation efforts. This makes the coupling between these two tools more than necessary to achieve this complementarily. Such coupling is not straightforward and there are several challenges to overcome which we could clarify by responding to the following questions:

- How can SOLAP and data warehouse integrate SDM algorithms?
- How to store in a data warehouse the knowledge extracted by an SDM method?
- How to execute on-line, sometimes large cubes, excavation algorithms of a certain complexity and time consuming?
- How to model in a multidimensional way complex data (geographic data)?
- How to make online analysis on the spatial data?

It is obvious that the implementation of such coupling will not be without difficulties and will have to be carried out under the respect of certain constraints. Indeed, unlike queries on data cubes, the OLAP-based search should take the context into account. At a given time, the user of a cube displays facts with different levels of granularity. This instant photography of the data can be used to perform data mining. Our approach is inspired from the vision of Jiawei Han who integrated OLAP with data mining in one coherent system which he called OLAM (On Line Analytical Mining) in [4]. In another work [28], we have designed EPISOLAP-MINING by combining a multi criteria spatial decision support system SDSS_MCA_DM (Spatial Decision Support System based on Multi Criteria Analysis and Data Mining) [29] and SOLAM sub-system. The purpose of the present article is precisely presenting this later module. An overview of SOLAM module is illustrated on Fig. 1.

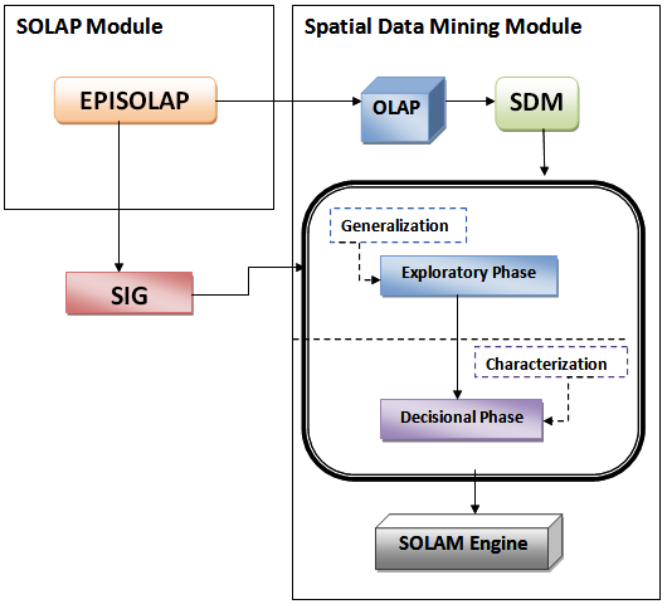


Fig. 1. SOLAM sub-system of EPISOLAP MINING.

The proposed SOLAM sub system follows two fundamental steps (Exploratory step and decisional step) by following the general approach of SDM proposed in [29]:

A. Exploratory Phase

Allows a synthetic description (global auto-correlation index, generalization, density, smoothing), to discover the discrepancies giving the local specificities (local auto-correlation or local factor analysis) or to look for clusters. This first phase guides the decision-making phase. Our choice in this exploratory phase was on the spatial generalization method.

B. Decisional Phase

In this phase, more detailed analysis is made to explain the differences or to characterize the groups (characterization, classification rules or association rules). Our choice in this decisional phase was on characterization method.

IV. Description of the Spatial Generalization Process

In this study, we have chosen **non-spatial dominant generalization**; this one does not use a spatial hierarchy but generates less detailed locations by fusion of spatial objects.

The principle of this process is simple, an attribute oriented induction (AOI) method [30] is performed using thematic hierarchies, but keeping their spatial description. This induction produces values of homogeneous attributes for several objects. These objects are then merged.

The comparison process breaks down into four processes, starting with the process of data collection, the choice of dimension relevance analysis, the process of applying synchronous generalization and

finally, the presentation of driven comparison. Fig. 2 illustrates step 1 and 2 and Fig. 3 illustrates step 3 and 4.

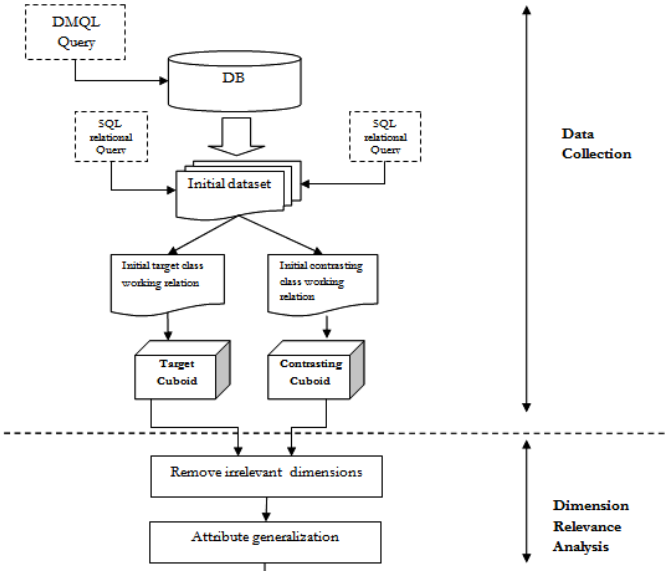


Fig. 2. Step 1 and 2 of comparison process.

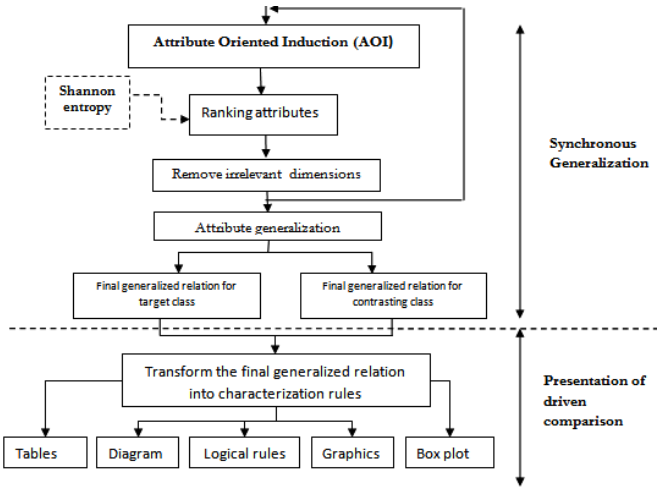


Fig. 3. Step 3 and 4 of comparison process.

A. Step 1: Data Collection

The relevant data of the database is collected by applying SQL query which partitions the data set into one target class and one or a set of contrasting class(es).

B. Step 2: Dimension Relevance Analysis

Only the relevant dimensions are included in the subsequent analysis after applying dimension relevance analysis on target and contrasting class(es). The method leads on ranking the relevance of the dimensions and only the more relevant dimensions will be included in data comparison process.

C. Step 3: Synchronous Generalization

The basic AOI algorithm used in this step is that proposed in [21]. The AOI algorithm breaks down into two processes, starting with the process of ***grouping the attributes*** according to the different concepts hierarchies and then the process of ***merging identical tuples***.

The AOI algorithm is performed by substituting lower level concepts with its corresponding prime level concepts, which leads to a prime relation, by eliminating duplicated tuples and accumulating the counts in the retained generalized tuples.

The ***prime target cuboid*** and the ***prime contrasting cuboid*** (s) arise from the generalization process that is performed on the target class to the level controlled by a user/expert specified ***dimension threshold*** and on the contrasting class generalized to the same level. Discriminated or comparison rule is an assertion which discriminates (or compares) the concepts of one (target) class from another (contrasting) class [31]. This rule gives a discriminated criterion which can be used to predict the class membership of new data.

Attribute-Oriented Induction is robust and handles noise and/or exceptional cases elegantly because it incorporates statistical information (using count) and generates disjunctive rules. The association of count with each disjunction leads naturally to mining approximate rules, for which the conditions with negligible weight can be dropped in generalization and rule formation, since a negligible weight implies a minimal influence to the conclusion.

In our proposed approach, an alternative of comparison algorithm could be performed by calculating informational gain using the ID3 formula and taking into account the neighboring relation. This alternative of comparison algorithm is adapted for spatial generalization by integrating the neighborhood matrix as an input in the comparison algorithm (see Algorithm 1).

Algorithm 1. Comparison algorithm

$$I(s_1, s_2, \dots\dots\dots s_m) = -\sum_{i=1}^{m} \left(\left({}^{si}/_{s} \right) * \log \left({}^{si}/_{s} \right) \right.$$

//(where m= # of classes)

While n>= 1(where n= # of attributes) **do** {

$$= \sum_{j=1}^{v} \left(\left(\left(s_{1j}, s_{2j}, \dots\dots\dots\dots s_{mj} \right) / s * I\left(s_{1j}, s_{2j}, \dots\dots\dots s_{mj} \right) \right) \right)$$

//(where v= # of sub classes)

$$Gain(A1) = I(s_1, s_2, \dots\dots\dots s_m) - E(A1)$$

Min_Gain = Gain(A1)

For each attribute Ai(2≤i≤n) in the generalized relation GR **do{**

$$E(A_i) = \sum_{j=1}^{v} \left(\left(\left(s_{1j}, s_{2j}, \dots\dots\dots s_{mj} \right) / s \right) * I\left(s_{1j}, s_{2j}, \dots\dots\dots s_{mj} \right) \right)$$

$$Gain(A_i) = I(s_1, s_2, \dots\dots\dots s_m) - E(A_i)$$

If Gain(Ai) < Min_Gain **then** Min_Gain = Gain(Ai)

R=Ai

}

Remove R// drop the attribute that correspond to the minimum gain Min_Gain

n=n-1

If corresponding zones are neighbors

then merge identical tuples

Accumulate count%

Create merged zone

}

End

The non-spatial dominant generalization triggers the merge of the connected regions with the same thematic descriptions. The spatial merge generates a set of consolidated identical regions. Some approximation algorithms (such as smoothing or ignoring minor outliers) can be applied, and the generalization can terminate when the number of distinct merged regions reaches a small number defined by a specified ***generalization threshold***.

D. Step 4: Presentation of Driven Comparison

The final result of the process can be driven in different forms like tables, diagrams or even box plots or logic rules. The information in the two classes is used to generate qualitative or quantitative discriminant rules. The generalized relation is transformed into conjunctive normal form, and multiple tuples are transformed into disjunctive normal form. Interesting rules can often be discovered by following different paths leading to several generalized relations for comparison.

The generalization process can be adjusted to the desired levels by performing roll up or drill down operations on the target and contrasting classes. Concept hierarchies can be created by users or domain experts, but also generated automatically based on value distribution or some clustering algorithms.

The induction process described above can be viewed as a data generalization and compression process, which compresses an initial relation into a usually much smaller prime relation, expressed at high concept levels.

V. Case Study

A. Delimitating the Study Area

Our study area is composed of 26 communes of the provinces of Oran. They represent the 26 outbreaks of tuberculosis epidemic previously detected by "EPISOLAP" system. Table II shows the level 0 of Generalization: Gen_0 with 26 zones.

B. Formulation of the Addressed Problem

In Table I, we give a list of the factors considered by studies in tuberculosis epidemiological monitoring. These factors will be analyzed and combined as attributes using the non-spatial dominant generalization method to derive a ranking of factors. We consider in our case study the following relevant attributes: name, incidence_rate[1], density, prec_const, inact_nbr, humidity.

TABLE I. Considered Attributes

N°	CRITERION	Type	Scale	Evaluation Method	Source
1	Incidence rate	Medical	[15,115]	Number of infectious cases	D.H.P[2]
2	The density Of population	Demographic	[630,9660]	Population mean, surface	O.N.S[3]
3	Number of precarious constructions	Demographic	[0,5125]	Number of slums	O.N.S
4	Number of inactive	Demographic	[112,9928]	Level of education, age	O.N.S
5	Humidity	Climatic	[1,4]	Distance according to the sea	O.N.M[4]

[1] It is calculated by the following formula:
Incidence rate=Number of infectious cases * 100000/(Population Mean)

[2] Direction of Health and Population of Oran

[3] Office National of Statistics of Oran

[4] Office National of Meteorology of Oran

C. Experimentation

The objective of the process is to derive the general characteristics of 26 tuples of epidemic outbreaks using analytical characterization to identify the most influential factors on the spread of tuberculosis in Oran. The methodology leads on the classification of factors (or attributes) based on the calculation of the information gain of each considered attribute. Gen (ai) is the level of concept hierarchy and the initial level of generalization is Gen_0. It is the level of generalization before applying any generalization operation and ai is the removed attribute.

Step 1: Data collection

At the beginning, a DMQL query (Data Mining Query Language) is executed in order to collect the task-relevant data (Fig. 4). The objective of the data mining process is to compare high and low level of epidemiological outbreaks using discriminant rules.

```
Use TBC_DB
Mine comparison as "High_vs_Low outbreaks"
In relevance to name, incid_rate, density, Prec_Const,
Inact_Nbr, Humidity
For High_level _of_risck
Where class in high
Versus low_level_of_risck
Where class in low
Analyse cont%
From Outbreaks
```

Fig. 4. Collect the task-relevant data.

The result of the DMQL query is illustrated on Table II.

TABLE II. Example of 26 Tuples of Epidemic Outbreak

N°	Name	Incidence rates	Density of population	Number of precarious constructions	Inactive number	Humidity
1	Oran	37	9660	5125	9928	1
2	Gdyel	44	413	101	760	2
3	Bir El Djir	36	5330	1305	2683	2
4	HassiBounif	49	1976	116	1356	3
5	EsSenia	53	2144	570	1345	2
.....	.....	.....	.....	.....	.....	.....
26	Ain Biya	25	89	5	395	2

After that, we divide our data set into two classes, target class and contrasting class by executing two SQL queries us it is shown on Fig. 5. The target class contains the low level of risk outbreaks and contrasting class contains the high level of risk outbreaks. The results of the two SQL queries are illustrated on Table III.

```
Use TBC_DB
Select name, incid_rate, density,
Prec_Const, Inact_Nbr, Humidity
From Outbreaks
Where class in {Low}
```

```
Use TBC_DB
Select name, incid_rate, density,
Prec_Const, Inact_Nbr, Humidity
From Outbreaks
Where class in {High}
```

Fig. 5. Initial contrasting class relation and initial target class relation.

TABLE III. CANDIDATE RELATION FOR TARGET AND CONTRASTING CLASSES

Name	Class	Inac_rate	Density	Prec_ Const	Inact_Nbr	Hum	Count
A. Turck		31	1005	197	671	1	12
……..	**Low**	……..	……..	……..	……..	……..	……..
Tafraoui		35	63	5	289	4	5
Ain kerma		79	71	0	122	1	6
……..	**High**	……..	……..	……..	……..	……..	……..
S.Chami		55	1661	1282	2324	2	64

Step 2: Dimension relevance analysis

In the second step dimension, relevance analysis is applied by using analytical generalization in order to eliminate irrelevant attributes.

Attribute removal: We start by dropping the less relevant attribute, the name of outbreaks does not have any importance. Then we can drop it; after that we generalize some identical tuples.

Attribute generalization: it consists of discretization of continuous attributes into different categories according to the concept hierarchies of Fig. 6 and accumulate counts

- generalize incidence_rate to incidence_rate_C
- generalize density in density_C
- generalize prec_const in prec_const_C
- generalize inact_nbr in inact_nbr_C
- generalize Humudity in Hum_C

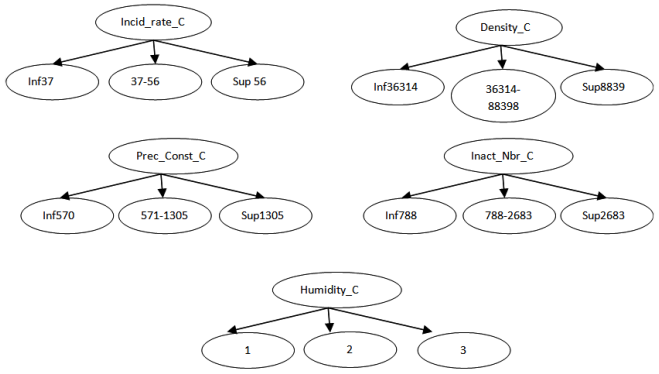


Fig. 6. Conceptual hierarchies.

Candidate relation: incid_rate_C, density_C, prec_const_C, inact_nbr_C and Hum_C.

The resulting outbreaks dataset is illustrated in Table III. Hereafter, name outbreaks will be deleted in order to create a new table (see Table IV), the algorithm processes by looking for identical tuples in order to be merged in new created sectors if only the corresponding zones are neighbors. Therefore, only the last ones are neighbors, then they are merged in new sector S6.

TABLE IV. TABLE OF THE FIRST LEVEL OF GENERALIZATION PROCESS

Name	Class	Incid_ rate_C	Density _C	Prec_ Const _C	Inact_ Nbr_C	Hum_C	Count	Grouped zones
A.Turck		37-56	inf36314	inf570	inf788	1	12	A.Turck
……..	**Low**	……..	……..	……..	……..	……	……	……
Tafraoui		37-56	inf36314	inf570	inf788	3	5	Tafraoui
A.kerma		sup57	inf36314	inf570	inf788	1	6	A. kerma
……..		……..	……..	……..	……..	……	……	……..
Bir El Djir	**High**	Sup57	36314-88398	571-1305	788-2683	2	63	S6
……..		……..	……..	……..	……..	……	……	……..
S.Chami		Sup57	36314 - 88398	571-1305	788-2683	2	64	S6

Step 3 Synchronous Generalization

Synchronous Generalization is applied on target and contrasting classes simultaneously by using the comparison algorithm described in section IV (Algorithm 1). The entropy of each attribute is computed and classified in Table V. The number between brackets is the number of generalization operations executed in each level for both low and high zones.

TABLE V. TABLE OF GAIN AND NUMBER OF MERGING ZONES

Removed attribute	Level	Zones	Low zones	High zones	Gain(Ai)
name	Gen_0	26	16(0)	10(1)	-
Density_C	Gen_1	25	16(0)	9(0)	0,44280177
Inact_nbr_C	Gen_2	24	16(0)	8(1)	0,503416154
Prec_Const_C	Gen_3	23	16(0)	7(1)	0,562716866
Incidence_rate_C	Gen_4	17	10(2)	7(0)	0,593837294
Hum_C	Gen_5	6	3(2)	3(2)	0,601812721
Class_C	Gen_6	2	1(1)	1(1)	-

We start by removing irrelevant/weakly relevant attributes from candidate relation having density_C. The iterative process is executed according to different generalization steps. Each step corresponds to one removal attribute dropped by the order given in Table V. The rest of spatial generalization steps is given in the appendix.

Step 4: Presentation of the driven comparison

According to the forth level of spatial generalization, neighboring zones S6, S7 are merged in S8. The generalized area is formed in the center of the study area because of the same characteristics of merged zones, therefore we can extract the following discriminant rule describing the high level of epidemiological risk:

$$\forall(x) High \leftarrow Inci\_rate\_C(x) \in sup57 \cap Hum\_C \in 2$$

Fig. 7 summarizes the different steps of generalization by giving the removed attribute corresponding to each level of generalization and the geographic representation of the resulting merged zones. The spatial generalization tree is composed of geographic areas which are the administrative divisions grouped in new sectors that are created by the operation of regionalization. The final two sectors are the high sector and the low sector.

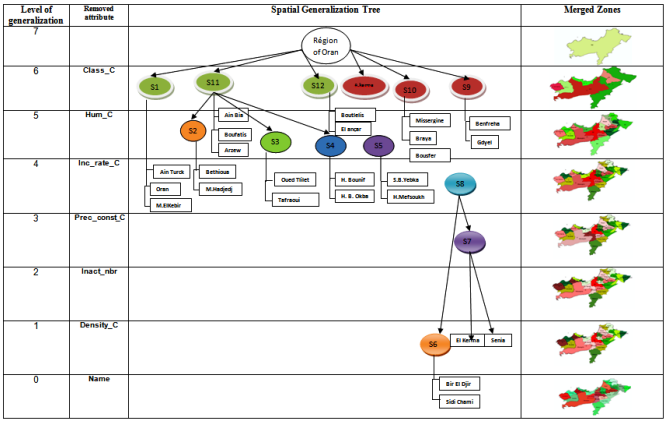


Fig. 7. Spatial Generalization tree.

Geographical summarization of spatial generalization is illustrated in the steps of Fig. 8.

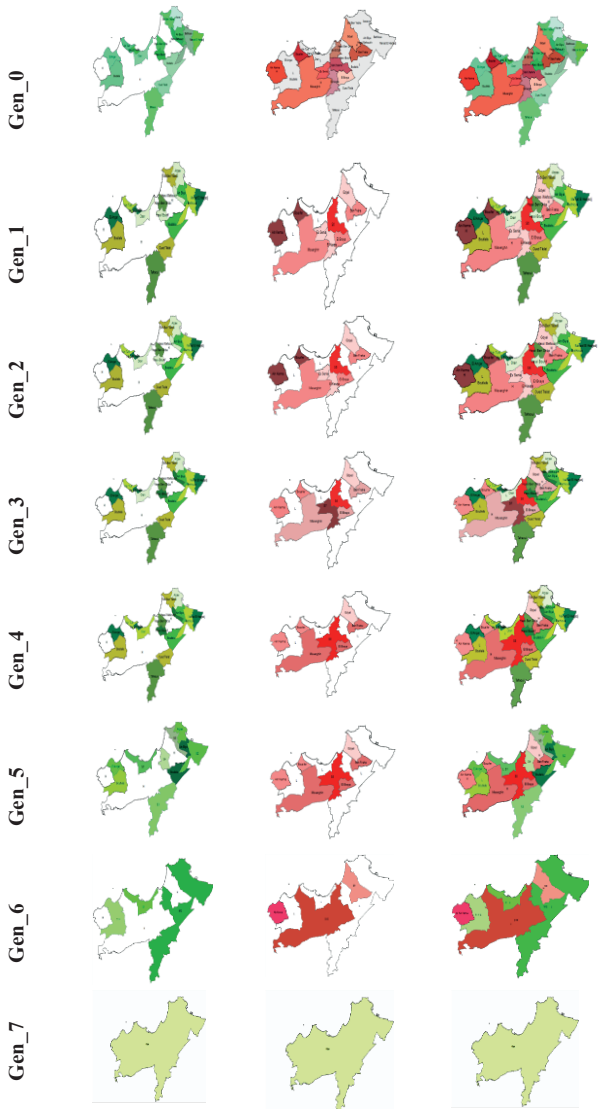


Fig. 8. Visualization of spatial generalization process.

VI. Conclusion

This article presents a new approach SOLAM based on spatial generalization applied according to the general characteristics of locations in lieu of the conventional spatial aggregation in SOLAP systems.

We have adopted the non dominant spatial generalization method based on the attribute-oriented induction method. This method has enriched SOLAP to no longer be limited to a simple aggregation function according to the administrative division of the study area but which may be based on the characteristics of the geographical regions. SOLAM engine helps classifying demographic and socio-environmental factors according to their levels of relevance. Thus, epidemiologists can focus their eradication actions and long-term prevention strategies by eliminating factors in order of relevance. EPISOLAM system has a new feature of sectorization (regionalization) allowing increasing to one level of generalization of geographic areas, whenever an irrelevant attribute is removed from the database.

SOLAM or semantic SOLAP is based on the comparison process performed by AOI algorithm which is tailored in our study by introducing the information gain in the process of identifying the most relevant attributes. This current research studies the primary kernel of this category of decision support systems, which merits further studies to be more comprehensive. The approach can be generalized to various other cases involving spatial analysis problem of multidimensional information. The application of spatial or non-spatial dominant generalization performed by Attribute-Oriented-Induction approach based on multidimensional information for a particular geographic context can guide the application of confirmatory spatial analysis techniques to data related to this zone.

There are many important issues that should be explored in further research. First, there is a question of identifying other measures and rules that may capture patterns in data not already captured by characterization rules. An efficient test for spatial autocorrelation would fix this lack of information. Global and local autocorrelation analyses seem very useful in analyzing space and/or space-time phenomena. That could be introduced by calculating geo-statistical indexes (e.g. Moran and Lisa indexes respectively) in each level of generalization. The experimental results will be reported in the near future. Secondly, we will try to introduce time in the study of autocorrelation by applying geo-statistics techniques to EPISOLAP-MINING system in view to analyze autocorrelation between locations in different periods of epidemic peak. Also, we would like to extend spatial generalization process to a space-time generalization process which means extending AOI algorithm by introducing the neighborhood matrix extended on time as an input of the algorithm instead of only the spatial neighborhood matrix. We hope to explore this idea in a later paper. Finally, the dataset that we have presented in our case study was very small and it was used only for a validation purpose of our proposal. It might be fruitful to explore the behavior of the suggested algorithm when the dataset is very large in terms of number of recorded zones and for long periods of time.

Appendix

The resulting outbreaks dataset is illustrated in Table VI. Hereafter, density_C attribute will be deleted and identical tuples will be merged in order to create the Table VII. However, there is no new identical tuples for contrasting class in current outbreaks dataset and none of identical tuples are neighbors for the target class, therefore we keep the same table VI in table VII.

TABLE VI. Table of the Second Level of Generalization Process

Name	Class	Incid_ rate_C	Density _C	Prec_ Const _C	Inact_ Nbr_C	Hum_C	Count
A.Turck		37-56	inf36314	inf570	inf788	1	12
……..	**Low**	……..	……..	……..	……..	……	……
Tafraoui		37-56	inf36314	inf570	inf788	3	5
Ain kerma	**High**	sup57	inf36314	inf570	inf788	1	6
……..		……..	……..	……..	……..	…….	……..
Gdyel		37-56	inf36314	inf570	inf788	2	17

The resulting outbreaks dataset is illustrated in Table VII. Hereafter, Inact_Nbr_C attribute will be deleted and identical tuples will be merged in order to create the table VIII as follows: There are 2 identical tuples in the contrasting class which are neighbors then they are merged in the new sector S7.

TABLE VII. TABLE OF THE THIRD LEVEL OF GENERALIZATION PROCESS

name	Class	Incid_ rate_C	Prec_Const _C	Inact_ Nbr_C	Hum_C	Count	Grouped zones
A.ElTurck		37-56	inf570	inf788	1	12	A.ElTurck
……..	**Low**	……..	……..	……..	……..	……	……
Tafraoui		37-56	inf570	inf788	3	5	Tafraoui
Ain kerma		sup57	inf570	inf788	1	6	Ain kerma
……..		……..	……..	……..	……..	……	……
El kerma	**High**	sup57	inf570	Inf788	2	13	S7
Essenia		sup57	inf570	788-2683	2	13	S7
Gdyel		37-56	Inf570	Inf788	2	17	Gdyel

The resulting outbreaks dataset is illustrated in Table VIII. Hereafter, Prec_Const_C attribute will be deleted and identical tuples will be merged in order to create the Table IX as follows: There are 2 identical tuples in the contrasting class which are neighbors then they are merged in the new sector S8.

TABLE VIII. TABLE OF THE FORTH LEVEL OF GENERALIZATION PROCESS

name	Class	Incid_ rate_C	Prec_Const _C	Hum_C	Count	Grouped zones
A.ElTurck		37-56	inf570	1	12	A.ElTurck
……..	**Low**	……..	……..	……..	……	……..
Tafraoui		37-56	inf570	3	5	Tafraoui
Ain kerma		sup57	inf570	1	6	Ain kerma
……..		……..	……..	……..	……	……..
S6	**High**	sup57	571-1305	2	127	S8
S7		sup57	inf570	2	57	S8
Gdyel		37-56	Inf570	2	17	Gdyel

TABLE IX. TABLE OF THE FIFTH LEVEL OF GENERALIZATION PROCESS

name	Class	Hum_C	Incid_rate_C	Count	Grouped zones
ATurck		1	37-56	12	S1
Bethioua		1	37-56	8	S2
Arzew		1	37-56	34	Arzew
Oran		1	37-56	288	S1
El Ancor		1	inf37	2	El Ancor
M Kebir		1	inf37	6	S1
M. Hadjaj		1	inf37	4	S2
……..	**Low**	……..	……..	……..	……..
O.Tlelat		3	inf37	3	S3
S.B.yebka		3	inf37	2	S5
Hassi Ben Okba		3	37-56	5	S4
Tafraoui		3	37-56	5	S3
H.Bounif		3	sup57	31	S4
H.Mefsoukh		3	sup57	6	S5
Ain kerma		1	sup57	6	Ain kerma
……..	**High**	……..	……	……..	……..
Gdyel		2	37-56	17	Gdyel

The resulting outbreaks dataset is illustrated in Table IX. Hereafter, Incid_rate_C attribute will be deleted and identical tuples will be merged in order to create the Table X. There are 5 new merged tuples with two or three identical tuples corresponding to neighbor zones. Therefore, they are merged in new sectors S1, S2, S3, S4 and S5.

TABLE X. TABLE OF THE SIXTH LEVEL OF GENERALIZATION PROCESS

name	Class	Hum_C	Count	Grouped zones
S1		1	306	S1
S2		1	12	S11
Arzew		1	34	S11
El Ancor		1	2	S12
Ain El Bia		2	8	S11
Boufatis	**Low**	2	3	S11
Boutlelis		3	6	S12
S3		3	8	S11
S4		3	36	S11
S5		3	8	S11
Ain kerma		1	6	Ain kerma
Bousfer		1	8	S10
Benfreha		3	12	S9
El Braya	**High**	3	5	S10
Misserguine		3	31	S10
S8		2	195	S10
Gdyel		2	17	S9

The resulting outbreaks dataset is illustrated in Table X. Hereafter, Hum_C attribute will be deleted and identical tuples will be merged in order to create the Table XI. There are 4 new merged tuples containing several identical tuples corresponding to neighbor zones. Therefore, they are merged in new sectors S9, S10, S11 and S12.

The resulting outbreaks dataset is illustrated in Table XI. Hereafter, Class_C attribute will be deleted and identical tuples will be merged in order to create a new table. There are tuples with identical tuples but could not be merged in new sectors because they are not neighbors, then we keep Table XI.

TABLE XI. TABLE OF THE SEVENTH LEVEL OF GENERALIZATION PROCESS

Name	Class	Count	Grouped zones
S1		306	S1
S11	**Low**	109	S11
S12		8	S12
Ain kerma		6	Ain kerma
S10	**High**	239	S10
S9		29	S9

REFERENCES

[1] Yvan Bédard, Marie-Josée Proulx, Sonia Rivest. Enrichissement de l'OLAP pour l'analyse géographique: exemples de réalisations et différentes possibilités technologiques. Soumis à la première journée francophone sur les entrepôts de données et analyse en ligne, Lyon, 10 juin 2005.

[2] Bimonte,S. A. Tchounikine, M. Miquel, F. Pinet. (2010). When Spatial Analysis Meets OLAP: Multidimensional Model and Operators. International Journal of DataWarehousing and Mining, 6(4), 2010, pp 33 to 60.

[3] Zemri, F. A., Hamdadou, D & Bouamrane, K. (2013). Towards a Spatio-Temporal Decision Support System for Epidemiological Monitoring: coupling SOLAP and datawarehouse. Accepted paper in the International

Conference on Software Engineering, Databases and Expert Systems (SEDEXS'12). November 13-14, 2013. Settat- Morocco.

[4] Han, J. (1998). Towards on-line analytical mining in large databases. ACM Sigmod Record, 27(1), pp. 97 to107.

[5] Han J., Chee S., Chiang J., (1998) Issues for On-Line Analytical Mining of Data Warehouses, in Proceedings of the 1998 SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery (DMKD'1998), pp. 2 :1–2 :5, Seattle, Washington, USA.

[6] Sonia Rivest, Yvan Bedard & Pierre Marchand. Toward better support for spatial decision making: defining the characteristics of spatial on line abalytical processing (SOLAP). GEOMATICA Vol 55. No. 4, 2001, pp. 539 to 555.

[7] Ganesh Viswanathan & Marcus Schneider. OLAP Formulation for Supporting Complex Spatial Objects in Data Warhouses. 13th International Conference, DaWaK 2011, Toulouse, France, August 29-September 2, 2011. Proceedings (pp.39-50)

[8] Ganesh Viswanathan, Marcus Schneider. BigCube: A MetaModel for Managing Multidimensional Data. In: Proceedings of the 19th Int. Conf. on Software Engineering and Data Engineering (SEDE). (2010). pp. 237 to242.

[9] Maria Luisa Damiani & Stefano Spaccapietra. Spatial Data Warehouse Modelling. This paper appears in the publication, Processing and Managing Complex Data for Decision Support edited by Jerome Darmont and Omar Boussaid © 2006, Idea Group Inc.

[10] Fatiha Amenzougarene. Extension du modèle multidimensionnel aux faitsqualitatifs. Application à l'analyseenligne des gènes des chantiersurbains. Thèse de doctorat de l'université de Versailles Saint Quentin en Yvelines. Présentéetsoutenuepubliquement le 19 septembre 2014.

[11] E. Malinowski & E. Zimanyi. OLAP Hierarchies: A conceptual Perspective. Conference: Advanced Information Systems Engineering, 16th International Conference, CAiSE 2004, Riga, Latvia, June 7-11, 2004, Proceedings.

[12] Ibtisam Ferrahi, Sandro Bimonte, & Kamel Boukhalfa. Conception logique et physique des hierarchies spatiales non strictesdans les entrepôts de données spatiales. Dans le proceeding de SAGEO '2016 Nice, 6-9 décembre 2016.

[13] ChengZhi Xu & Phillip C-Y Sheu. SOLAP Based on Novel Spatial Dimentions. Y. Wu(Ed): Software Engineering and Knowledge Engineering, AISC 114, pp.383-391. Springer-Verglag Berlin Heidelberg 2012.

[14] Torben Bach Pedersen, JunminGu, ArieShoshani, Christian S.Jensen. Object-extended OLAP quessrying. In the journal of Data & Knowledge Engineering, Vol. 68, pp. 453-480, 2009.

[15] Abdallah Bensalloua Charef, Hamdadou Djamila. Users Integrity Constraints in SOLAP Systems. Application in Agroforestry. International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 5, No. 1, pp. 47-56, 2018.

[16] Taher Omran Ahmed. Spatial On-Line Analytical Processing (SOLAP): Overview and Current Trends. 2008 International Conference on Advanced Computer Theory and Engineering.

[17] Eveline Bernier, Yvan Bédard, Thierry Badard, Frédiric Hubert. UMapIT (Unrestricted Mapping Interactive Tool): Merging the datacube paradigm with an occurrence-based approach to support on-demand web mapping. Centre de recherche en geomatiquesnUniversité Laval de Quebec. Canada.

[18] S. Bimonte, A. Tchounikine and M. Miquel "Spatial OLAP: Open Issues and a Web Based Prototype" In the Proc. of 10th AGILE International Conference on Geographic Information Science, 2007, Aalborg University, Denmark.

[19] Han, J. 1997. OLAP Mining: An Integration of OLAP with Data Mining. In Proceedings of the 7th IFIP 2.6 Working Conference on Database Semantics (DS-7), pp. 1 to9.

[20] Han, Jiawei. And Yongjian Fu, Y. 1995. Exploration of the power of attribute-oriented induction in data mining. Simon Fraster University. In U. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy, eds. Advances in Knowledge Discovery and Data Mining, pp.399 to421

[21] Jiawei Han, YandongCai and Nick Cercone (1992). Knowledge Discovery in Databases: An Attribute-Oriented Approach. in proceedings of the 18th VLDB Conference. Vancouver, British Columbia, Canada.

[22] S. Warnars, Measuring Interesting rules in characteristic rule. Proceeding of the 2nd International Conference on Soft Computing, Intelligent System and Information Technology (ICSIIT), Bali, Indonesia, (2010) pp. 152 to156.

[23] Han J., Koperski K., and Stefanovic N. : GeoMiner : A system Prototype for Spatial Data Mining, Proc. 1997 ACM-SIGMOD IntConf on Management of Data (SIGMOD'97), Tucson, Arizona, System prototype demonstration, 1997.

[24] Roddick J.F and Brian G. L. Spatio-Temporal Data Mining Paradigms and Methodologies in Geographic Data Mining and Knowledge Discovery. Second edition. 2009.

[25] W. Lu, J. Han, and B. C. Ooi. Discovery of general knowledge in large spatial databases. Proc. Far East Workshop on Geographic Information Systems, pp. 275-289, Singapore, June 1993.

[26] E. M. Knorr and R. Ng. Extraction of spatial proximity patterns by concept generalization. Proc. 2nd Int. Conf Knowledge Discovery and Data Mining (KDD-96), pp. 347-350, Portland, OR, USA, August 1996.

[27] Y. Bédard, Amélioration des capacités décisionnelles des SIG par l'ajout d'un module SOLAP. Université de Provence, Centre de Mathématiques et Informatique, LSIS, Marseille, 8 avril.2004.

[28] Zemri et al, 2015. Vers un système d'Aide à la Décision Multicritères et Spatiotemporel pour la Surveillance Epidémiologique. Accepted paper in the Extraction et Gestion des Connaissances, EGC'2015. Actes de l'atelier GAST – Gestion et Analyse de données Spatiales et Temporelles. January 27 2015. Luxembourg.

[29] Zemri Farah Amina & Hmdadou Djamila. Integration of Data Mining Techniques in Multi Criteria Spatial Decision Support System for Epidemiological Monitoring. Accepted paper in International Journal of Healthcare Information Systems an Informatics (IJHISI). June 2017.

[30] Han, Jiawei. And Yongjian Fu, Y. 1995. Exploration of the power of attribute-oriented Induction in data mining. Simon Fraster University. In U. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy, eds. Advances in Knowledge Discovery and Data Mining, pp. 399 to421.

[31] Warnars. S. Mining Patterns with Attribute Oriented Induction. In the Proceedings of the International Conference on Database, Data Warehouse, Data Mining and Big Data (DDDMBD), Jakarta, Indonesia 2015.

Farah Amina Zemri

Phd Student in computer sciences. She had her master in Diagnostic, Decision making and Interactive interfaces in 2011. She is specialized in Spatio temporal decision making systems, spatial datamining techniques and Geographical Information Systems.

Djamila Hamdadou

Specialised in Decision Support Systems, Multicriteria Analysis and Collaborative and spatial decisional Systems. She is professor and leads a research team at the laboratory of computer science of Oran (LIO) at the University of Oran1 in Algeria.

Genetic-Moth Swarm Algorithm for Optimal Placement and Capacity of Renewable DG Sources in Distribution Systems

Emad A. Mohamed[1,2]*, Al-Attar Ali Mohamed[2], Yasunori Mitani[1]

[1] Department of Electrical Engineering, Kyushu Institute of Technology, 1-1 Sensui-cho, Tobata-ku, Kitakyushu-shi, Fukuoka 804-8550 (Japan)
[2] Department of Electrical Engineering, Aswan University, 81542 Aswan (Egypt)

Received 13 August 2019 | Accepted 19 November 2018 | Published 30 October 2019

Abstract

This paper presents a hybrid approach based on the Genetic Algorithm (GA) and Moth Swarm Algorithm (MSA), namely Genetic Moth Swarm Algorithm (GMSA), for determining the optimal location and sizing of renewable distributed generation (DG) sources on radial distribution networks (RDN). Minimizing the electrical power loss within the framework of system operation and under security constraints is the main objective of this study. In the proposed technique, the global search ability has been regulated by the incorporation of GA operations with adaptive mutation operator on the reconnaissance phase using genetic pathfinder moths. In addition, the selection of artificial light sources has been expanded over the swarm. The representation of individuals within the three phases of MSA has been modified in terms of quality and ratio. Elite individuals have been used to play different roles in order to reduce the design space and thus increase the exploitation ability. The developed GMSA has been applied on different scales of standard RDN of the (33 and 69-bus) power systems. Firstly, the most adequate buses for installing DGs are suggested using Voltage Stability Index (VSI). Then the proposed GMSA is applied to reduce real power generation, power loss, and total system cost, in addition, to improve the minimum bus voltage and the annual net saving by selecting the DGs size and their locations. Furthermore, GMSA is compared with other literature methods under several power system constraints and conditions, in single and multi-objective optimization space. The computational results prove the effectiveness and superiority of the GMSA with respect to power loss reduction and voltage profile enhancement using a minimum size of renewable DG units.

Keywords

Radial Distribution System, Renewable Distributed Generation, Power Loss, Genetic-Moth Swarm Algorithm.

DOI: 10.9781/ijimai.2019.10.005

I. Introduction

In recent years, the continuous development of electrical loads, especially due to industrial plants and human activities, results in increased number of new transmission lines, power plants, distribution networks and interconnection between different power systems. This effect leads to higher currents and power losses accompanied by voltage drop. Distribution system (DS) is an essential part of this power system problem as it connects loads to the transmission line at substations. About 70% of the power system losses are occurring at distribution system [1]. Therefore, the reduction of the losses in DS is the main concern nowadays. Hence, the world directed to use new generation sources of renewable energy resources (RERs) such as photovoltaic (PV), wind turbines and biomass energy, which are considered economically for supplying energy to electrical grids and suitable for power generation in remote areas [2], [3]. There are many potential benefits of DGs depending on their size and location. Normally, the real power loss and the voltage profile are the base objectives. Some other technical parameters may accompany this base objective such as reactive power requirement, reliability and efficiency of distribution network, emission, load-ability, voltage stability, DG capacity maximization, or economy oriented objectives [4]. There are different types of DG units, which can be classified based on whether they generate or absorb reactive power along with active power generation to (a) type A-DG units or P-type, which produces active power only such as PV (b) type B-DG units or Q-type, which produces reactive power only, like capacitor banks (c) type C-DG units or PQ+-type, which produces active and reactive power like synchronous generators (d) Type D-DG units or PQ--type which produces active power and consumes reactive power, like wind power induction generators.

The random placement of DGs and capacitors in DS can cause more voltage drop and higher power losses than losses without them [5], [6]. Therefore, determining the proper placement and capacity of DGs in DS becomes a crucial task for obtaining their maximum possible advantages. Several techniques have been proposed in recent years to determine the optimal locations and sizes of DGs in DS such as Ref. [7], which discussed the adaptive protection using neural networks for high penetration of DGs, but this technique takes very long training time. Ref. [8] made a very hard work to get the effective signals for optimal ratings of RERs as the objective function and

* Corresponding author.
E-mail address: q595506a@mail.kyutech.jp

constraints are designed using fuzzy sets. In Ref. [9] authors discussed the achievement of the trade-off between the reliability improvement and DG capacity by examining the load shedding.

Recently, numerous optimization algorithms handled the problem of DGs locations and sizing in DS. Artificial bee colony (ABC) [10], Genetic Algorithm (GA) [11], cuckoo search algorithm (CSA) [12], mixed integer nonlinear programming (MINLP) [13], Differential Evolution (DE) [14], flower pollination algorithm (FPA) [15]. Although these heuristic algorithms have been implemented simply and free derivative, they need numerous iterations to guarantee that the solution is converged. Hence, these techniques are computationally intensive. Furthermore, some studies used hybrid algorithms with analytical to combine their features and eliminate the shortage like, simulated annealing uses Loss Sensitivity Factor (LSF) in [16], PSO uses sensitivity analysis in [17], and hybrid PSO in [18]. There is another type of hybridization, which is combining metaheuristic algorithms together such as, genetic algorithm (GA) with imperialist competitive algorithm [19], ant colony optimization with artificial bee colony (HACO) [20], hybrid grey wolf optimizer (HGWO) [21], backtracking search optimization algorithm (BSOA) [22], and in [23], which used particle ant bee colony with harmony search (PABC). Other studies used the analytical approach such as in [24], which uses efficient analytical with optimal power flow (EA-OPF), an improved analytical (IA) method in [25] and machine learning method in [26]. In addition to Naresh, who used an analytical expression for optimum location for DG [27]. Most of previous techniques use a simple single objective function for minimizing the power losses except [19], [21], [23] that use a multi-objective functions to reduce real losses and improve voltage stability. Further, only few methods deal with the renewable DGs like in [20], [22], ant lion optimization (ALO) algorithm in Ref. [28], and backtracking search (BSA) algorithm in Ref. [29]. The above mentioned algorithms seem to be efficient. However they may not guarantee reaching the optimal value and face difficulty in escaping from the local minimum as the power losses face nonlinear equality constraints. This makes the problem non-convex.

A new hybrid GMSA is developed based on the incorporation of GA operations with adaptive mutation operator on the reconnaissance phase using genetic pathfinder moths and the expanding of artificial light sources over the swarm. The GMSA has some advantages over the other swarm algorithms such as (i) its simplicity and flexibility as it can be applied to different problems without changing the main algorithm structure. (ii) ability on avoiding the trap in local minima. (iii) achieving fast convergence characteristics [30]. Ref. [30] determined the optimal sizes and locations of DGs without considering the different types of DGs. In this paper, three types of DG units including PV, WTG, and capacitor bank based DGs are embedded in distribution system optimally for minimizing the power losses. A sensitivity analysis based-Voltage Stability Index (VSI) has been performed to determine the most candidate locations for inclusion the compensation devices in DS to reduce the search space of optimization techniques and simulation time. Then, the hybrid approach based on the genetic algorithm (GA) and moth swarm algorithm (MSA) [31], is presented to determine the optimal renewable DG capacity and locations in the DS to minimize the system power losses, and maintain the voltage profile for various electrical distribution systems. It is tested on standard distribution systems i.e., (33 and 69 -bus). In addition, the obtained results from the proposed approach are compared with those obtained from other algorithms to confirm its superiority. The article is organized as follows; section II provides the objective function formulation. GMSA algorithm is represented in section III. In section IV, the implementing of GMSA code for solving the DGs allocation problem has been presented. Section V shows the numerical results of the proposed technique applied on multiple standard systems. The last section concludes the results and advantages of the proposed method.

II. Problem Formulation

A. Load Flow Calculation

Radial distribution networks (RDN) creates some negative conditions such as radial meshed networks, unbalanced operation, high R/X ratios and distributed generation. Due to these problems, the Newton Raphson, Gauss Siedel and other conventional load flow algorithms are not effective to solve the load flow calculation of the distribution systems [32]. Therefore, the modern algorithm called backward/forward sweep [32] is used in this work to analyze the power flow in the tested IEEE distribution systems. Fig. 1 shows a single line diagram of RDN. The active power flow (P_{k+1}) and reactive power flow (Q_{k+1}) in RDN including DG unit at bus (k+1) are calculated by (1) and (2):

$$P_{k+1} = P_k - P_{L(k+1)} - R_k \frac{P_k^2 + Q_k^2}{V_k^2} + P_{DG} \tag{1}$$

$$Q_{k+1} = Q_k - Q_{L(k+1)} - X_k \frac{P_k^2 + Q_k^2}{V_k^2} + Q_{DG} \tag{2}$$

where, k is the sending end and $k+1$ is the receiving end. Voltages of a transmission line and real power losses in the line can be calculated from (3), (4), and (5) respectively:

$$V_{k+1}^2 = V_k^2 - 2(R_k.P_k + X_k.Q_k) + (R_k^2 + X_k^2).\frac{(P_k^2 + Q_k^2)}{V_k^2} \tag{3}$$

$$P_{loss(k,k+1)} = R_k * \frac{(P_k^2 + Q_k^2)}{V_k^2} \tag{4}$$

$$Q_{loss(k,k+1)} = X_k * \frac{(P_k + Q_k)}{V_k^2} \tag{5}$$

The total system loss is calculated by summing all line losses in the system as shown in (6):

$$P_{Tloss} = \sum_{k=1}^{n-1} P_{loss(k,k+1)} \tag{6}$$

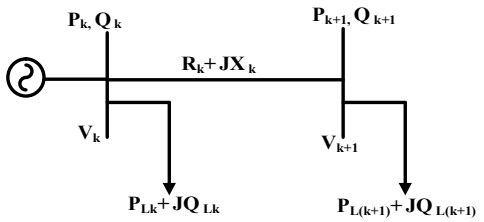


Fig. 1. Simple radial distribution system.

The system security level is important and can be determined using the voltage stability index as follows:

$$VSI_{(k+1)} = |V_k|^4 - 4(P_{k+1}X_k - Q_{k+1}R_k)^2 - 4(Pk_{m+1}X_k + Q_{k+1}R_k)|V_k|^2 \tag{7}$$

VSI should be high in order to improve the voltage profile and this can be achieved by minimizing the voltage deviations (VD) as follows:

$$VD = \sum_{k=1}^{n} (V_m - V_{ref})^2 \tag{8}$$

where, n is the number of buses and V_{ref} is the reference voltage, which commonly is 1 p.u.

B. Objective Functions

The main aim of the optimal DG placement problem is to minimize the voltage deviation, reduce the real power losses and improve the system voltage stability. There is a contrasting relation between these objectives, as clearly identified and numerically obtained by [33]. Hence, the multi-objective functions have been performed by using the following mathematical statements:

$$f_1 = \frac{\sum_{k=1}^{n_l} (P_{loss}(k))_{After\ \ DG}}{\sum_{k=1}^{n_l} (P_{loss}(k))_{Before\ \ DG}} \tag{9}$$

$$f_2 = \frac{\sum_{k=1}^{n_b} (VD)_{After\ \ DG}}{\sum_{k=1}^{n_b} (VD)_{Before\ \ DG}} \tag{10}$$

$$f_3 = \frac{1}{\sum_{k=1}^{n_b} (|VSI(k)|)_{After\ \ DG}} \tag{11}$$

where, n_l is number of branches in RDN and n_b is number of buses. The weighted sum method is used to evaluate the effectiveness of the proposed approach for optimal placement and sizing of DG units. The concept Pareto strategy is not appropriate for such purpose, where the challenge in multi-objective optimization based on Pareto strategy is to find the Pareto optimal point that meets the decision maker's given preferences. From the perspective of mathematical optimization, the weighted sum method allows the multi-objective to be cast as a single-objective mathematical optimization problem resulting in only one solution, in addition to its lower computational cost (CPU-time). These advantages are more proper for real world problems. Therefore, the generalized objective function based on weighted sum method can be formulated as follows:

$$f_t = w_1 f_1 + w_2 f_2 + w_3 f_3 \tag{12}$$

where, w_1, w_2, and w_3 are weighting factors. The value of any weighting factor is selected based on the relative importance on the related objective function with others objective functions. The sum of the absolute values of the weight factors in (12) subjected to all impacts should equal one:

$$|w_1| + |w_2| + |w_3| = 1 \tag{13}$$

C. Constraint Conditions

The multi-objective functions are subjected to the following constraints:

1. Active and Reactive Power Balance

The active and reactive power flow constraints, which represent the equality constraints could be established for maintaining the balance between generation and consumption.

$$P_{sys} + \sum_{k=1}^{N_{DG}} P_{DG}(i) = \sum_{k=1}^{n} P_d(k) + \sum_{j=1}^{n_b} P_{Tloss}(j) \tag{14}$$

$$Q_{sys} + \sum_{i=1}^{N_{DG}} Q_{DG}(i) = \sum_{i=1}^{k} Q_d(i) + \sum_{j=1}^{n_b} Q_{Tloss}(j) \tag{15}$$

2. Voltage Constraints

The buses voltages are the inequality constraints. The bus voltage magnitude of each bus must be maintained within the following range:

$$V_{\min} \le V_k \le V_{\max} \tag{16}$$

where V_{max} and V_{min} are the maximum and minimum values of bus (k) voltages. The lower and upper values are taken as 0.9 and 1.05 Pu, respectively.

3. DG Capacity Limits

The constraints of DG capacities are as follows:

$$\sum_{k=1}^{N_{DG}} P_{DG}(k) \le \frac{3}{4} \times \left(\sum_{k=1}^{n} P_d(k) + \sum_{j=1}^{n_b} P_{loss}(j) \right) \tag{17}$$

$$\sum_{k=1}^{N_{DG}} Q_{DG}(k) \le \frac{3}{4} \times \left(\sum_{k=1}^{n} Q_d(k) + \sum_{j=1}^{n_b} Q_{loss}(j) \right) \tag{18}$$

$$P_{DG}^{\min} \le P_{DG}^{k} \le P_{DG}^{\max} \tag{19}$$

$$pf_{DG}^{\min} \le pf_{DG}^{k} \le pf_{DG}^{\max} \tag{20}$$

$$Q_{DG}^{\min} \le Q_{DG}^{k} \le Q_{DG}^{\max} \tag{21}$$

where, $P_{DG}^{\min}$ and $P_{DG}^{\max}$ are the minimum and maximum real outputs of the DG source. $pf_{DG}^{\min}$ and $pf_{DG}^{\max}$ are the minimum and maximum power factor of the DG source.

The input control vector x_c is composed of independent adjustable variables for each DG units. Each DG has three input control variables: location (L), power factor (PF) and injecting active power (P_{DG}). Multiple DG units can be installed in a system as follows:

$$X_c = \begin{bmatrix} L_{C1}, P_{C1}, \dots, L_{C_{N_C}}, P_{C_{N_C}}, L_{DG1}, P_{DG1}, PF_{DG1}, \dots, \\ L_{DG_{N_{DG}}}, P_{DG_{N_{DG}}}, PF_{DG_{N_{DG}}} \end{bmatrix} \tag{22}$$

In this paper, in the case of capacitor banks, the PF is zero and for PV units, the PF is considered to be unity thus the DG unit only delivers active power. While, in the case of wind, the DG unit delivers active and reactive power.

D. Equality and Inequality Constraints Treatment

Power-flow equations, equality constraints (14) and (15), can be satisfied during the process of power-flow calculation. In the encoding period, the inequality constraints (16)–(21) can be satisfied through adding penalty function into the objective function in such a way that it penalizes any violation of the constraints. Consequently, the constrained optimization problem is then converted into an unconstrained form.

III. Overview of GMSA

A. Genetic Algorithms

Genetic algorithms (GAs) were initially introduced by John Holland as the main global optimization technique. These algorithms

have been applied successfully to solve a large number of problems in different real-world fields by simulating the natural evolution systems. The recombination operation produces offspring that carry a combination of genetic material information from each parent where crossover operations are applied to exchange the chromosomes. The natural selection determines the evolution where the fittest survives with higher probability. Therefore, a suitable selection strategy is then used to determine the solutions that survives to the next generation based on their fitness values. The mutation operation is the main genetic operator that can achieve some diversity in the population.

The steps of the MSA technique are discussed below.

B. Moth Swarm Algorithm

The moth swarm algorithm has been presented in 2017 by Al-Attar et al. [30]. It is inspired by the orientation of moths towards moonlight. The available solution of an optimization problem using MSA is performed by the light source position, and its fitness is the luminescence intensity of the light source. Furthermore, the proposed method consists of three main groups, the first one is called pathfinders which are considered a small group of moths over the available space of the optimization. The main target of this group is to guide the locomotion of the main swarm by discriminating the best positions as light sources. Prospectors group is the second one which has a tendency to expatiate in a non-uniform spiral path within the section of the light sources determined by the pathfinders. The last one is the onlookers, this group of moths move directly to the global solution which has been acquired by the prospectors.

C. Genetic Moth Swarm Algorithm (GMSA)

The proposed hybrid based algorithm aims to integrate advantages of the well-known GA in terms of sharing information and global search ability to find the optimal value of a given function using the following steps:

1. Initialization

Initially, the positions of moths are randomly created for dimensional (d) and population number (n) as seen in (23).

$$x_{ij} = rand[0,1].(x_j^{\max} - x_j^{\min}) + x_j^{\min} \quad \forall i \in \{1,2,...,n\}, j \in \{1,2,...,d\} \tag{23}$$

where, $x_j^{\max}$ and $x_j^{\min}$ are the upper and lower limits, respectively.

Afterwards, the type of each moth is selected based on the determined fitness. Consequently, the worst moths are selected as pathfinders that are modified to act genetically in the following reconnaissance phase. In the next two phases, the best individuals of the swarm are regarded as prospectors and onlookers, respectively, according to their fitness. In addition, each moth in the modified algorithm has its own light source which is available to share with others in the swarm.

2. Genetic Reconnaissance Phase

The moths may be concentrated in the regions which seem to provide good performance. Therefore, the swarm quality for reconnaissance may be decreased during processing the optimization and this process may lead to a stagnation case. To avoid the early convergence and enhance the solution diversity, a part of the swarm is compelled to determine the less congested area. The pathfinder moths that perform this role are manipulated to evolve by the genetic operators, with the size of (n_p=*floor* (n/2)) selected from the worst-performing individuals in the swarm. The crossover and mutation operators of GA are applied on all moths in the swarm to improve the pathfinder group. Therefore, after the sorting of the population, the first half of the individuals that have better luminescence intensity values are regarded as candidate parents (elite individuals). The size of the elite individuals can be simply calculated using (n_e=n-n_p).

The probability distribution function (pdf) is used to select parents, which is increased as the fitness of the individual is greater. Therefore, two of the moths from the elite individuals are randomly selected as parents for one pathfinder moth. In order to perform the possible mating, a single crossover point is identified on both parents' vectors at random. The elite individuals are then divided at this point to exchange their tails thereby giving birth to the new child pathfinder (x_p). This ensures that the best candidates (local optima) are copying into the next generation. After the reproduction operation, a mutation operator based on normal distribution is applied on these offspring in order to increase their diversity and increase the ability to jump out of suboptimal/local solutions. For exploitation purpose, an adaptive mutation rate (m_{rate}) is proposed to decrease through all iterations T as follows:

$$m_{rate} = 0.05.(1 - t/T) \tag{24}$$

The fitness value of the genetic pathfinder solution, x_p^{t+1}, is determined after finishing the last procedure. The structure of worst half of the old population is then redesigned by comparing the fitness of these offspring with that of their previous positions $f(x_p^t)$. The suitable solutions that have the highest luminescence intensity are chosen to retain for the next generation, which is used for minimization problems as follows:

$$\overrightarrow{x_p^{t+1}} = \begin{cases} \overrightarrow{x_p^t} & if \quad f(\overrightarrow{x_p^{t+1}}) \geq f(\overrightarrow{x_p^t}) \\ \overrightarrow{v_p^t} & if \quad f(\overrightarrow{x_p^{t+1}}) < f(\overrightarrow{x_p^t}) \end{cases} \tag{25}$$

Finally in this phase, the light sources are elected from among the combined population (survivors of the previous equation and their parents) to continue as guidance of the next phases. Therefore, the moths are changed dynamically in the GMSA model where any pathfinder moth uplifts to become prospector or onlooker moth if it discovers a solution with more luminescence than the existing light sources. That means the new lighting sources will be presented at the end of this stage. The probability p_i of selecting the i_{th} moth as a light source is proportional to the corresponding fitness, which can be calculated as follows:

$$p_i = \frac{f(x_i)}{\sum_{i=1}^{n} f(x_i)} \qquad \forall i \in \{1,2,...,n\} \tag{26}$$

3. Transverse Orientation

Individuals that have been selected as elites or parents have another role at this stage as prospectors. The number of these moths n_f is proposed to decrease with time progress as follows:

$$n_f = round\left((n - n_p) \times (1 - \frac{t}{T})\right) \tag{27}$$

After the pathfinders have finished their search, the information about luminescence intensity is shared with prospectors, which attempt to update their positions in order to discover new light sources. Each prospector moth X_j is soared into the logarithmic spiral path as shown in Fig. 2(a) to make a deep search around the corresponding artificial light source X_i, which is chosen on the basis of the probability P_i using (26). The new position of jth prospector moth, can be expressed mathematically as follows:

$$x_j^{t+1} = \left|x_j^t - x_i^t\right|.e^{\theta}.\cos 2\pi\theta + x_i^t \quad \forall\ j \in \{1,2,...,n_f\};\ \forall i \in \{1,2,...,n\} \tag{28}$$

where, $\theta \in [r, 1]$ is a random number to define the spiral shape and

(r = -1-t/T). The GMSA is dealing with each variable according to the previous formula as an integrated unit. At the end of this stage, only moonlight is updated. It should be noted that all moths in the modified swarm cooperate to discover new sources of light, which increases the area of selection and prevents from falling into local solutions and thus increases the efficiency of the proposed algorithm.

4. Celestial Navigation

The diminishing of the number of prospectors during the optimization process increases the onlookers number (n_o=n_e-n_f). This may lead to an increase in the speed of the convergence rate of GMSA towards the global solution. The onlookers are the moths that have the lowest luminescent sources in the parent group. Their main aim for traveling directly to the moon is the most shining solution Fig. 2(b). In the GMSA, the onlookers are forced to search for the hot spots of the prospectors effectively. These onlookers are divided into the two following parts:

The first part, with the size of n_G=*round* (n_o/2), walks according to Gaussian distributions. The new onlooker moth in this sub-group $x_i^{(t+1)}$ moves with series steps of Gaussian walks, which can be described as follows:

$$x_i^{t+1} = x_i^t + \varepsilon_1 + \left[\varepsilon_2 \times best_g^t - \varepsilon_3 \times x_i^t\right] \quad \forall i \in \{1,2,...,n_G\} \tag{29}$$

$$\varepsilon_1 \sim random(size(d)) \oplus \quad N\left(best_g^t, \frac{\log t}{t} \times (x_i^t - best_g^t)\right) \tag{30}$$

Where ε_1 is a random number generated from Gaussian distribution, ε_2 and ε_3 are random samples drawn from a uniform distribution within the interval [0,1]. $best_g$ is the global best solution (moonlight) obtained in the transverse orientation phase. Based on many optimization algorithms, there is a memory to transfer information from the current generation to the next generation. However, the moths may fall into the fire in the real world due to the lack of an evolutionary memory. This is due to the fact that the performance of moths is intensely affected by the short-term memory and the associative learning between the moths. Therefore, the second part of onlooker moths (n_A=n_e-n_G) will sweep toward the moonlight using associative learning immediate memory (ALIM) to imitate the actual behavior of moths in nature. The instantaneous memory is initialized from the continuous uniform of Gaussian distribution on the range from (x_i^{min} - x_i^t) to (x_i^{max} - x_i^t). The updating equation of this type can be completed in form:

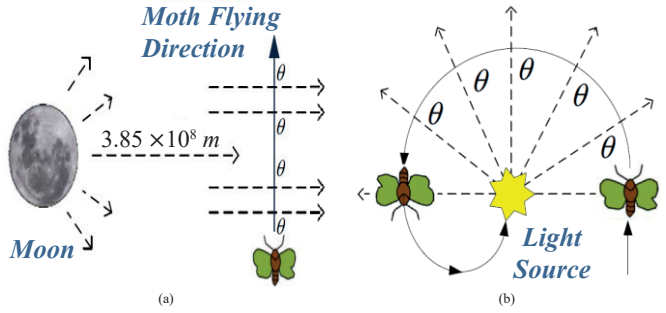


Fig. 2. Orientation behavior of moth swarm: (a) Moth flying in a spiral path into nearby light source (b) Moth flying in a fixed angle relative to moonlight.

$$x_i^{t+1} = x_i^t + 0.001.N\left[x_i^{min} - x_i^t, x_i^{max} - x_i^t\right] + (1 - \frac{g}{G}).r_1.(best_p^t - x_i^t) + 2g/G.r_2.(best_g^t - x_i^t) \; \forall i \in \{1,2,...,n_A\}$$

Where, r_1 and r_2 are random numbers within the interval [0, 1], 2g/G is the social factor, (1-g/G) is the cognitive factor and $best_p$ is a light source selected from the modified swarm based on the probability p_i.

IV. Results and Discussion

To evaluate the validity and efficiency of the proposed GMSA method against power loss minimization, the distribution systems of 33 and 69-bus have been applied for this simulation. The MATLAB 8.6 ® is used and run on a personal computer that has core i5 processor, 2.50 GHz, and 4 GB RAM to implement the GMSA technique for the optimal renewable DGs placement and sizing problems. The backward/forward sweep load flow program is used to solve the equations iteratively and update the voltage profile. The parameters of the GMSA are adopted after many trails and errors for all the test cases of RDNs mentioned in Table I appendix (A). Three types of DG units including P-type, Q-type, and PQ--type are considered in this study. Each type is applied to the three cases of one DG, two DG, and three DG units. The GMSA is compared with all other types of algorithms such as analytical, metaheuristic methods, classical, and hybrid approaches.

A. 33-Bus Test System

To evaluate the impact of the proposed hybrid GMSA on the medium network of the RDN, the 33-bus system has been tested. Fig. 3 shows the single line diagram of this system. The system rated voltage is 12.66 kV with 100 MVA base. The total real and reactive power demands are 3,715 kW and 2,300 kVAR respectively. The load and line data are given in [34]. Load flow calculation is run before using DG units, the minimum bus voltage is registered as 0.9036 p.u at node 18 and the total active power loss at nominal load is 210.98 kW. The best locations and sizes of the three types of PV, WTG, and capacitor banks that are captured by the proposed GMSA, and all obtained results, are listed in Table II appendix (A), such as, total power loss, minimum bus voltage, VD, VSI, and loss reduction percentage.

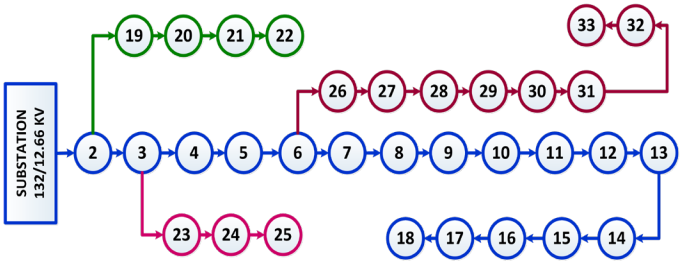


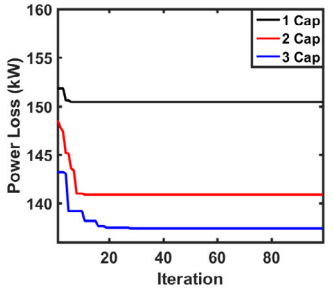
Fig. 3. Single line diagram of the 33-bus RDN [34].

1. Case 1: Q-Type DG

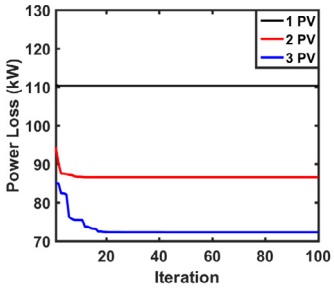
In this case, Table III appendix (A) shows a comparison of different algorithms for the Q-type with three cases of 1DG, 2DGs, and 3DGs. For single DG unit, most methods selected bus 30 as the best location for the DG unit with different size values. GMSA produced better solutions, whereas the real power loss is diminished to 150.43 kW as 28.7% of the base case by using 1200 kVAR of capacitance. It is considered the lowest value compared to the other methods. Moreover, The GMSA has increased the minimum voltage value to 0.9175 p.u after compensation. For two DG units allocation, the optimal locations selected by the GMSA method are buses 12 and 30 with 450 and 1050 kVAR, respectively. The reduction in power losses is 140.87kW, which is the best comparing with 141.83kW for HGWO, 141.94kW for Hybrid method, 143.11kW for MSA, and 151.12 kW with GA. Furthermore, the minimum bus voltage enhanced from 0.9175 p.u with one DG to 09332 p.u with two DGs. In the case of determining three DG units, the GMSA performed well as the active power loss is diminished to 137.46 kW. This value is the least of all other methods as seen in Table III. In addition, the system voltage profile is improved and the worst bus voltage is enhanced to 0.9334 p.u. It is shown that the proposed GMSA is more effective than the other approaches in case of Q-type DG under the medium scale of distribution system.

2. Case 2: P-Type DG

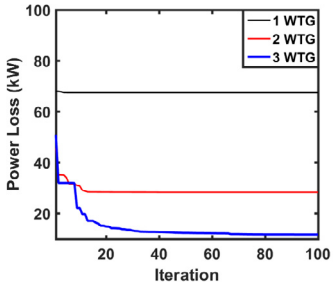
Table IV appendix (A) shows the optimal locations and capacities of PV units by the proposed GMSA method compared to different algorithms for the same three cases (1 DG, 2 DGs, and 3 DGs). The GMSA presented the best solutions as the power losses reduced to 110.267 kW with only one PV unit installed at bus 6. This value is considered the best value comparing with other techniques and also better than the all three cases of Q-type DG. Moreover, it enhanced the voltage profile as the minimum bus voltage at bus 18 increased from 0.9036 p.u to 0.9427 p.u. For 2 PV units case, the optimal bus places are at 13 and 30 for most methods. However, the proposed technique reduced the active power loss to 86.58 kW compared to 87.17 kW with PSO, 87.164 kW with HGWO, and 87.172 with EA. It is also observed from the results that the VD is enhanced to be 0.6723 p.u while the voltage stability is also enhanced to be 29.4035 p.u. In the last case of installing three PV units, GMSA selected buses 13, 24, and 30 to locate the PV units with 801, 1091, 1053 kW, respectively, which helps in reducing the real power loss to 72.299 kW and increasing the minimum bus voltage to 0.9712 p.u.



(a)



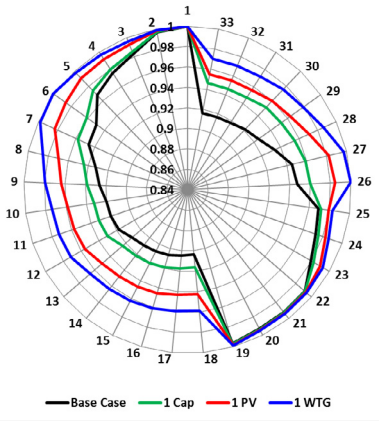
(b)



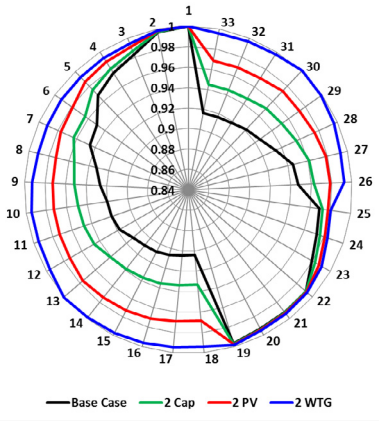
(c)

Fig. 4. Power loss convergence rate of 33-bus system using GMSA for different DG types (a) Capacitor banks, (b) PV, (c) WTG.

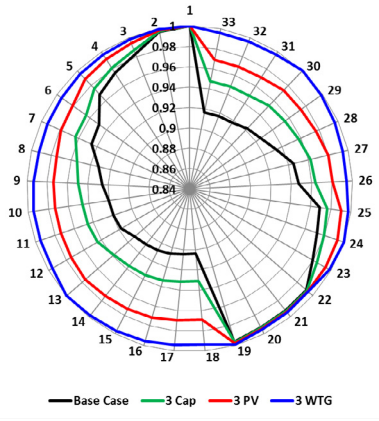
These results prove the superiority of the proposed GMSA compared to other methods as seen in Table IV. It is also observed that the power loss is minimized significantly as compared with the Q-type DGs, this helps in improvement in the system voltage profile.



(a)



(b)



(c)

Fig. 5. 33-Bus voltage profile level with (a) Single DG (b) Two DGs (c) Three DGs.

3. Case 3: PQ--Type DG

Table V appendix (A) shows the results for the PQ+ -type such as, wind turbine generators (WTG), which injects both active and reactive power for the three cases. The results stated that this type of DGs is the best as only one WTG installed at bus 6 with the size of 3105 kVA and 0.82 pf using the GMSA. This one unit improved the minimum bus voltage value from 0.90369 p.u to 0.9586 p.u and reduced the total power loss to about 67.42kW, which is considered the least value compared to the three capacitor banks and the three PV units. Furthermore, it is the best compared with other algorithms as seen in Table V. Moreover, increasing the number of WTGs, leads to minimizing more in power loss. It is clearly seen from the results that

two WTG diminished the power losses to 28.33 kW, while three WTGs reduced it to 11.68 kW, which is the best-minimized power losses value for all cases of different DGs. Further, the voltage stability is enhanced to be 31.53 p.u and the VD is minimized to 0.1223 p.u.

For this case of 33 bus system, the GMSA, HGWO, and EA produce better solutions compared to the other methods, whereas the best power loss value obtained by BSOA is much more than the rest of algorithms. The GMSA has a speedy and smooth rate of convergence to the minimum value without any oscillations and settles down early as shown in Fig. 4. On the other hand, the WTG has the best effect on the system performance as one WTG gave better results than 3 units of capacitor banks or 3 PV units. Fig. 5(a, b, c) shows a comparison between the different types of DGs in terms of voltage profile improvement.

B. 69-Bus Test System

To investigate the effectiveness of the proposed GMSA on a large scale of RDN, it is applied on the 69-bus system, which consists of 69 buses and 68 branches as shown in Fig. 6. This system is operated with 100 MVA base, 12.66 kV rated voltage, and the total system load is (1.896MW+j1.347MVAR). All data of lines and loads are given in [35]. The total real power loss for the base case without using capacitors or DGs is found at 224.99 kW with the lowest bus voltage at bus 65 is 0.9092 p.u. The best locations and sizes of the three types of PV, WTG, and capacitor banks that are obtained by the proposed GMSA and all results are listed in Table VI appendix (A) in terms of power losses, minimum and maximum bus voltages, VD, VSI, and the percentage of loss reduction.

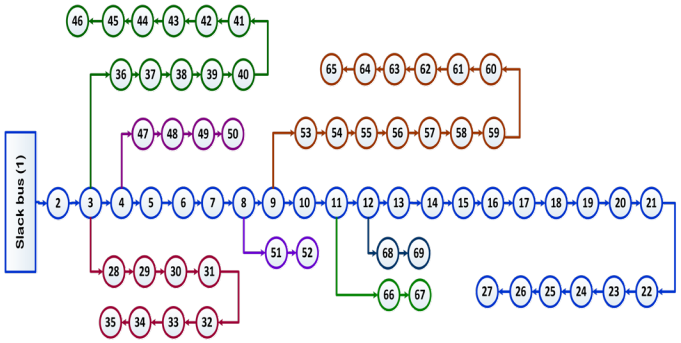


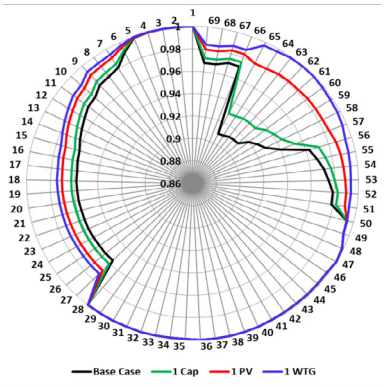
Fig. 6. Single line diagram of the 69-bus RDN [35].

1. Case 1: Q-Type DG

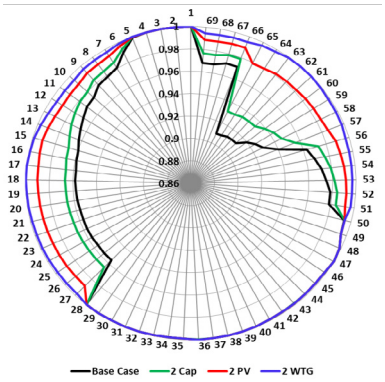
In this case, using the capacitors banks as Q-type helps in reducing the 69 bus system power loss by 32.61% from the base case using one DG unit at bus 61 with the size of 1200 kVAR, using the GMSA. It is considered the lowest value compared to the other methods. Moreover, the GMSA has increased the minimum bus voltage value to 0.9296 p.u after compensation. While, the power losses are reduced to 35.27% and 35.83% with two and three capacitors, respectively. Table VII appendix (A) summarized the obtained results by GMSA technique in case of installing 1 capacitor bank, 2 capacitor banks, and 3 capacitor banks units compared to other four algorithms. The results stated that as the number of capacitor banks units increased, the minimizing of power loss increased and consequently improved the whole system profile.

2. Case 2: P-Type DG

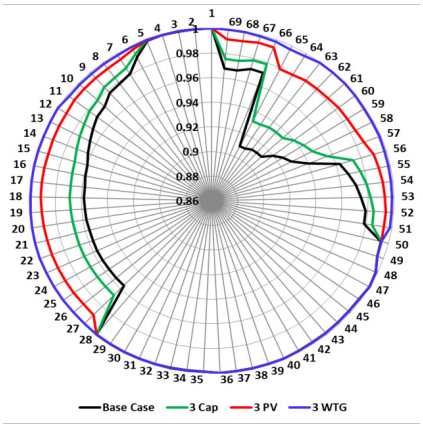
As for the previous system 33-bus, the installing of PV units enhances the voltage profile of the 69 bus system. It is seen from Table VIII appendix (A) that the GMSA selected bus 61 to install 1.87 MW PV unit, which reduces the power losses to 82.4 kW compared to 83.22 kW with MFO, HGWO and PSO, and 83.34 kW with EA. These four algorithms came in the second order after the GMSA method. While, a hybrid approach, MINLP, and IA came after that with 83.37, 83.38 and 83.44 kW, respectively. Moreover, GMSA enhanced the minimum bus voltage from 0.9036 p.u to 0.9686 p.u. For 2 PV units, the optimal bus locations are the same at 17 and 61 for all the methods, except the GMSA at 15 and 61. The power loss is 71.37 kW by GMSA. It is the lowest of all techniques as seen in Table VIII. In the last case of installing three PV units, GMSA selected buses 11, 17, and 61 to place the PV units with 526, 380.7, 1718 kW, respectively, which helps in reducing the real power loss to 68.974 kW and increasing the minimum bus voltage to 0.9799 p.u. It is also observed from the results that the VD is enhanced to be 0.4471 p.u, while the voltage stability is also enhanced to be 66.2363 p.u.



(a)

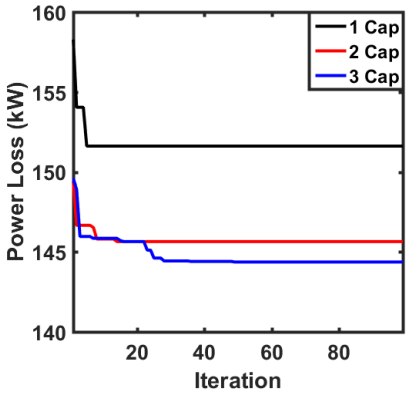


(b)

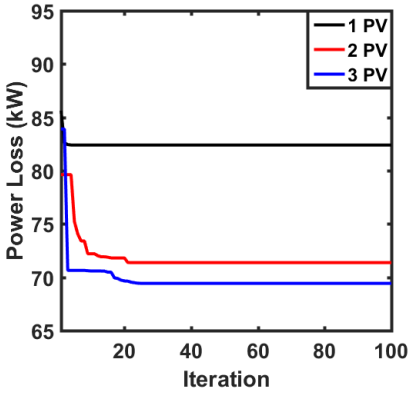


(c)

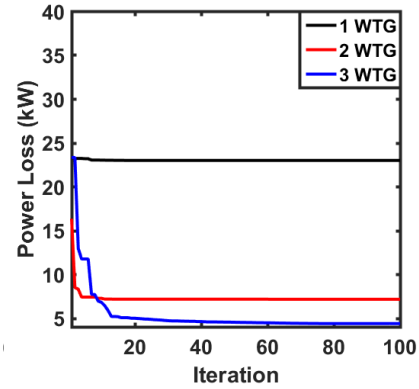
Fig. 7. 69-Bus voltage profile with (a) Single DG (b) Two DGs (c) Three DGs.



(a)



(b)



(c)

Fig. 8. Power loss convergence rate of 69-bus system using different DG types by GMSA.

3. Case 3: PQ--Type DG

To verify the effectiveness of the proposed GMSA method, it has been applied to assign the best locations of one, two, and three WTGs. In case of inclusion of one WTG DG, the kW loss is reduced to 22.98 kW and the VD is enhanced to be 0.5825 p.u while the voltage stability is also enhanced to be 65.7382 p.u. Furthermore, it increased the minimum bus voltage to 0.9728 p.u. In this case, it can be seen from Table IX appendix (A) that the results of the GMSA are the best compared to all other methods in terms of power loss and minimum bus voltage. Further, it can be noted that the use of only one WTG is better than the usage of 3 capacitor banks or 3 PV units. Moreover, increasing the number of WTGs, leads to more reduction in power loss. The results stated that two WTG reduced the power losses to 7.144 kW, while three WTGs reduced it to 4.21 kW, which is the best-minimized power losses value for all cases and types of DGs. Further, the voltage stability is also enhanced to be 67.7559 p.u with minimizing the VD to 0.0617 p.u. The proposed hybrid algorithm provides a significant improvement of bus voltage profile and power loss in the case of PQ--type as compared with other cases. Fig. 7(a, b, c) shows a comparison between the different types of DGs in terms of voltage profile improvement.

The results stated that the GMSA method performed better than the other algorithms over all cases of the 69-bus system. In addition, the best performance of the proposed GMSA is noted by the flat and stable convergence curves of total real power losses as shown in Fig. 8.

V. Conclusion

In this article, the exploitation ability of the GMSA, in terms of quick convergence and fast execution time, has been maintained by using the best moths in the swarm to perform that role in the phases of the transverse orientation and celestial navigation. The tradeoff between the global and local search has been regulated by introducing an adaptive mutation operation of GA on the pathfinders as the largest population group in the swarm. In addition, individuals have cooperated to produce the light sources for the guidance of the transverse orientation phase, which assists the exploration ability in such exploitation phase and enhances the solution diversity. The complexity of reconnaissance phase has been reduced. The GA operations increased the information sharing and the performance of the proposed algorithm.

The new GMSA approach has been successfully applied on multiple systems (33 and 69-bus systems) for solving the problem of renewable energy distributed generation sources placement and ratings for minimizing the total power losses. A sensitivity analysis based-Voltage Stability Index (VSI) has been performed to determine the best candidate locations for inclusion of the compensation devices in distribution systems to reduce the search space of GMSA and simulation time. Furthermore, a detailed comparison has been conducted with other best results of alternative methods in the literature. The P-type, Q-type and PQ--type DG units have been used in this study. The PQ--type presented the best results for all three cases of the two test systems. This is due to the variability in the solution vector and flexibility of power factor, which leads to more availability to select of other variables. The proposed GMSA can improve the voltage profile at each bus in these systems. Consequently, enhances the performance of distribution networks. GMSA method presented a desirable and superior performance with stable convergence. Results stated that the proposed GMSA minimized the objective function, and provided remarkable results compared to other algorithms. Hence, the applications of the proposed GMSA method can be considered as the most recent optimization algorithms for the network reconfiguration and dealing with the protection coordination system in presence of capacitors banks and renewable energy distributed generation sources during grid faults are the future scope of this work.

Appendix (A)

Nomenclature

P_k	Real power flow from bus k
Q_k	Reactive power flow from bus k
P_{Lk}	Real power load connected at bus k
Q_{Lk}	Reactive power load connected at bus k
$P_{L(k+1)}$	Real power load connected at bus k+1
$Q_{L(k+1)}$	Reactive power load connected at bus k+1
R_k	Resistance connected between buses k and k+1
X_k	Reactance connected between buses k and k+1
V_k	Voltage at bus k
V_{k+1}	Voltage at bus k+1
P_{sys}	Network active power
Q_{sys}	Network reactive power
ε1	Random samples drawn from Gaussian stochastic distribution
Qfc	Reactive power compensation
Vmin	Minimum bus voltage value
*V*max	Maximum bus voltage value
PT loss	Tap setting of transformer
*n*p	Number of pathfinders moths
μt	Variation coefficient
σ^t_j	Dispersal degree
*best*g	The global best solution
r1, r2	Random number within the interval [0, 1]
Pj	The real power loss during jth load level
n	The number of candidate buses
*Q*fc	The size of the shunt capacitor
ε2, ε3	Random numbers distributed uniformly within the interval [0,1]

TABLE I. Control-Parameters Values for the Different Algorithms

Parameter	Value (s)
Maximum iteration	100
Number of Search Agents	50
Number of Pathfinders	20
Murate	0.05
Weighting factor (W1)	0.5
Weighting factor (W2)	0.25
Weighting factor (W3)	0.25
Mix rate parameter (mixrate)	1.0
Voltage limits	$0.9 \le V_k \le 1.05$ PU
$P_{DG}^{\min} \le P_{DG}^{k} \le P_{DG}^{\max}$	$0 \le P_{DG}^{k} \le 5MW$ 33-Bus
	$0 \le P_{DG}^{k} \le 5MW$ 69-Bus
$pf_{DG}^{\min} \le pf_{DG}^{k} \le pf_{DG}^{\max}$	$0 \le pf_{DG}^{k} \le 0$ Q-type
	$1.0 \le pf_{DG}^{k} \le 1.0$ P-type
	$0.7 \le pf_{DG}^{k} \le 1.0$ PQ--type

TABLE II. Optimal Locations and Rating of Renewable DGs for Units Using GMSA for 33-Bus System

Type		V_{min} (p.u)	V_{max} (p.u)*	P_{loss} (kW)	% Loss reduction	VSI (p.u)	VD (p.u)	Optimal bus no, optimal DG (kVA), optimal (pf)
Without DG		0.9036	0.9971	210.98	—	25.5401	1.8044	—
1 DG	**1 Cap 1 PV 1 WT**	0.9175	0.9976	150.426	28.7%	26.7764	1.4259	30(1200)
		0.9427	0.9980	110.267	47.74%	28.6451	0.9235	6(2589)
		0.9586	1.000	67.418	68.05%	29.8458	0.5646	6(3105), (0.82)
2 DG	**2 Cap 2 PV 2 WT**	0.9332	0.9978	140.876	33.23%	27.2855	1.2702	30(1050), 12(450)
		0.9687	0.9984	86.58	58.96%	29.4035	0.6723	30(1157), 13(851)
		0.9805	1.0009	28.326	86.57	31.2752	0.1867	30(1556, 0.73), 13(935, 0.9)
3 DG	**3 Cap 3 PV 3 WT**	0.9334	0.9979	137.466	34.84%	27.3357	1.2568	24(450), 12(450), 30(1050)
		0.9712	0.9989	72.299	65.73%	29.6328	0.6124	24(1091), 13(801), 30(1053)
		0.9924	1.0006	11.684	94.46%	31.5347	0.1223	24(1187, 0.9), 13(877, 0.9), 30(1441, 0.72)

TABLE III. Comparison Results of Different Optimization Techniques for Q-type DG Units for 33-Bus System

Technique	1 DG unit				2 DG units				3 DG units			
	V_{min} (p.u)	Optimal (bus no, DG size kVA)	Total P_{loss} (kW)	CPU time (sec)	V_{min} (p.u)	Optimal (bus no, DG size kVA)	Total P_{loss} (kW)	CPU time (sec)	V_{min} (p.u)	Optimal (bus no, DG size kVA)	Total P_{loss} (kW)	CPU time (sec)
Base Case	0.9036	—	210.98	—	0.9036	—	210.98	—	0.9036	—	210.98	—
GMSA	**0.9175**	**30(1200)**	**150.426**	**6.2**	**09332**	**30(1050) 12(450)**	**140.87**	**6.22**	**0.9334**	**24(450), 12(450) 30(1050)**	**137.46**	**6.26**
GA	0.9173	29(1350)	153.121	**7.31**	0.9159	21(150) 30(1200)	151.12	**7.34**	0.9333	6(750), 13(350) 31(750)	142.07	**7.48**
MSA	0.9159	30(1200)	151.497	**6.51**	0.9297	8(750) 30(900)	143.11	**6.67**	0.9298	2(150), 12(1050) 30(450)	141.71	**6.93**
Hybrid [17]	0.9161	30(1230)	151.41	**NA**	0.9336	30(1040) 12(430)	141.94	**NA**	0.9335	24(450), 12(450) 30(1050)	138.37	**NA**
HGWO [20]	0.9163	30(1258)	151.36	**NA**	0.9338	30(1054) 12(467)	141.83	**NA**	0.9334	24(450), 12(450) 30(1050)	138.25	**NA**

TABLE IV. Comparison Results of Different Optimization Techniques for P-Type DG Units for 33-Bus System

Technique	1 DG unit				2 DG units				3 DG units			
	V_{min} (p.u)	Optimal (bus no, DG size kVA)	Total P_{loss} (kW)	CPU time (sec)	V_{min} (p.u)	Optimal (bus no, DG size kVA)	Total P_{loss} (kW)	CPU time (sec)	V_{min} (p.u)	Optimal (bus no, DG size kVA)	Total P_{loss} (kW)	CPU time (sec)
Base Case	0.9036	—	210.98	—	0.9036	—	210.98	—	0.9036	—	210.98	—
GMSA	**0.9427**	**6(2589)**	**110.267**	6.55	**0.9687**	**30(1157) 13(851)**	**86.58**	6.81	**0.9712**	**24(1091), 13(801) 30(1053)**	**72.299**	6.61
BSOA [21]	0.9441	8(1857.5)	118.12	20.4	0.9665	31(924) 13(880)	89.34	23.54	NA	NA	NA	24.95
PSO [16]	NA	6(2590)	111.03	NA	NA	30(1160) 13(850)	87.17	NA	NA	14(770), 24(1090) 30(1070)	72.79	NA
Hybrid [17]	0.9451	6(2490)	111.17	NA	0.9687	13(830) 30(1110)	87.28	NA	0.9713	13(790), 24(1070) 30(1010)	72.89	NA
HGWO [20]	0.9455	6(2590)	111.018	NA	0.9714	13(852) 30(1158)	87.164	NA	0.9715	13(802), 24(1090) 30(1054)	72.784	NA
IA [24]	0.9425	6(2601)	111.1	NA	0.9539	6(1800) 14(720)	91.63	NA	0.9690	6(900), 12(900) 31(720)	81.05	NA
EA [23]	NA	6(2530)	111.07	NA	NA	13(844) 30(1149)	87.172	NA	NA	13(798), 24(1099) 30(1050)	72.787	NA
MINLP [12]	0.9424	6(2590)	111.01	NA	0.9685	13(850) 30(1150)	87.16	NA	0.9687	13(800), 24(1090), 30(1050)	72.79	NA

TABLE V. COMPARISON RESULTS OF DIFFERENT OPTIMIZATION TECHNIQUES FOR PQ--TYPE DG UNITS FOR 33-BUS SYSTEM

Technique	1 DG unit				2 DG units				3 DG units			
	V_{min} (p.u)	Optimal (bus no, DG size kVA, pf)	Total P_{loss} (kW)	CPU time (sec)	V_{min} (p.u)	Optimal (bus no, DG size kVA)	Total P_{loss} (kW)	CPU time (sec)	V_{min} (p.u)	Optimal (bus no, DG size kVA, pf)	Total P_{loss} (kW)	CPU time (sec)
Base Case	0.9036	—	210.98	—	0.9036	—	210.98	—	0.9036	—	210.98	—
GMSA	**0.9586**	**6(3105, 0.82)**	**67.418**	**6.79**	**0.9805**	**30(1556,0.73) 13(935, 0.9)**	**28.326**	**6.77**	**0.9924**	**24(1187, 0.9), 13(877, 0.9), 30(1441, 0.72)**	**11.684**	**7.03**
BSOA [21]	0.9549	8(2265.5, 0.82)	82.78	36.87	0.9796	13(777,0.89) 29(1032,0.7)	31.98	49.80	NA	NA	NA	56.50
PSO [16]	?	6(3035, 0.82)	67.928	NA		13(914,0.91) 30(1535,0.73)	28.56	NA		24(1188, 0.9), 13(863, 0.91), 30(1431, 0.71)	11.76	NA
Hybrid [17]	0.9587	6(3028, 0.82)	67.937	NA	0.9801	13(1039,0.91) 30(1508,0.72)	28.98	NA	0.9923	24(1186, 0.9), 13(873, 0.9), 30(1441, 0.72)	11.76	NA
HGWO [20]	0.9585	6(3106, 0.82)	67.855	NA	0.9802	13(932,0.9), 30(1558,0.72)	28.5	NA	0.9922	24(1182, 0.9), 13(878, 0.89), 30(1439, 0.71)	11.74	NA
IA [24]	0.9425	6(3107, 0.82)	67.85	NA	0.9539	6(2195,0.82) 30(1098,0.82)	44.39	NA	0.9690	6(1098,0.82),14(768, 0.82), 30(1098,0.82)	22.29	NA
MINLP [12]	0.9424	6(3105, 0.82)	67.85	NA	0.9685	(13, 0.88), (30, 0.8), (2477)	29.31	NA	0.9687	13(0.87), 24(0.88), 30(0.8), (3481)	12.74	NA
EA [23]	NA	6(3119, 0.82)	67.87	NA	NA	13(938,0.9) 30(1537,0.73)	28.52	NA	NA	24(1189,0.9),13(886, 0.9), 30(1450,0.71)	11.8	NA

TABLE. VI. OPTIMAL LOCATIONS AND RATING OF RENEWABLE DGS FOR UNITS USING GMSA FOR 69-BUS SYSTEM

Type		V_{min} (p.u)	V_{max} (p.u)*	P_{loss} (kW)	% Loss reduction	VSI (p.u)	VD (p.u)	Optimal bus no, optimal DG (kVA), optimal (pf)
Without DG		0.9092	0.9999	224.99	—	61.2183	1.8374	—
1 DG	**1 Cap**	0.9296	0.9999	151.617	32.61%	62.2409	1.5361	61(1200)
	1 PV	0.9686	0.9999	82.4	63.38%	64.6524	0.8645	61(1872)
	1 WT	0.9728	0.9999	22.98	89.79%	65.7382	0.5825	61(2243.7, 0.81)
2 DG	**2 Cap**	0.9315	0.9999	145.646	35.27%	62.6248	1.4293	61(1200), 12(600)
	2 PV	0.9792	0.9999	71.371	68.28%	66.0147	0.5041	61(1777), 15(554)
	2 WT	0.9944	1.0003	7.144	96.82%	67.4868	0.1289	61(2131, 0.81), 17(630.8, 0.83)
3 DG	**3 Cap**	0.9318	0.9999	144.369	35.83%	62.7862	1.3844	61(1200), 53(350), 17(350)
	3 PV	0.9799	1.0002	68.974	69.34%	66.2363	0.4471	61(1718), 17(380.7), 11(526)
	3 WT	0.9943	1.003	4.21	98.13%	67.7559	0.0617	61(2058,0.81), 17(459,0.85), 11(603.7,0.8)

TABLE. VII. COMPARISON RESULTS OF DIFFERENT OPTIMIZATION TECHNIQUES FOR Q-TYPE DG UNITS FOR 69-BUS SYSTEM

Technique	1 DG unit				2 DG units				3 DG units			
	V_{min} (p.u)	Optimal (bus no, DG size kVA)	Total P_{loss} (kW)	CPU time (sec)	V_{min} (p.u)	Optimal (bus no, DG size kVA)	Total P_{loss} (kW)	CPU time (sec)	V_{min} (p.u)	Optimal (bus no, DG size kVA)	Total P_{loss} (kW)	CPU time (sec)
Base Case	0.9092	—	224.99	—	0.9092	—	224.99		0.9092	—	224.99	
GMSA	**0.9312**	**61(1200)**	**151.617**	14.3	**0.9315**	**61(1200) 12(600)**	**145.65**	13.76	**0.9318**	**61(1200) 53(350), 17(350)**	**144.37**	14.66
GA [29]	0.9311	61(1350)	152.07	NA	0.9310	61(1200) 66(600)	147.63	NA	0.9308	12(600), 45(150) 61(1200)	146.72	NA
MSA [29]	0.9310	61(1350)	152.05	NA	0.9288	61(1200) 12(600)	146.69	NA	0.9299	2(1050), 17(350) 61(1200)	146.61	11.42
Hybrid [17]	NA	61(1290)	152.1	NA	NA	61(1240) 18(350)	146.49	NA	NA	11(330), 18(250) 61(1190)	145.28	NA
HGWO [20]	0.9311	61(1330)	152.04	NA	0.9315	61(1277) 17(364)	146.44	NA	0.9317	11(412), 21(230) 61(1231)	145.12	NA

TABLE. VIII. Comparison Results of Different Optimization Techniques for P-type DG Units for 69-Bus System

Technique	1 DG unit				2 DG units				3 DG units			
	V_{min} (p.u)	Optimal (bus no, DG size kVA)	Total P_{loss} (kW)	CPU time (sec)	V_{min} (p.u)	Optimal (bus no, DG size kVA)	Total P_{loss} (kW)	CPU time (sec)	V_{min} (p.u)	Optimal (bus no, DG size kVA)	Total P_{loss} (kW)	CPU time (sec)
Base Case	0.9092	—	224.99	—	0.9092	—	224.99	—	0.9092	—	224.99	—
GMSA	**0.9686**	**61(1872)**	**82.4**	**14.66**	**0.9792**	**61(1777) 15(554)**	**71.371**	**14.3**	**0.9799**	**61(1718) 17(380.7), 11(526)**	**68.974**	**14.69**
IA [24]	0.9692	61(1900)	83.44	NA	0.9765	61(1700) 17(510)	72.13	NA	0.9785	11(340), 17(510) 61(1700)	70.16	NA
PSO [16]	0.9681	61(1870)	83.222	**NA**	0.9806	17(1780) 61(1530)	71.68	**NA**	0.9806	11(460), 17(440) 61(1700)	69.52	**NA**
Hybrid [17]	NA	61(1810)	83.372	**NA**	NA	17(520) 61(1720)	71.82	**NA**	NA	12(496), 22(311) 61(1735)	69.56	**NA**
HGWO [20]	0.9682	61(1872)	83.222	**NA**	0.9799	17(531) 61(1781)	71.674	**NA**	0.9799	11(527), 17(380) 61(1781)	69.425	**NA**
MINLP [12]	0.9682	61(1870)	83.38	**NA**	0.9789	17(530) 61(1780)	71.83	**NA**	0.9790	11(530), 17(380) 61(1720)	69.59	**NA**
EA [23]	NA	61(1878)	83.23	NA	NA	17(534) 61(1795)	71.68	NA	NA	11(467), 18(380) 61(1795)	69.62	NA

TABLE. IX. Comparison Results of Different Optimization Techniques for PQ--type DG Units for 69-Bus System

Technique	1 DG unit				2 DG units				3 DG units			
	V_{min} (p.u)	Optimal (bus no, DG size kVA, pf)	Total P_{loss} (kW)	CPU time (sec)	V_{min} (p.u)	Optimal (bus no, DG size kVA)	Total P_{loss} (kW)	CPU time (sec)	V_{min} (p.u)	Optimal (bus no, DG size kVA, pf)	Total P_{loss} (kW)	CPU time (sec)
Base Case	0.9092	—	224.99		0.9092	—	224.99	—	0.9092	—	224.99	—
GMSA	**0.9728**	**61(2243.7, 0.81)**	**22.98**	**14.67**	**0.9944**	**61(2131,0.81) 17(630.8,.83)**	**7.144**	**14.94**	**0.9943**	**61(2058,0.81), 17(459,0.85), 11(603.7,0.8)**	**4.21**	**15.18**
MINLP [12]	0.9682	(61, 2244, 0.81)	23.16	NA	0.9789	(17,658,0.82) (61,2196,0.82)	7.44	NA	0.9790	(11,607,0.813), (50,1058,0.82) (61,1058,0.82)	4.26	NA
PSO [16]	0.9724	(61, 2240, 0.81)	23.16	NA	0.9943	(17,630,0.82) (61,2130,0.81)	7.2	NA	0.9940	(11,600,0.83), (18,460, 0.81) (61,2060,0.81)	4.28	NA
Hybrid [17]	NA	(61, 2200, 0.82)	23.92	NA	NA	(17,630,0.82) (61,2120,0.81)	7.21	NA	NA	(18,480,0.77), (61,2060,0.83) (66,530,0.82)	4.30	NA
HGWO [20]	0.9724	(61, 2246, 0.81)	23.16	NA	0.9941	(17,628,0.82) (61,2127,0.81)	7.2	NA	0.9942	(11,614,0.81), (18,452, 0.83) (61,2056,0.81)	4.26	NA
IA [24]	0.9732	61(2243, 0.82)	22.62	NA	0.9944	61(2195,0.82) 17(659, 0.82)	7.25	NA	0.9939	61(2073,0.82), 17(622,0.82) 50(829,0.82)	4.95	NA
EA [25]	0.9731	61(2290, 0.82)	23.26	NA	0.9942	17(643,0.83) 61(2189,0.82)	7.35	NA	0.9943	11(668,0.82), 18(458, 0.83), 61(2113,0.82)	4.48	NA

References

[1] P. Dinakara, V. C. Veera, and T. Gowri Manohar, "Whale optimization algorithm for optimal sizing of renewable resources for loss reduction in distribution systems", Renewable: Wind, Water, and Solar, Vol. 4, No.1, 2017, pp. 1.

[2] L. Freris, D. Infield, Renewable energy in power system. Wiltshire: A John Wiley & Sons; 2008.

[3] M. Saejia and I. Ngamroo "Stabilization of microgrid with intermittent renewable energy sources by SMES with optimal coil size", *Elsevier, Physica C*, 2011, pp. 1385–1389.

[4] HA, M. P., Huy, P. D., & Ramachandaramurthy, V. K. "A review of the optimal allocation of distributed generation: Objectives, constraints, methods, and algorithms". Renewable and Sustainable Energy Reviews, 75, 2017, pp. 293-312.

[5] T. Griffin, K. Tomsovic, D. Secrest, A. Law, "Placement of dispersed generation systems for reduced losses. In: 33th proceeding Hawaii int. Conf. on System sciences", 2000. pp. 1-9.

[6] W. Ahmed, A. Selim, S. Kamel, J. Yu, F. Jurado, "Probabilistic Load Flow Solution Considering Optimal Allocation of SVC in Radial Distribution System", *International Journal of Interactive Multimedia and Artificial Intelligence*. 2018;5(3):152-161.

[7] N. Rezaei, MR.Haghifam, "Protection Scheme for A Distribution system with distributed generation using neural networks", *Int. J Electr Power Energy Syst*, May 2008;30(4):235-241.

[8] TJ. Hashim, A. Mohamed, "Fuzzy logic based coordinated voltage control for distribution network with distributed generations", *Int. J. of Electrical, Computer, Energetic*, Electron Commun Eng 2013;7(7):369-374.

[9] A. Awad, T. El-Fouly, M. Salama, "Optimal distributed generation allocation and load shedding for improving distribution system reliability", *Electr Power Components Syst*, 2014;42(6):576-584.

[10] F. S. Abu-Mouti and M. El-Hawary, "Optimal distributed generation allocation and sizing in distribution systems via artificial bee colony algorithm," *IEEE transactions on power delivery*, vol. 26, 2011, pp. 2090-2101.

[11] T. Shukla, S. Singh, V. Srinivasarao, and K. Naik, "Optimal sizing of distributed generation placed on radial distribution systems," *Electric power components and systems,* vol. 38, pp. 260-274, 2010.

[12] W. Tan, M. Hassan, M. Majid, and H. A. Rahman, "Allocation and sizing of DG using Cuckoo Search algorithm," in Power and Energy (PECon), 2012 IEEE International Conference on, 2012, pp. 133-138.

[13] S. Kaur, G. Kumbhar, and J. Sharma, "A MINLP technique for optimal placement of multiple DG units in distribution systems," *International Journal of Electrical Power & Energy Systems*, vol. 63, Dec. 2014, pp. 609-617.

[14] L. Arya, A. Koshti, and S. Choube, "Distributed generation planning using differential evolution accounting voltage stability consideration," *International Journal of electrical power & energy systems*, vol. 42, pp. 196-207, 2012.

[15] E. S. Oda, A. A. Abdelsalam, M. N. Abdel-Wahab, and M. M. El-Saadawi, "Distributed generations planning using flower pollination algorithm for enhancing distribution system voltage stability," *Ain Shams Engineering Journal*, 2015.

[16] S. K. Injeti at al., "A novel approach to identify optimal access point and capacity of multiple DGs in a small, medium and large-scale radial distribution systems," *Int. Journal of Electrical Power & Energy Systems*, vol. 45, no. 1, Feb. 2013, pp. 142-151.

[17] B.B. Zad, H. Hasanvand, J. Lobry, and F. Vallée, "Optimal reactive power control of DGs for voltage regulation of MV distribution systems using sensitivity analysis method and PSO algorithm," *International Journal of Electrical Power & Energy Systems*, vol. 68, Jun. 2015, pp. 52-60.

[18] S. Kansal, V. Kumar, and B. Tyagi, "Hybrid approach for optimal placement of multiple DGs of multiple types in distribution networks," *International Journal of Electrical Power & Energy Systems*, vol. 75, Feb. 2016, pp. 226-235.

[19] M.H. Moradi, A. Zeinalzadeh, Y. Mohammadi, and M. Abedini, "An efficient hybrid method for solving the optimal siting and sizing problem of DG and shunt capacitor banks simultaneously based on an imperialist competitive algorithm and genetic algorithm," *International Journal of Electrical Power & Energy Systems*, vol. 54, Jan. 2014, pp. 101-111.

[20] M. Kefayat, A.L. Ara, and S.N. Niaki, "A hybrid of ant colony optimization and artificial bee colony algorithm for probabilistic optimal placement and sizing of distributed energy resources," *Energy Conversion and Management*, vol. 92, Mar. 2015, pp.149-161.

[21] R. Sanjay, Jayabarathi, T. Raghunathan, V. Ramesh, and Nadarajah, "Optimal Allocation of Distributed Generation Using Hybrid Grey Wolf Optimizer", *IEEE Access*, 2017, pp. 14807 – 14818.

[22] El-Fergany A. Optimal allocation of multi-type distributed generators using backtracking search optimization algorithm. *Int. J Electr Power Energy Syst*. Vol. 64, January 2015, pp. 1197-1205.

[23] K. Muthukumar, and S. Jayalalitha, "Optimal placement and sizing of distributed generators and shunt capacitors for power loss minimization in radial distribution networks using hybrid heuristic search optimization technique," *International Journal of Electrical Power & Energy Systems*, vol. 78, Jun. 2016, pp. 299-319.

[24] K. Mahmoud, N. Yorino, and A. Ahmed, "Optimal distributed generation allocation in distribution systems for loss minimization," *IEEE Transactions on Power Systems*, vol. 31, no. 2, Mar. 2016, pp. 960-969.

[25] D.Q. Hung, and N. Mithulananthan, "Multiple distributed generator placement in primary distribution networks for loss reduction," *IEEE Transactions on industrial electronics*, vol. 60, no. 4, Apr. 201, pp. 1700-1708.

[26] K. Mahmoud and M. Abdel-Nasser, "Fast-yet-Accurate Energy Loss Assessment Approach for Analyzing/Sizing PV in Distribution Systems Using Machine Learning," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 13, July. 2019, pp. 1025-1033.

[27] Naresh Acharya, Pukar Mahat, N. Mithulananthan, "An analytical approach for DG allocation in primary distribution network", *Electrical Power and Energy Systems*, vol. 28, 2006, pp. 669-678.

[28] E. Ali, S. A. Elazim, and A. Abdelaziz, "Ant Lion Optimization Algorithm for Renewable Distributed Generations," Energy, vol. 116, pp. 445-458, 2016.

[29] A. El-Fergany, "Optimal allocation of multi-type distributed generators using backtracking search optimization algorithm," *International Journal of Electrical Power & Energy Systems*, vol. 64, pp. 1197-1205, 2015.

[30] Mohamed, E. A., Mohamed, A. A. A., & Mitani, Y, "Hybrid GMSA for Optimal Placement and Sizing of Distributed Generation and Shunt Capacitors," *Journal of Engineering Science & Technology Review*, 11(1).

[31] EA. Mohamed, A. Mohamed -, and Y. Mitani, "MSA for Optimal Reconfiguration and Capacitor Allocation in Radial/Ring Distribution Networks," *International Journal of Interactive Multimedia and Artificial Intelligence*. 2018;5(1):107-122.

[32] J. A. Micheline Ruba and S. Ganesh "Power Flow Analysis for Radial Distribution Systems using Backward/Forward Sweep Method", *International Journal of Electrical and Computer Engineering*, Vol 8, No.10, 2014, pp. 1621 – 1625.

[33] W.Sheng, K. Y. Liu, Y. Liu, X. Meng, and Y. Li, "Optimal placement and sizing of distributed generation via an improved nondominated sorting genetic algorithm II," *IEEE Transactions on Power Delivery*, 2015, 30(2), 569-578.

[34] B. Venkatesh, and R. Ranjan, "Optimal radial distribution system reconfiguration using fuzzy adaptation of evolutionary programming," *International journal of electrical power & energy systems*, vol. 25, no. 10, pp. 775-780, Dec. 2013.

[35] J.S. Savier, and D. Das, "Impact of network reconfiguration on loss allocation of radial distribution systems," *IEEE Transactions on Power Delivery*, vol. 22, no. 4, pp. 2473-2480, Oct 2007.

Emad A. Mohamed

He received the B.Sc., M.Sc. and PhD degrees in electrical power engineering from Aswan University, Aswan, Egypt, and Kyushu Institute of Technology in 2005, 2013, and 2019 respectively. He was working as a Demonstrator in the Department of Electrical Engineering, Aswan Faculty of Engineering, Aswan University, Aswan, Egypt. Between November 2007 and August 2013, and he is an assistant lecturer from 2013 to 2019. Now, he is an Assistant Professor. He was in a Master Mobility scholarship at Faculté des Sciences et Technologies - Université de Lorraine, France – 1. The scholarship sponsored by FFEEBB ERASMUS MUNDUS. He was a research student from April to October 2015 in Kyushu University, Japan. His research interests are applications of superconductivity on power systems, power system stability, and protection.

Al-Attar Ali Mohamed

Al-Attar Ali Mohamed received his B.Eng., M.Eng, and Ph.D. from Aswan University, Aswan, Egypt in 2010, 2013 and 2017. Currently, he is an assistant professor sponsored in Faculty of Engineering, Aswan University. His research interests are in the area of analysis and design of power systems, optimization, and controls.

Yasunori Mitani

Yasunori Mitani received B.Sc., M.Sc., and D.Eng. degrees in Electrical Engineering from Osaka University, Japan in 1981, 1983 and 1986, respectively. He was a visiting research associate at the University of California, Berkeley, from 1994 to 1995. He is currently a professor at the department of electrical and electronics engineering, Kyushu Institute of Technology (KIT), Japan. At present, he is the head of environmental management center of KIT and vice dean of a graduate school of engineering, KIT. He is also the President of the Institute of Electrical Engineers of Japan (IEEJ), Power and Energy Society.

Performance Enhancement of Wind Farms Using Tuned SSSC Based on Artificial Neural Network

Yousry Ibrahim[1], Salah Kamel[1,2], Ahmed Rashad[1], Loai Nasrat[1], Francisco Jurado[3]*

[1] Department of Electrical Engineering, Faculty of Engineering, Aswan University, 81542 Aswan (Egypt)
[2] Department of Electrical Engineering, University of Jaén, 23700 EPS Linares, Jaén (Spain)
[3] State Key Laboratory of Power Transmission Equipment and System Security and New Technology, Chongqing University, Chongqing 400030 (China)

Received 4 December 2018 | Accepted 1 May 2019 | Published 9 May 2019

Abstract

Recently, power systems are confronting a lot of challenges. Increasing the dependence on renewable energy sources especially wind energy and its impact on the stability of electrical systems are the most important challenges. Flexible alternating current transmission systems (FACTS) can be used to improve the relationship between wind farms and electrical grids. The performance of these FACTS depends on the parameters of its control system. These parameters can be tuned using modern methods like Artificial Neural Network (ANN). In this paper, ANN is used to improve the performance of static synchronous series compensator (SSSC) integrated into combined wind farm (CWF). This CWF is composed of squirrel cage induction generators (SCIG) and doubly fed induction generators (DFIG) wind turbines. This wind farm is collecting the advantage of SCIG and DFIG wind turbines. To view out the motivation of this paper, a comparison is done among the performances of combined wind farm (CWF) with ANN-SSSC, CWF with ordinary SSSC and CWF with SSSC tune by Multi-objective genetic algorithm (MOGA SSSC). The root mean square Error (RMSE) is used to evaluate the results. The results illustrate that the performance of CWF can be improved using SSSC adjusted by ANN.

Keywords

Squirrel Cage Induction Generator (SCIG), Doubly Fed Induction Generator (DFIG), Combined Wind Farm (CWF), Static Synchronous Series Compensator (SSSC), Artificial Neural Network (ANN).

DOI: 10.9781/ijimai.2019.05.001

I. Introduction

Renewable energy is an important source for the power generation. Solar energy, and wind energy are the most famous forms of this technology. Wind energy plays an important role in producing electric power in all the world so that its injection on the grid represents a wide range of studies. This injection depends on the induction generator of the wind turbines. There are two types of induction generator, first type is squirrel cage induction generators (SCIG) which are suitable to fixed speed wind turbines and second type is doubly fed induction generators (DFIG) that are used with variable speed wind turbines. The stability of wind farms is affected by the exchange in the reactive power between the interconnected grid and the wind farms. The compensation devices of the reactive power consider a fundamental element in SCIG wind turbines (SCIG-WT). The flexible alternating current transmission systems (FACTS) were used to damp power oscillation and, enhance power stability. In Ref. [1] a dual STATCOM had been used to damp power oscillations. Tuning parameters of SSSC had been proposed in [2] to damp power oscillations. In Ref [3] a unified power flow controller has been used to damp power oscillations between two areas. The SSSC used to damp power oscillation, enhance power stability and control the power flow of DFIG-WF is studied in [4]-[5].The effect of FACTS such as Static VAR Compensator (SVC), Static Synchronous Compensator (STATCM) and SSSC on the performance of wind farms were studied in [6]-[10]. The impact of SSSC on the performance of different types of wind farms had been discussed in [11].

The main advantage of Artificial intelligence (AI) is solving complex problems in less time and with high precision, such as using optimization methods to solve the complex control problem. Also, AI can easily predict and take the correct decisions with little margin of error. It can be used for predicting the change in wind speed and its impact on stability of power system. In this paper, AI has been used to predict and determine the optimal value of the control gains of SSSC which can enhance the performance of CWF. On other side, AI represents high technology so that it is storage costly. In last years, Artificial Intelligence (AI) has been used extensively in improving the performance of FACTS and enhancing the performance of wind farms interconnected grid. A genetic algorithm has been implemented to tune different type of FACTS interconnected wind farms and photovoltaic solar plant in [12]. In ref [13] [14] multi-objective genetic algorithm is used to improve the performance of DFIG. Also, multi-objective genetic algorithm is used to find the optimal gains of SSSC in [15]. Adaptive-network-based fuzzy inference system (ANFIS), ANN and genetic algorithm are proposed in [16] to improve the reactive power control of STATCOM. The whale optimization algorithm, genetic algorithm and ANN were used in [17] to determine the optimal parameters of STATCOM integrated with CWF. In Ref [18] particle swarm optimization is used to tune and damp power system oscillation of DFIG wind farms integrated with SSSC. A new control strategy based on ANFIS is proposed in [19] to improve the performance of DFIG wind farm integrated with SSSC.

* Corresponding author.
E-mail address: fjurado@ujaen.es

This paper aims to improve the performance of CWF which is based on SCIG and DFIG using SSSC controlled by ANN (ANN-SSSC). Also, in this paper the control parameters which had been investigated in [15] are used for implementing ANN. Moreover, a comparison is done between the performances of CWF with ordinary SSSC, CWF associated with SSSC tuned by multi-objective genetic algorithm (SSSC MOGA) investigated in [15] and CWF associated with proposed ANN-SSSC during three phase-faults.

The rest of the paper is organized as follows. Section II presents a brief summary of ANN. Section III presents modelling of wind turbines. Section IV explains the construction, operation and control system of SSSC. Section V introduces the proposed ANN control, which is applied to SSSC. The last two sections present the results and conclusion.

II. Artificial Neural Network (ANN)

The artificial neural network is a modest simulation of the effect, form and content of the neural network found in the human brain. It consists of nodes called neurons and connected together by bonds called weights. Each set of neurons forms a single layer; the ANN is composed of different types of layers. From Fig. 1, it can be observed that it consists of input layer, hidden layer (processing element) and output layer. The hidden layer could be single layer or multi-layers. The input signal is passed from input layer to the output layer through the hidden layer. The input is transferred to the neurons through weight matrix W. The output can be given by [20]:

$$Y_{out} = \sum_{j=1\to n}^{i=1\to n} x_i w_{ji} \quad (1)$$

Where Y_{out} represents the output of ANN, x is input signal which starts from 1 to n inputs and w_{ji} represents the synaptic weights between neurons.

III. Modeling of Wind Turbines

The mathematical model of wind turbines was discussed in several articles on wind energy. Fig. 2 shows the equivalent circuit of induction generator. The direct and quadratic (d-q) illustration of IG with respect to the synchronous frame can be illustrated as flows [21] [22]:

$$\begin{cases} v_{ds} = p\varphi_{ds} - \omega_s\varphi_{qs} + R_s i_{ds} \\ v_{qs} = p\varphi_{qs} - \omega_s\varphi_{ds} + R_s i_{qs} \end{cases} \quad (2)$$

$$\begin{cases} v_{dr} = p\varphi_{dr} - s\omega_s\varphi_{qr} + R_r i_{dr} \\ v_{qr} = p\varphi_{qr} - s\omega_s\varphi_{dr} + R_r i_{qr} \end{cases} \quad (3)$$

$$\begin{bmatrix} \varphi_{ds} \\ \varphi_{qs} \\ \varphi_{dr} \\ \varphi_{qr} \end{bmatrix} = \begin{bmatrix} i_{ds} \\ i_{qs} \\ i_{dr} \\ i_{qr} \end{bmatrix} \cdot \begin{bmatrix} (L_{ls}+L_m) & 0 & L_m & 0 \\ 0 & (L_{ls}+L_m) & 0 & L_m \\ L_m & 0 & (L_{lr}+L_m) & 0 \\ 0 & L_m & 0 & (L_{lr}+L_m) \end{bmatrix} \quad (4)$$

$$\begin{bmatrix} i_{ds} \\ i_{qs} \\ i_{dr} \\ i_{qr} \end{bmatrix} = \frac{1}{(L_{ls}+L_m)(L_{lr}+L_m)-L_m^2} \begin{bmatrix} (L_{lr}+L_m) & 0 & -L_m & 0 \\ 0 & (L_{lr}+L_m) & 0 & -L_m \\ -L_m & 0 & (L_{ls}+L_m) & 0 \\ 0 & -L_m & 0 & (L_{ls}+L_m) \end{bmatrix} \cdot \begin{bmatrix} \varphi_{ds} \\ \varphi_{qs} \\ \varphi_{dr} \\ \varphi_{qr} \end{bmatrix} \quad (5)$$

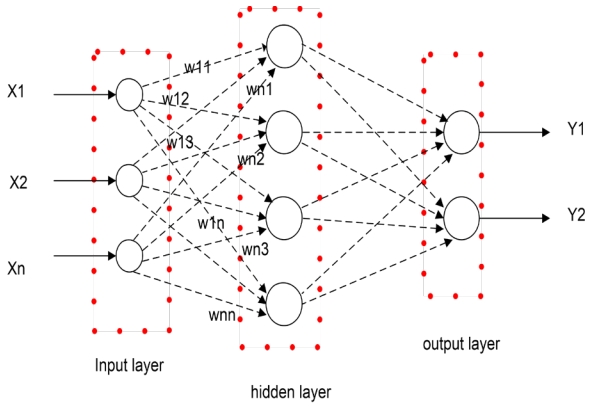


Fig. 1. ANN principle operation.

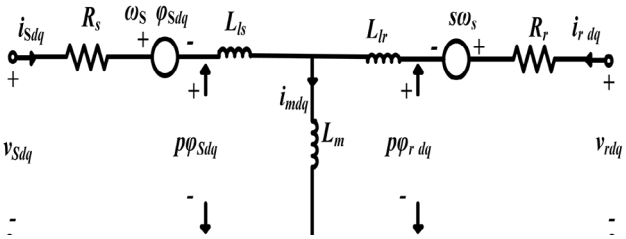
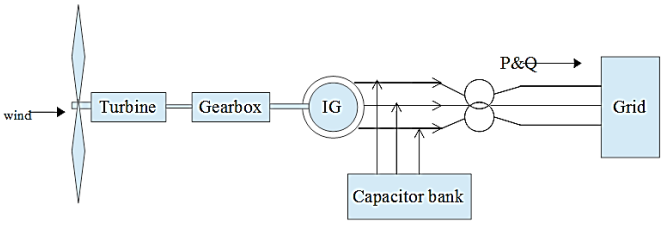
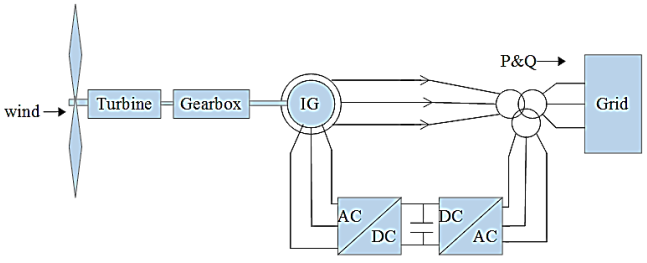


Fig. 2. Equivalent circuit of IG generators.

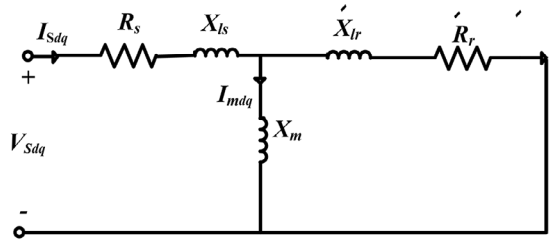
Equations (2) to (5) and Fig. 2 can be applied to a general modeling of IG. The voltage of rotor of SCIG is equal to zero because the rotor is a short circuit and there is no connection between SCIG's rotor and the grid as shown in Fig. 3 (a). While the rotor in DFIG is a wound rotor so the rotor current and voltage is taken into account. The AC/DC/AC converters are used to connect rotor of DFIG's to the grid through as illustrated in Fig. 3 (b). The equivalent circuits of DFIG and SCIG are illustrated in Fig. 3 (c) and (d).



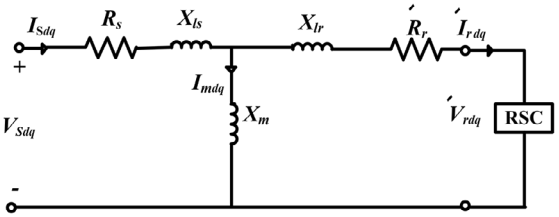
(a) A block diagram of SCIG interconnected grid



(b) A block diagram of DFIG interconnected grid



(c) Equivalent circuit of SCIG during steady stat



(d) Equivalent circuit of DFIG during steady stat

Fig. 3. Equivalent circuits and block diagrams of SCIG and DFIG.

The extracted power from the wind by wind farms is given by:

$$P_{wt} = \frac{1}{2}\rho A v^3 C_p$$

$$C_p = 0.517\left(\frac{116}{\lambda_i} - 0.4\beta - 5\right)e^{\frac{-0.0068}{\lambda_i}} + 0.0068\lambda$$

$$\frac{1}{\lambda_i} = \frac{1}{116}\left(\frac{1}{\lambda+0.08\beta} - \frac{0.035}{\beta^3+1}\right)$$

$$\lambda = \frac{\omega_r l_b}{v} \tag{6}$$

Where, v is the wind speed, ρ is the air density, C_p is the power coefficient, A is the area swept by the turbine blades. l_b is the blade length or rotor radius, ω_r is the rotor speed and it is equal to1.22Kg/m3. C_p is a function on the pitch angle β and the tip speed ratio λ.

The electrical torque of SCIG and DFIG is given by:

$$T_e = \frac{3PL_m}{2L_r}\left(i_{qs}\varphi_{dr} - i_{ds}\varphi_{qr}\right) \tag{7}$$

$$T_e = \frac{3P}{2}\left(i_{qs}\varphi_{ds} - i_{ds}\varphi_{qs}\right) \tag{8}$$

The pitch angle control method is used to control the rotor speed of wind turbine in order to keep the output power inside permissible limits. Fig. 4 illustrates a schematic diagram of pitch angle control system [23].

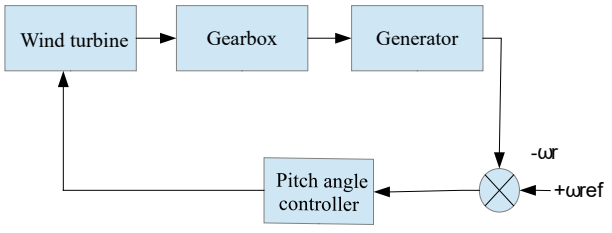
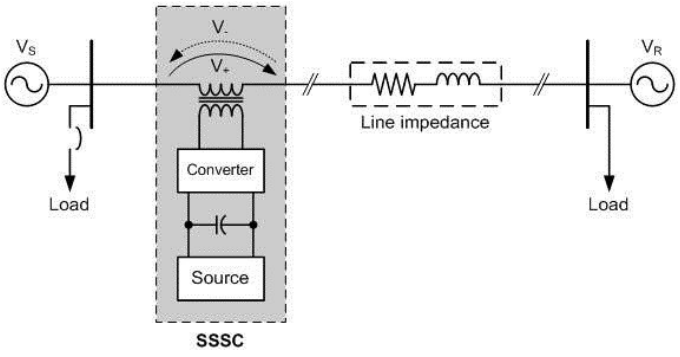


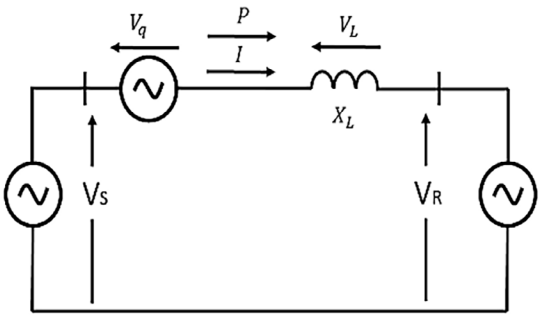
Fig. 4. A block diagram of pitch angle control system.

IV. Static Synchronous Series Compensator (SSSC)

The Static Synchronous Series Compensator (SSSC) belongs to the series devices of FACTS controller [23]. It is a series connected with the transmission line so that it injects a series voltage which is in phase quadrature with the line current. The block diagram of SSSC and its equivalent circuit are shown in Fig. 5. As shown in Fig. 5. The SSSC is connected in series with the transmission line of an electrical grid by a coupling transformer.



(a) Schematic diagram of SSSC



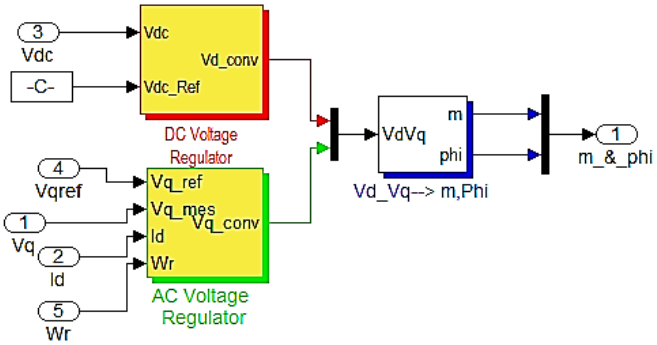
(b) Equivalent circuit of SSSC

Fig. 5. Principle operation of SSSC.

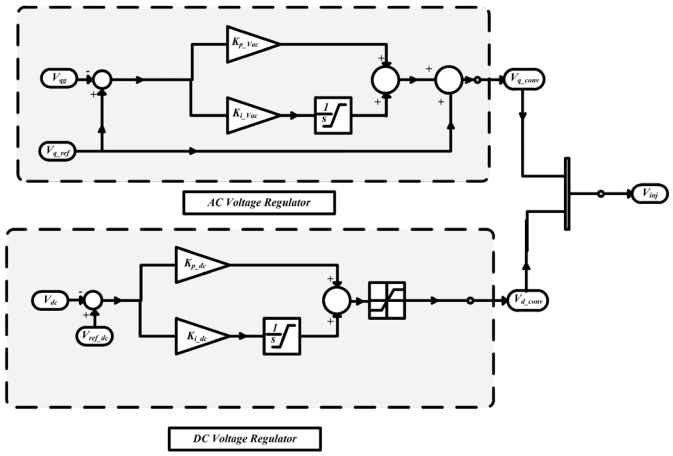
As shown in Fig. 5 (b) the SSSC injects voltage in series with the line. This voltage injected may be capacitive or inductive, if the injection voltage overrides the voltage drop (VL), the transferred power will be reflected in the direction.

A. SSSC Control System

The main components of SSSC controller are AC voltage regulator and DC voltage regulator. Fig. 6 (a) shows the MATLAB Simulink model of ordinary SSSC control system. Fig. 6 (b) illustrates the PI controller of AC voltage regulator and DC voltage of ordinary SSSC. The initial values of AC voltage (Kp_Vac and Ki_Vac) and gains of DC voltage (Kp_Vdc, Ki_Vdc) are specified by MATLAB Simulink model.



(a) MATLAB model of SSSC control system



(b) PI controller of DC and AC voltage regulators

Fig. 6. SSSC controller.

V. Applying the ANN Method

In this work the ANN is based on multi-layer feed-forward network. The multi-layer is divided into three layers input, hidden and output layer. Fig. 7 represents the flowchart of adjusting SSSC's parameters using ANN. As shown in Fig. 7, the neural fitting tool (NFTOOL) and the sample range are based on the value of control parameters of SSSC tuned by multi-objective genetic algorithm (SSSC MOGA) investigated in [15]. Also, Fig. 7 illustrates that the input signal is the change in voltage at the point of connection and the output layer represents the control parameters of SSSC (AC voltage regulator (Kp-vac and Ki-vac) and DC voltage regulator (Kp-vdc and Ki-vdc)). In this work, Levenberg-Marquardt algorithm is used for training the value of control parameters of SSSC tuned by multi-objective genetic algorithm (SSSC MOGA) investigated in [15].

The neural fitting tool (NFTOOL) is composed of sets of processes: training, validation and testing. The application divides input and target into three groups as follows: 70% is applied for training. 15% is applied to validate and 15% is applied to test. Table I shows the parameters of NFTOOL and illustrates the mean square error (MSE) and regression value (R) of training, validation and testing.

TABLE I. The Value of Gains of AC Voltage, DC Voltage Regulators Obtained Using Moga

Process	Selected percentage	Number of samples	MES	R value
training,	70%	6385	4.790e-5	1.859e-1
validation	30%	1362	4.8334e-5	1.909e-1
testing	30%	1362	4.8614e-5	1.826e-1
Type of algorithm	Levenberg-Marquardt and multi-objective genetic algorithm investigated in [15]			
Number of hidden neurons	20 neurons			

In this study NFTOOL is a feed forward network with input, output and multi-layers. Fig. 8 (a) illustrates the neural network size. Fig. 8 (b) illustrates training, validation and testing samples. Fig. 8 (c) illustrates results of training operation. Fig. 9 illustrates best validation performance. Fig. 10 illustrates the ANN controller of DC and AC voltage regulators.

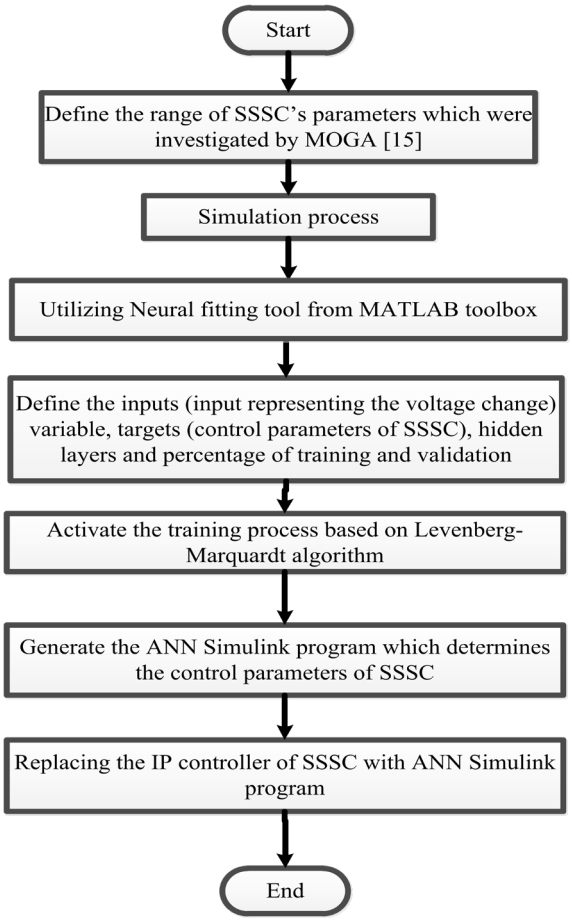
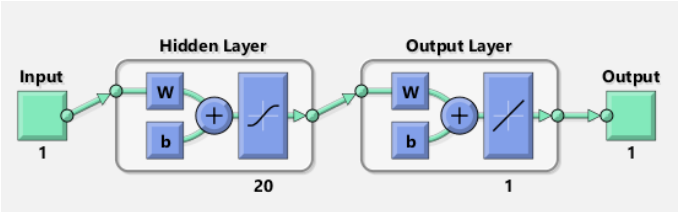
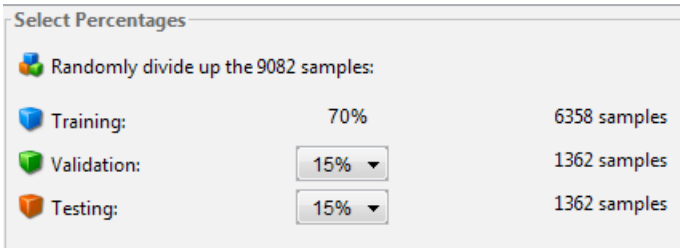


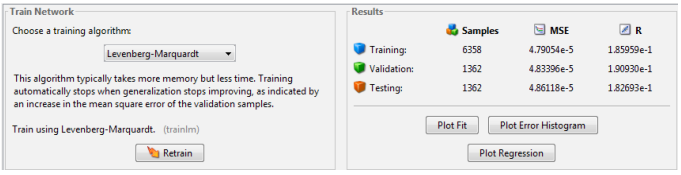
Fig. 7. The flow chart of adjusting SSSC's parameters using ANN operation.



(a) Neural network size



(b) Training, Validation and testing samples



(c) Results of training operation

Fig. 8. The size, training operation and the results.

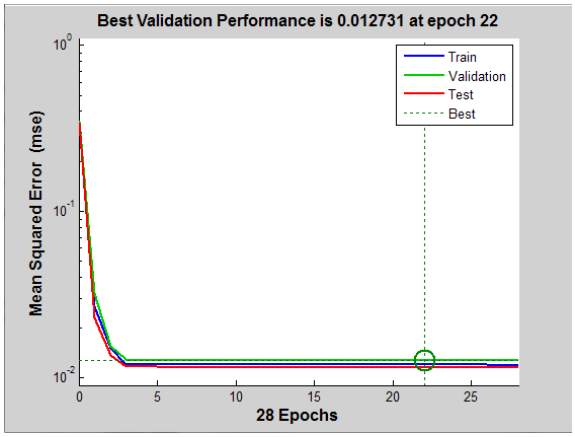


Fig. 9. The performance's convergence of the studied system with ANN.

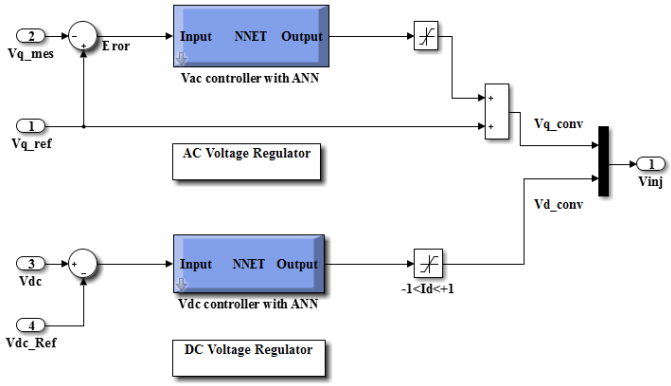


Fig. 10. ANN controller of DC and AC voltage regulators.

VI. Studied System Description

The studied system contains six wind turbines, each one produces 1.5 MW and 575v. The wind turbines are divided into three SCIG fixed-speed wind turbines and three DFIG variable speed wind turbines. Fig. 11 shows the block diagram of the studied system. A three phase fault is applied at 25 s and removes [e1] the fault after time equal 25.15 s.

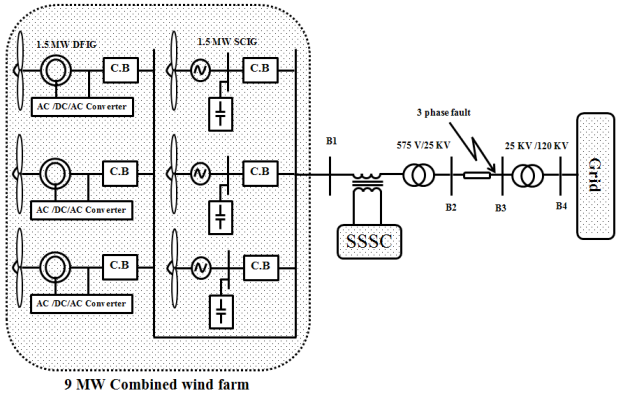


Fig.11. The block diagram of the system.

VII. Simulation Results

The simulation studied the performance of CWF with ANN-SSSC and CWF with ordinary SSSC (PI-SSSC) during three-phase fault. The voltage, reactive power and active power are measured at the point connection between the interconnected grid and wind farms.

The value of the control parameters of ordinary SSSC, MOGA SSSC and the proposed ANN SSC are shown in Table II.

TABLE II. The Value Of Gains of AC Voltage, DC Voltage Regulators of Ordinary SSSC, SSSC Moga and Ann SSSC

Type of SSSC	Regulator	Gain	Gain value
ordinary SSSC	AC voltage	Kp_Vac, Ki_Vac,	3.75e-3 0.1875
	DC voltage	Kp_Vdc, Ki_Vdc,	0.1e-3 20e-3
MOGA SSSC	AC voltage	Kp_Vac, Ki_Vac,	0.24742 0.98783
	DC voltage	Kp_Vdc Kp_Vdc,	0.06939. 0.95957
proposed ANN SSC	AC voltage	Kp_Vac, Ki_Vac,	0.1107 0.8665
	DC voltage	Kp_Vdc Kp_Vdc,	0.0753 0.8686

A. Impact of Three Fault

As illustrated in Fig. 12 the voltage with ANN-SSSC is 0.81 pu while the voltage with ordinary SSSC is 0.73 pu and the voltage with MOGA-SSSC is 0.75 pu. This means that the voltage of CWF has been improved with ANN-SSSC more than with the ordinary SSSC.

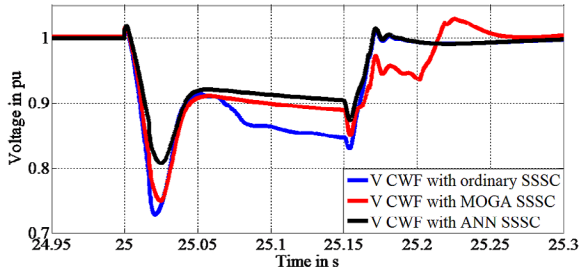


Fig. 12. The voltage of CWF with ANN-SSSC and CWF with ordinary SSSC during three phase fault.

This enhancement in voltage of CWF with ANN SSSC is due to the enhancement in the performance of SSSC when it is controlled by ANN. This can be observed by monitoring the injected voltage of SSSC in case of ordinary, MOGA and ANN as shown in Fig. 13.

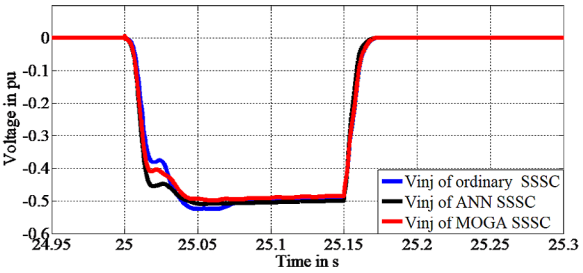


Fig. 13. The injected voltage of SSSC in case of ordinary, MOGA and ANN.

As illustrated in Fig. 14, the injected voltage of SSSC has been increased when it is controlled by ANN specially at the begging of fault period. This will decrease the reactive power absorbed by the CWF with ANN SSSC during fault.

As illustrated in Fig. 14 the reactive power of CWF with ANN-SSSC is-2.8 MVAR while the reactive power with ordinary SSSC is -4.75 MVAR and -3.83 MVAR for MOGA SSSC especially at the beginning of fault period. This means that the performance of CWF has been improved with ANN-SSSC more than the ordinary SSSC.

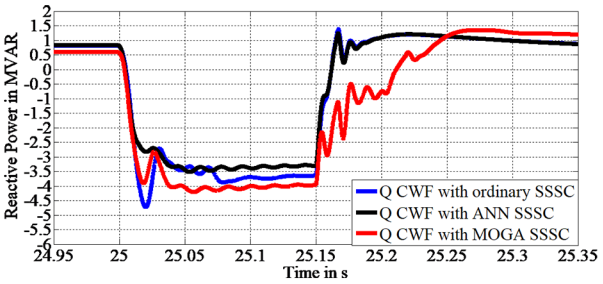


Fig. 14. The reactive power of CWF with ANN-SSSC and CWF with ordinary SSSC during three phase fault.

As illustrated in Fig. 15 the active power of CWF with ANN-SSSC has the highest value of output power especially at the beginning of fault period.

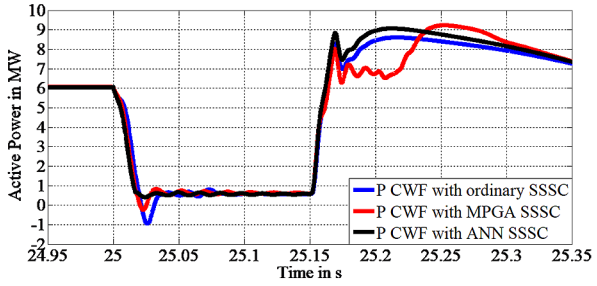


Fig. 15. The active power of CWF with ANN-SSSC and CWF with ordinary SSSC during three phase fault.

In order to justify the good performance of the proposed method, root mean square Error (RMSE) is used to measure the impact of ordinary SSSC, MOGA SSSC and ANN SSSC on the performance of CWF during the fault. RMSE is used to measure the error between the reference voltage (Vref = 1 pu) and the actual voltage at the point of common connection during the fault condition. Table III shows the RMSE of the three cases during fault period and after fault clearance.

TABLE III. The Value of RMSE During Fault Period and After Fault Clearance

Period	Ordinary SSSC	MOGA SSSC	ANN SSSC
RMSE during fault period	0.2063	0.1959	0.1805
RMSE after fault clearance	0.1918	0.1721	0.1405

From Table III, it can be observed that the CWF with ANN SSSC has the lowest RMSE. This means that the ANN managed to tune SSSC and gives the best result.

VIII. Conclusions

This paper has presented the design of a CWF which consists of two types of induction generator, the first one is SCIG fixed speed wind turbines and the second one is DFIG variable speed wind turbines. ANN has been used in order to adjust SSSC's parameters to enhance the performance of combined wind farm (CWF). In addition, this paper includes a comparison among the performances of combined wind farm (CWF) with ANN-SSSC, with performances of CWF with ordinary SSSC and performances of CWF with SSSC tune by Multi-objective genetic algorithm (MOGA SSSC). The performances of CWF with ANN-SSSC, ordinary and SSSC MOGA SSSC have been studied during the three-phase fault. The obtained results showed that the adjusted SSSC using ANN had enhanced the active power, the voltage and the reactive power of CWF, particularly during the three-phase fault.

IX. Future Work

This paper opens the door for many future works, for example:

1. Using ANN hybrid with different methods of optimization to determine optimal values for different types of FACTS to improve the performance of wind stations.
2. Use different types of ANN like ANFIS with different methods of optimization to determine optimal values for different types of FACTS to improve wind station performance.

Acknowledgment

The authors gratefully acknowledge the contribution of the NSFC (China)-ASRT (Egypt) Joint Research Fund, Project No. 51861145406 for providing partial research funding to the work reported in this research.

References

[1] D. Ananth and G. N. Kumar, "Mitigation of voltage dip and power system oscillations damping using dual STATCOM for grid connected DFIG," *Ain Shams Engineering Journal,* 2015.

[2] M. G. Jolfaei, A. M. Sharaf, S. M. Shariatmadar, and M. B. Poudeh, "A hybrid PSS–SSSC GA-stabilization scheme for damping power system small signal oscillations," *International Journal of Electrical Power & Energy Systems,* vol. 75, pp. 337-344, 2016.

[3] D. Ananth, G. N. Kumar, D. D. Chowdary, and K. A. Naidu, "Damping of Power System Oscillations and Control of Voltage Dip by Using STATCOM and UPFC." International Journal of Pure and Applied Mathematics, vol.114(10), pp 487-496, August 2017.

[4] P. Bhatt, "Short term active power support from DFIG with coordinated control of SSSC and SMES in restructured power system," in *Power and Energy Engineering Conference (APPEEC), 2014 IEEE PES Asia-Pacific*, 2014, pp. 1-6.

[5] L. Wang and Q.-S. Vo, "Power flow control and stability improvement of connecting an offshore wind farm to a one-machine infinite-bus system using a static synchronous series compensator," *IEEE transactions on sustainable energy,* vol. 4, pp. 358-369, 2013.

[6] H. Heydari-Doostabad, M. R. Khalghani, and M. H. Khooban, "A novel control system design to improve LVRT capability of fixed speed wind turbines using STATCOM in presence of voltage fault," *International Journal of Electrical Power & Energy Systems,* vol. 77, pp. 280-286, 2016.

[7] B. Boubekeur, A. Gherbi, and H. Mellah, "Application of STATCOM to increase transient stability of wind farm," *American Journal of Electrical Power and Energy Systems,* vol. 2, pp. 50-56, 2013.

[8] R. M. Pereira, C. M. Ferreira, and F. M. Barbosa, "Comparative study of STATCOM and SVC performance on dynamic voltage collapse of an electric power system with wind generation," *IEEE Latin America Transactions,* vol. 12, pp. 138-145, 2014.

[9] H. A. Mohammadpour, A. Ghaderi, H. Mohammadpour, and M. H. Ali, "Low voltage ride-through enhancement of fixed-speed wind farms using series FACTS controllers," *Sustainable Energy Technologies and Assessments,* vol. 9, pp. 12-21, 2015.

[10] A. M. Rashad and S. Kamel, "Enhancement of Hybrid Wind Farm performance using tuned SSSC based on Multi-Objective Genetic Algorithm," in *Power Systems Conference (MEPCON), 2016 Eighteenth International Middle East*, 2016, pp. 786-791.

[11] A. Rashad, S. Kamel, and F. Jurado, "Stability improvement of power systems connected with developed wind farms using SSSC controller," *Ain Shams Engineering Journal,* 2017.

[12] O. Nadjemi, T. Nacer, A. Hamidat, H. Salhi, " Optimal hybrid PV/wind energy system sizing: Application of cuckoo search algorithm for Algerian dairy farms," Renewable and Sustainable Energy Reviews, vol 70, April

2017, pp 1352-1365.

[13] Elkasem A. H. A., S. Kamel, A. Rashad, and F. Jurado, "Optimal Performance of Doubly Fed Induction Generator Wind Farm Using Multi-Objective Genetic Algorithm," *International Journal of Interactive Multimedia and Artificial Intelligence,* vol. In Press, pp. 1-6, April 2019.

[14] A. H. Elkasem, S. Kamel, A. Rashad, and F. Jurado, "Optimal Performance of DFIG Integrated with Different Power System Areas Using Multi-Objective Genetic Algorithm," in *2018 Twentieth International Middle East Power Systems Conference (MEPCON)*, 2018, pp. 672-678.

[15] A. M. Rashad and S. Kamel, "Enhancement of Hybrid Wind Farm performance using tuned SSSC based on Multi-Objective Genetic Algorithm," in *2016 Eighteenth International Middle East Power Systems Conference (MEPCON)*, 2016, pp. 786-791.

[16] N. K. Saxena and A. Kumar, "Reactive power control in decentralized hybrid power system with STATCOM using GA, ANN and ANFIS methods," *International Journal of Electrical Power & Energy Systems,* vol. 83, pp. 175-187, 2016.

[17] A. Rashad, S. Kamel, F. Jurado, M. Abdel-Nasser, and K. Mahmoud, "ANN-Based STATCOM Tuning for Performance Enhancement of Combined Wind Farms," *Electric Power Components and Systems,* pp. 1-17, 2019.

[18] J. Bhukya and V. Mahajan, "Optimization of damping controller for PSS and SSSC to improve stability of interconnected system with DFIG based wind farm," *International Journal of Electrical Power & Energy Systems,* vol. 108, pp. 314-335, 2019.

[19] D.-N. Truong and V.-T. Ngo, "Designed damping controller for SSSC to improve stability of a hybrid offshore wind farms considering time delay," *International Journal of Electrical Power & Energy Systems,* vol. 65, pp. 425-431, 2015.

[20] B. Minasny and A. McBratney, "The neuro-m method for fitting neural network parametric pedotransfer functions," *Soil Science Society of America Journal,* vol. 66, pp. 352-361, 2002.

[21] B. Wu, Y. Lang, N. Zargari, and S. Kouro, *Power conversion and control of wind energy systems*: John Wiley & Sons, 2011.

[22] M. G. Simoes and F. A. Farret, *Alternative energy systems: design and analysis with induction generators* vol. 13: CRC press, 2011.

[23] Z. Chen, "Wind turbine drive train systems," *Wind Energy Systems, Optimising design and construction for safe and reliable operation,* pp. 208-246, 2010.

Yousry Ibrahim

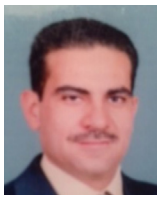
Yousry Ibrahim received the B.Eng. from Faculty of Engineering, Aswan University, Egypt in 2013. He is currently pursuing his MSc degree in Department of Electrical Engineering, Aswan Faculty of Engineering, Aswan University. His research activities include renewable energy and power system optimization.

Salah Kamel

Salah Kamel received the international PhD degree from University of Jaen, Spain (Main) and Aalborg University, Denmark (Host) in Jan. 2014. He is an Assistant Professor in Electrical Engineering Department, Aswan University. Also, he is a leader for power systems research group in the Advanced Power Systems Research Laboratory (APSR Lab), Aswan, Egypt. He is currently a Postdoctoral Research Fellow in State Key Laboratory of Power Transmission Equipment and System Security and New Technology, School of Electrical Engineering, Chongqing University, Chongqing, China. His research activities include power system modeling, analysis and simulation, and applications of power electronics to power systems and power quality.

Ahmed Rashad

Ahmed Rashad received the B.Eng. from Faculty of Energy Engineering, Aswan University, Egypt and M.Sc. degree in electrical power engineering from Faculty of Engineering, South Valley University, Egypt in 2013. He is currently pursuing his jointly-supervised Ph.D. degree in Department of Electrical Engineering, Aswan Faculty of Engineering, Aswan University, Egypt and University of Jaen, Linares, Jaen, Spain. His research activities include power system modeling, analysis and simulation, renewable energy and smart grid technologies, applications of power electronics to power systems and power quality.



Loai Saad Eldeen Nasrat

Loai Saad Eldeen Nasrat received his B.Sc, M.Sc, and Ph.D degrees in Electrical Engineering from Ain Shams University, Egypt in 1988, 1994, and 1999, respectively. He received a Diploma in Computers and Programming from American University in Cairo, 1994. Since 2002, he has been associated with the Department of Electrical Engineering, Aswan University, Egypt as an Assistant Professor, and since 2008, as an Associate Professor, and since 2013, as a Professor. He is the Vice Dean of Faculty of Engineering for Students and Education Affaires, Aswan University, 2008- till present and the Head of electrical engineering department, Faculty of Engineering, Aswan University from 2006 to 2008. He is currently the Dean of Faculty of Engineering, Aswan University, Egypt. He worked as a part time professor in various Egyptian universities; October 6 University, Alfauom University, Arab Academy for Science, Technology & Maritime Transport, Cinema High Institute, Alazhar University, New Cairo Academy. His research interests include power systems quality and renewable energy.

Francisco Jurado

Francisco Jurado obtained the MSc and PhD degrees from the UNED, Madrid, Spain, in 1995 and 1999 respectively. He is Full Professor at the Department of Electrical Engineering of the University of Jaén, Spain. His research activities have focused on two topics: power systems and renewable energy.

Voltage Stability Enhancement Based on Optimal Allocation of Shunt Compensation Devices Using Lightning Attachment Procedure Optimization

Salah Kamel[1,2]*, Heba Youssef[3]

[1] Department of Electrical Engineering, Faculty of Engineering, Aswan University, 81542 Aswan (Egypt)
[2] State Key Laboratory of Power Transmission Equipment & System Security and New Technology, Chongqing University, 400030 Chongqing (China)
[3] Upper Egypt Electricity Distribution Company, Ministry of Electricity, Aswan (Egypt)

Received 19 December 2018 | Accepted 26 July 2019 | Published 30 October 2019

unir
LA UNIVERSIDAD EN INTERNET

Abstract

This paper proposes a combined approach to determine the optimal allocation of different shunt compensation devices (shunt capacitor, static var compensator, and static synchronous compensator) in power systems. The developed approach is a combination between Lightning Attachment Procedure Optimization (LAPO) and loss sensitivity indices (LSIs). Different objective functions such as enhancement of voltage stability index, improvement of voltage profile and minimization of total power losses are considered. Two loss sensitivity indices (LSIs) are developed to reduce the search space in all buses and the total computation time. The developed algorithm is validated using standard IEEE 14-bus and IEEE 30-bus test systems. The developed algorithm successes to achieve the objective functions with the better performance compared with other well-known optimization techniques such as Teaching learning-based optimization (TLBO), genetic algorithm (GA) and particle swarm optimization (PSO).

Keywords

Voltage Stability, Shunt Compensation Devices, Optimal Power Flow, Lightning Attachment Procedure Optimization, Sensitivity Indices.

DOI: 10.9781/ijimai.2019.10.001

I. Introduction

SUNT compensation is basically a powerful tool to improve voltage profile and increase the steady state transmittable power. Voltage stability problems are solved by providing sufficient reactive power at the appropriate location, thus improving the voltage profile and reducing power loss [1].

There are various kinds of shunt compensation devices and each one has its own characteristic and boundary [2]. The shunt capacitor (C), static var compensator (SVC), and static synchronous compensator (STATCOM) are used to improve the static voltage stability margin and power transfer efficiency. However, the performance of SVC and STACOM is better in terms of reducing power losses and improving voltage profile [3]. To achieve the benefits of loss reduction and voltage profile improvement, engineers are in need to determine the optimal location and size of the shunt compensation devices.

Several techniques have been presented for solving the problem of the optimal capacitor placement in power system such as analytical, numerical programming, heuristic, artificial intelligence-based techniques [4], Combined Optimization Approach [5], Ant-lion optimization (ALO) [6, 7], Analytical Technique [8]-[10], combined algorithm based on Loss Sensitivity Factor and Salp Swarm Algorithm [11], combined algorithm based on Fuzzy Loss Sensitivity Factor with Sine Cosine Algorithm [12]. The genetic algorithm (GA), simulated annealing (SA), artificial immune system (AIS), Pareto Envelope-based Selection Algorithm II (PESA-II) with fuzzy logic decision maker and particle swarm optimization (PSO) have been used to determine the optimal placement of SVC in power system [13] [14]. In [14], global harmony search algorithm (NGHS) has been used to determine the optimal allocation of STATCOM. In addition, different optimization techniques have been used for determining the optimal allocation of D-STATCOM in distribution systems [15]-[21]. However, the determination of optimal allocation of shunt compensation devices is still hot topic and needs more effort to achieve the maximum benefits from installation of these devices in electrical power systems.

This paper uses a new optimization technique, referred to as the Lightning Attachment Procedure Optimization (LAPO), to determine the optimal placement and sizing of shunt Var compensators in power systems. Lightning Attachment Procedure Optimization (LAPO) is a new optimization method that simulates the attachment procedure of lightning. LAPO has robustness, high quality and is able to disband a lot of troubles [4].

Sensitivity analysis has been applied to determine the candidate buses in order to reduce the search space in all buses and the total computation time [22] [23].

In this paper, the optimal allocation of such compensation devices in power systems is determined using the developed optimization algorithm. In this algorithm, the candidate buses, which are considered

* Corresponding author.
E-mail address: skamel@aswu.edu.eg

the most suitable buses to connect with the compensation devices, are determined based on the loss sensitivity indices (LSIs). This step is necessary for reducing the search space and computation time. Then, LAPO is applied to compute the best size and determine the appropriate kind of shunt compensation devices in power systems. The results are contrasted with modern optimization techniques such as Teaching learning-based optimization (TLBO), genetic algorithm (GA) and particle swarm optimization (PSO) to confirm the applicability of the proposed technique. Two test systems are relied; the IEEE 14-bus and IEEE 30-bus system.

The rest of this paper is organized as follows: Section II presents an overview about the shunt compensation devices. The problem formulation is described in Section III. Section IV presents the sensitivity analysis. The LAPO algorithm is presented in Section V. Section VI presents the numerical results of the developed LAPO approach based on two standard test systems. Finally, the conclusions are presented in Section VII.

II. Shunt Compensation Devices

The capacitors are the conventional type of compensation devices that are considered relatively inexpensive. The capacitor injects reactive power and is connected in parallel with system buses. Its injected reactive power is proportional to the square of the terminal voltage. Flexible Alternating Current Transmission Systems (FACTS) are the modern type of compensation devices. They are widely used with electric power systems to improve the system loadability, security and maximize exploitation of transmission lines [24] [25]. There are different types of FACTS devices such as SVC and STATCOM.

SVC is capable to extend fast acting reactive power reparation in electrical systems. In other meaning, SVCs have their output edit for interchange inductive or capacitive current to control the voltage, power factor, harmonics and stabilizing the system. SVC includes two groups; the first group consists of a fixed capacitor (FC) and a thyristor-controlled reactor (TCR), while the second group consists of a thyristor-switched capacitor (TSC) and TCR [26] [27]. Schematic diagram of SVC is shown in Fig. 1.

STATCOM is an advanced device that depends on a power electronics voltage source converter (VSC) and it can control the injected reactive power to the system. In addition, it can extend active power when it is connected to a source of power. Schematic diagram of STATCOM is shown in Fig. 2.

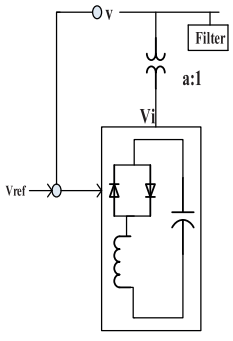


Fig. 1. SVC.

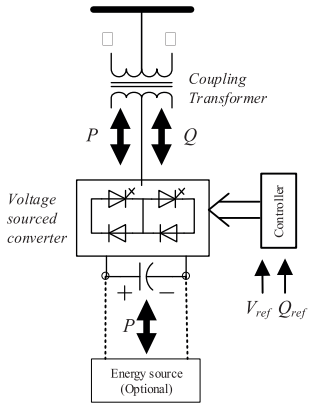


Fig. 2. STATCOM.

Commercially, shunt compensation devices (capacitor, SVC and STATCOM) have constraints. However, this study has been achieved under injected Q with limit value 50 MVar and the suitable types of shunt compensation devices are given in Table I [14].

TABLE I. Types of Appropriate Shunt Compensation Devices Based on the Injected Reactive Power

Shunt controller	Qmin (MVAR)	Qmax (MVAR)
Capacitor	-10	10
SVC	-30	30
STATCOM	-50	50

III. Objective Functions

A. Voltage Profile Refinement

The voltage profile enhancement is the first objective function, which seeks to minimize the voltage deviations of load buses which can be represented as:

$$F_1 = \sum_{i=1}^{NPQ} (V_i - 1) \quad (1)$$

where, NPQ is the number of load buses and V_i is the voltage of load bus i.

B. Voltage Stability Improvement

The System stability is indicated by voltage stability index (L_{max}), which varies between 0 to 1. This objective function is given by:

$$F_2 = min(L_{max}) = min\,(max\,(L_b)) \quad b= 1,2, \ldots, NPQ \quad (2)$$

The voltage stability index at bus j is given by (3).

$$L_j = \left|1 - \sum_{i=1}^{NPV} F_{ij} \frac{V_i}{V_j}\right| \quad (3)$$

where, V_i is the voltage of i^{th} generator bus and V_j is the voltage of load bus.

Calculate F_{ij} using Y_{bus} matrix (4) – (6).

$$\begin{bmatrix} I_G \\ I_L \end{bmatrix} = \begin{bmatrix} Y_{GG} & Y_{GL} \\ Y_{LG} & Y_{LL} \end{bmatrix} \begin{bmatrix} V_G \\ V_L \end{bmatrix} \quad (4)$$

where, the complex voltages and currents at the load and generator are V_L, V_G, I_L, I_G. Sub-matrices of system Y_{bus} are Y_{LL}, Y_{GG}, Y_{GL}, Y_{LG}. Eq. (4) can be written as in (5) and F_{ij} is given by (6).

$$\begin{bmatrix} V_L \\ I_G \end{bmatrix} = \begin{bmatrix} Z_{LL} & F_{LG} \\ K_{GL} & Y_{GG} \end{bmatrix} \begin{bmatrix} I_L \\ V_G \end{bmatrix} \quad (5)$$

$$F_{ij} = [F_{LG}] = -[Y_{LL}]^{-1}[Y_{LG}] \quad (6)$$

C. Real Power Losses Diminishing

Minimizing real power losses is the third objective function that is given by:

$$F_3 = P_{loss} = \sum_{i=1}^{NTL} G_{ib}(V_i^2 + V_b^2 - 2V_iV_bcos\delta_{ib}) \quad (7)$$

where, NTL is the number of transmission lines, G_{ib} is the conductance of the transmission ib and δ_{ib} is phase difference between voltages of buses i, b.

IV. Constraints

A. Equivalent Constraints

The balanced load flow equations are given by:

$$P_{Gi} - P_{Di} = |V_i| \sum_{b=1}^{NB} |V_j| (G_{ib}cos\delta_{ib} + B_{ib}sin\delta_{ib}) \quad (8)$$

$$Q_{Gi} - Q_{Di} = |V_i| \sum_{b=1}^{NB} |V_b| (G_{ib}sin\delta_{ib} - B_{ib}cos\delta_{ib}) \quad (9)$$

where, the generated reactive and active power at bus i are Q_{Gi} and

P_{Gi}, respectively, the reactive and active load demand at bus i are Q_{Di} and P_{Di}, respectively, the susceptance and conductance between buses b and i are B_{ib} and G_{ib}, respectively.

B. Inequality Constraints

The inequality constraints can be classified as follows:

1. Generated active and reactive powers, generators bus voltages, voltage magnitude of load buses and transformer tap settings which are within minimum and maximum values.
2. Apparent power flow in transmission lines $\geq S_{TL}^{min}$.
3. Compensation MVAR $(-50 \leq Q_{comd} \leq 50)$.

The operating constraints variables of the system must be considered in the objective functions. These variables can be easily limited in optimization solution using the quadratic penalty formulation of the objective functions for all dependent variables; therefore, the generalized objective function can be expressed as follows:

$$F_g(x,u) = F_i(x,u) + h_G\left(P_{G1} - P_{G1}^{lim}\right)^2 + h_Q \sum_{i=1}^{NPV}\left(Q_{Gi} - Q_{Gi}^{lim}\right)^2 + h_V \sum_{i=1}^{NPQ}\left(V_{Li} - V_{Li}^{lim}\right)^2 + h_S \sum_{i=1}^{NTL}\left(S_{Li} - S_{Li}^{lim}\right)^2 + +h_F\left(V_{sh} - V_{sh}^{lim}\right)^2 + h_t\left(\theta_{sh} - \theta_{sh}^{lim}\right)^2 \quad (10)$$

where, h_G, h_V, h_Q, h_S, h_F and h_t are the penalty factors. These values are high positive. x^{lim} acts as the limit values of the dependent variable as:

$$x^{lim} = \begin{cases} x^{max}; & x > x^{max} \\ x^{min}; & x < x^{min} \end{cases} \quad (11)$$

V. Sensitivity Analysis

The loss sensitivity analysis consists of two LSIs to determine the candidate buses for the existence of shunt reparation devices. Consequently, the search space in all the buses, and the simulation time will be reduced. Moreover, places are arranged according to their severity for efficient detecting the candidate load buses. Fig. 3 shows two-bus system connected by a line as a part of a large power system. Buses p and q represent the sending and receiving end buses, respectively.

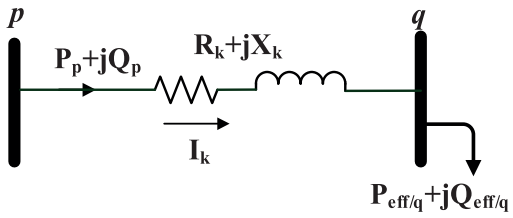


Fig. 3. Two-bus system.

The reactive and active powers at the node p can be given as:

$$P_p = P_{eff/q} + P_{lossk} \quad (12)$$

$$Q_p = Q_{eff/q} + Q_{lossk} \quad (13)$$

where, the power passes through line k are P_p and Q_p. The total reactive and active load power around bus q are $Q_{eff/q}$ and $P_{eff/q}$, respectively. The reactive and active power losses through line k are Q_{Lossk} and P_{Lossk}, respectively.

The current passing through line k from point p to point q is given as:

$$I_k = \frac{P_p - jQ_p}{V_p < -\delta_P} = \frac{V_p < \delta_p - V_q < \delta q}{R_k + jX_k} \quad (14)$$

where, the voltage magnitudes at point p and q are V_p and V_q, respectively, the voltage angles at point p and q are δ_p and δ_q, respectively, the reactance and resistance of line k are X_k and R_k respectively.

Eq. (13) can be manipulated as in (14).

$$V_p^2 - V_p V_q < \left(\delta_q - \delta_p\right) = (P_p - jQ_p)(R_k + jX_k) \quad (15)$$

The imaginary and real parts of (15) are equate, and then squared and added to (16):

$$V_q^2 = V_p^2 - 2 \times \left(P_p R_k + Q_p X_k\right) + \left(\frac{P_P^2 + Q_P^2}{V_p^2}\right)(R_k^2 + X_k^2) \quad (16)$$

The reactive and active power through line k are given by:

$$P_{Lossk} = I_k^2 \times R_k = \left(\frac{P_{eff/q}^2 + Q_{eff/q}^2}{|V_q|^2}\right) \times R_k \quad (17)$$

$$Q_{Lossk} = I_k^2 \times X_k = \left(\frac{P_{eff/q}^2 + Q_{eff/q}^2}{|V_q|^2}\right) \times X_k \quad (18)$$

The total reactive and active power losses of power system are given by:

$$P_{Loss} = \sum_{k=1}^{NB} P_{Lossk} \quad (19)$$

$$Q_{Loss} = \sum_{k=1}^{NB} Q_{Lossk} \quad (20)$$

Calculate LSI_1 by the first derivative of P_{Lossk} in (19) with respect to $|V_q|$, as follows:

$$LSI_1 = \frac{\partial P_{Lossk}}{\partial |V_q|} = -2 \times \left(\frac{P_{eff/q}^2 + Q_{eff/q}^2}{|V_q^3|}\right) \times R_k \quad (21)$$

Values of LSI_1 are listed from the smaller to larger in ascending order. The optimal buses to locate the shunt compensation devices are the highest negative values in the LSI_1.

Calculate LSI_2 by the first derivative of P_{Lossk} in (19) with respect to $Q_{eff/q}$ as follows:

$$LSI_2 = \frac{\partial P_{Lossk}}{\partial Q_{eff/q}} = 2 \times \left(\frac{Q_{eff/q}}{|V_q^2|}\right) \times R_k \quad (22)$$

Values of LSI_2 are listed from the larger to smaller in descending order. The optimum buses to locate the shunt compensation devices are those with the highest positive values in the LSI_2. The optimum buses where the shunt compensation devices will be located are acquired from the top two lists by means of merger or union. The buses selection is approximately half of the total number of system buses (50% to 55%).

VI. Lighting Attachment Procedure Optimization (LAPO)

The Lightning Attachment Procedure Optimization (LAPO) [4] [28], is inspired by the nature of lightning attachment operation which contains the movement of the falling leader, spread of the rising leader, and the feature of lightning. Ultimate better result would be the lightning hit point. The suggestion method is free from any parameter setting and it is seldom stuck in the local best points.

A. Serious Phases of LAPO

1. Air Collapse on Cloud Surface

The cloud's charges classified to three stages are shown in Fig 4. In the top stage in the cloud a large value of positive charge is placed, in the down stage in the cloud a large value of negative charge is located, as well as a little positive charge. The lightning is created and a large value of electrical charge shifts across the earth, when the voltage gradient on the border of the cloud increases, because of, the potential is increased between the charge centers, the positive charges and negative charges.

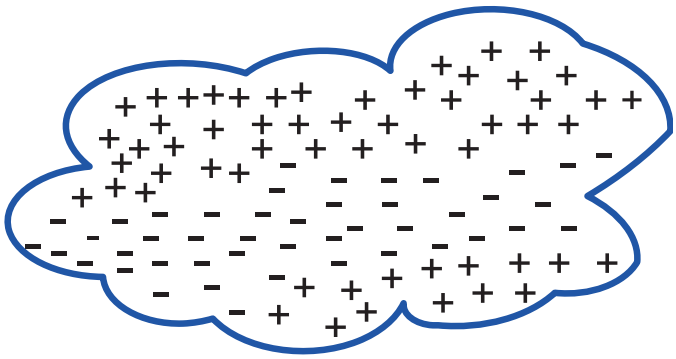


Fig. 4. Distribution of the charge in the cloud.

2. The Fall of Lighting Channel

In the cloud's edge air breakdown occurs, the lightning oncoming the earth in a stepwise motion. After each step, the lightning pauses, then shift to one or more other directions across the earth. With a view to know this operation, after each step, a hemisphere is seen beneath the leader tip with the center of leader tip and the radius of next step length (see Fig. 5) [4] [28]. There are more than one potential points on the face of this hemisphere, there are many of potential points which could be chosen as the next jump spot. The following jump spot is chosen randomly; yet, a spot with large value of electrical field between the line connecting the leader tip and the identical spot is more possible to be considered as the following jump.

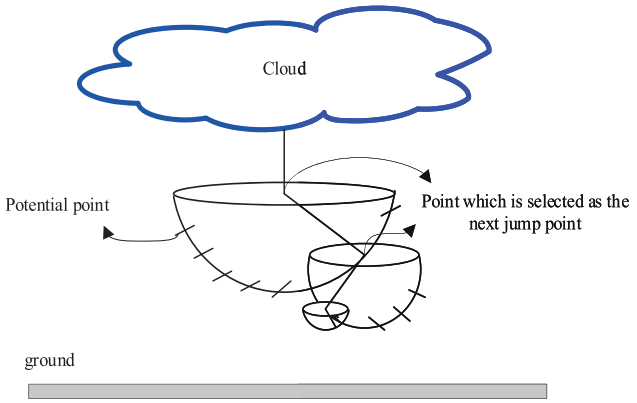


Fig. 5. The downward leader 's procedure of the gradient movement.

3. Section Fading

There are a lot of spots for the following jump of lightning, the charge of the top branch is classified into new branches. the new branches are created by the same steps. No air break-down happens when the charges of branch decrease more than a stringent value (1 μC) and the result is that no movement occurs. Thus, this branch would vanish as shown in Fig. 6.

4. Leader of the Rising

Existence of cloud implies existence of a large negative charge over the earth. This results in collecting of positive charges on the ground surface or earthed object beneath the cloud. In the heavy points, the high electric field produces air breakdown; thus, the heavy points start upward leader and spread through the air (see Fig. 7). As the downward leader is oncoming the ground, the upward leaders go across the downward leader quickly.

5. Ultimate Leaping

The ultimate jump happens when upward leader arrives to downward leader wand, the striking point would be the point from which the upward leader has started. In this situation, all the other branches disappear and charge of the cloud is naturalized through this channel.

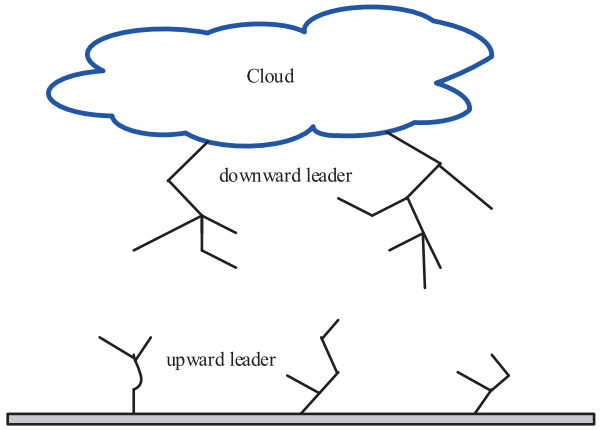


Fig. 6. Formation of the upward leader and diffusion through the downward leader.

B. Mathematical Steps of LAPO Algorithm

First step: Trail spot

Generated initial trail spots are placed at the cloud and earth edge. Several of these trail spots are the emitted lightning points, and some of them are the spot from which the upward leaders start. The trail spot is calculated by

$$Y^i_{trail\ spot} = Y^i_{min} + \left(Y^i_{max} - Y^i_{min}\right) * rand \tag{23}$$

Where *Ymin* and *Ymax* are minimal and maximum bound of variables, and rand is random variable in the range (0,1). The fitness function is evaluated depending on the objective function

$$F^i_{trail\ spot} = obj\ (Y^i_{trail\ spot}) \tag{24}$$

Second step: jump definition

The average of all trail spots and the fitness function according to trail spot's averages are calculated by Eqs:

$$Y_{avr} = mean\ (Y_{trail\ spot}) \tag{25}$$

$$F_{avr} = obj\ (Y_{avr}) \tag{26}$$

There are many potential spots for a test point, which the lightning can pass. Since the lightning was a random action, for test point i, a random point p is chosen between the population (i /= p). If the fitness of the point p is greater than that of the average value, the lightning leaps across this point, otherwise, the lightning shifts to another direction. And are given by

$$Y^i_{trail\ spot\_new} = Y^i_{trail\ spot} + rand * (Yav + (Y^p_{potential\ \ spot})) \tag{27}$$

If average fitness of point p is lower than the fitness point

$$Y^i_{trail\ spot\_new} = Y^i_{trail\ spot} - rand * (Yav + (Y^p_{potential\ \ spot})) \tag{28}$$

If average fitness of point p is higher than the fitness point.

Third Step: Section Fading

If the fitness function is higher than the prior point, the branch maintains; else, it fades, and are given by:

$$Y^i_{trail\ spot} = Y^i_{trail\ spot\_new}\ if\ \ F^i_{trail\ spot\_new} < F^i_{trail\ spot} \tag{29}$$

$$Y^i_{trail\ spot\_new} = Y^i_{trail\ spot}\ \ otherwise \tag{30}$$

This operation is executed for all the test points. In other words, in the first stage, all the remaining points are treated to shift down.

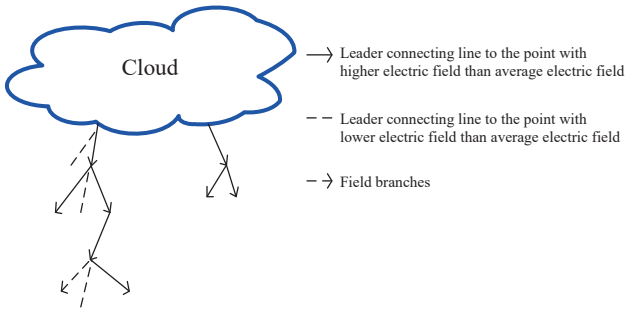


Fig. 7. Determine the next jump and determine the lightning path.

Fourth Step: Leader of the Rising

As explained in the prior steps, all the test points are treated as the downward leader and shifted down. In the second stage, all the test points are treated as the upward leader and moved above. The upward leader motion depends on the charge of the channels which is essentially spread exponentially, and is given by:

$$S = 1 - (\frac{j}{j_{max}}) * \exp(-\frac{j}{j_{max}}) \quad (31)$$

where the number of iterations is j, the maximum number of iterations is j_{max} , and next jump is s, which depends on the charge of the channel and the next point is given by:

$$Y_{trail\ sport\ new} = Y_{trail\ sport\ new} + rand * S(Y_{min} - Y_{max}) \quad (32)$$

Where Y_{min} and Y_{max} are the best and the worse solutions of the population.

Final Steps: Ultimate Leaping

The lightning operation pauses when the up leader and the down leader are gathering each other.

VII. Simulation Results

A. Validation Strategy & Used Parameters

The developed optimization technique is validated using two standard IEEE systems (IEEE 14-bus and IEEE 30-bus) with three various objective functions. The systems data are given in [29]. Three case studies are discussed for each system; (i) Case 1: One compensation device, (ii) Case 2: Two compensation devices, and (iii) Case 3: Three compensation devices.

The parameters of LAPO are population size = 20 and maximum number of iterations = 200. The parameters of TLBO are population specified size = 20 and maximum number of iterations = 200. The parameters of PSO are population size = 20, maximum iterations = 200, C1 =1.05, C2 = 1.1, ω_{max}= 1 and ω_{min}= 0.3 [30]. The penalty factors are specified to be 100. The simulations are executed for 40 trials and the optimum results are given below.

B. Obtained Results

Table II and Table III present the obtained LSIs for IEEE 14-bus and IEEE 30-bus systems, respectively. The candidate buses that have high opportunity for connecting with shunt compensation devices is at the beginning of LSIs table and vice versa for the buses located at the end of LSIs table. The candidate buses are 2,3,4,5,11,12,13 and 14 for IEEE 14-bus and 2,3,4,5,6,7,15,16,17,21,24,26,27,28 and 30 for IEEE 30-bus.

TABLE II. Order of Load Buses Depending on LSIS (IEEE 14-Bus Test System)

LSI_1		LSI_2	
Order	**Value**	**Order**	**Value**
2	-0.085021	5	0.010309
5	-0.083808	13	0.0035644
3	-0.049055	14	0.0015991
4	-0.025891	2	0.001478
13	-0.022128	4	0.00026831
14	-0.013788	12	0.00014508
12	0.0014145-	11	0.0001217
11	0.00049929 -	3	0.00011671
10	0.00022395	10	8.1518e-005
1	0	1	0
6	0	6	0
7	0	7	0
8	0	8	0
9	0	9	0

TABLE III. Arrangement of Load Buses Depending on LSIS for IEEE 30-Bus Test System

LSI_1		LSI_2	
Order	**Value**	**Order**	**Value**
6	-0.24189	6	0.0047946
4	-0.21416	15	0.0030409
2	-0.10295	28	0.0022654
3	-0.065464	2	0.0015673
5	-0.062066	7	0.0010889
15	-0.02248	24	0.00078636
30	-0.013854	21	0.00063556
7	-0.0089256	30	0.00058199
28	-0.007415	17	0.00053991
27	-0.0040301	27	0.00045571
24	-0.003341	26	0.00028421
21	-0.0021812	16	0.00021524
8	-0.0020607	23	0.00016113
29	-0.0017742	3	0.00013632
17	-0.0017486	22	0.00013588
14	-0.0014727	14	0.00013301
16	-0.0010829	29	0.00012081
26	-0.0009206	25	9.2603e-005
18	-0.0007789	18	5.8691e-005
22	-0.0006286	8	3.0762e-005
23	-0.0006245	5	2.8648e-005
25	-0.000151	20	2.0106e-005
20	-0.0001430	4	1.3554e-005
19	-9.9446e-01	19	6.3026e-006
1	0	1	0
9	0	9	0
10	0	10	0
11	0	11	0
12	0	12	0
13	0	13	0

The results obtained with LAPO are compared with those obtained by other optimization techniques. Tables IV, V and VI present the results for VSI, voltage deviation and real power losses, respectively. From these tables, it can be observed that in case of not using FACTS, the maximum voltage stability index (L_{max}) is 0.0669 p.u, voltage deviation is 0.0272 p.u and active power losses is 2.8178 MW, while the maximum voltage stability index after determining the optimal allocation of different shunt compensation devices decreases from 0.0669 p.u to 0.0645 p.u, voltage deviation decreases from 0.0272 p.u to 0.0068 p.u, and active power losses decreases from 2.8178 MW to 2.7571 MW for IEEE 14-bus and IEEE 30-bus, in case of not using FACTS, the maximum voltage stability index (L_{max}) is 0.1240 p.u, voltage deviation is 0.0866 p.u and active power losses is 3.0896 MW, while the maximum voltage stability index after determining the optimal allocation of different shunt compensation devices decreases from 0.1240 p.u to 0.0923 p.u, voltage deviation decreases from 0.0866 p.u to 0.0643 p.u, and active power losses decreases from 3.0896 MW to 2.8087 MW. Fig. 8 gives convergence characteristics of various optimization methods for VSI without compensation devices of IEEE 14-bus system. Fig. 9 shows the convergence characteristics of various optimization methods for power losses with one shunt compensation devices of IEEE 14-bus system. From this figure, it can be observed that the performance of developed algorithm is competing with other optimization techniques, while TLBO is considered the faster one. Fig. 10 gives the convergence characteristics of various optimization methods for power losses without shunt compensation devices of IEEE 30-bus system. From this figure, it can be observed that the performance of LAPO is competing with other optimization techniques, while GA is considered the worst one. Fig. 11 gives the convergence characteristics of various optimization methods for VSI with two shunt compensation devices of IEEE 30-bus system. It can be observed that the TLBO is the faster one compared with other optimization techniques. Fig. 12 gives convergence characteristics of various optimization methods for VSI with three shunt compensation devices of IEEE 30-bus system. Fig. 13 shows convergence characteristics of various optimization methods for VDD with three shunt compensation devices of IEEE 30-bus system. Fig. 14 gives convergence characteristics of various optimization methods for power losses with three shunt compensation devices of IEEE 30-bus system. Fig. 15 gives convergence characteristics for power losses using LAPO of IEEE 30-bus system. Fig.16 gives convergence characteristics for VSI using LAPO of IEEE 30-bus system. From Fig. 12 and Fig. 13, it can be observed that the performance of the developed algorithm is competing with other optimization techniques.

C. Outstanding Features of Developed Algorithm

The results obtained by LAPO are comparable with those obtained by the well-known optimization techniques. This verifies the applicability of LAPO for power system studies as it gives a minimum objective function compared with TLBO, PSO and GA techniques.

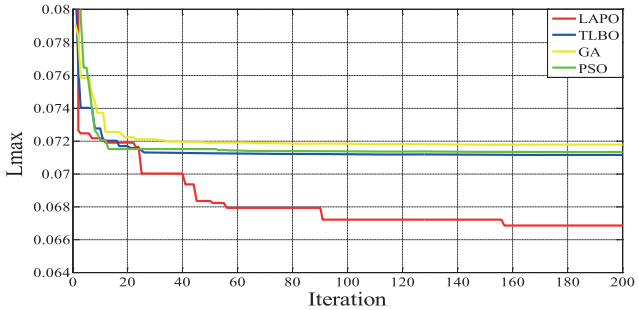


Fig. 8. Convergence characteristics of various optimization methods for VSI without compensation devices of IEEE 14-bus system.

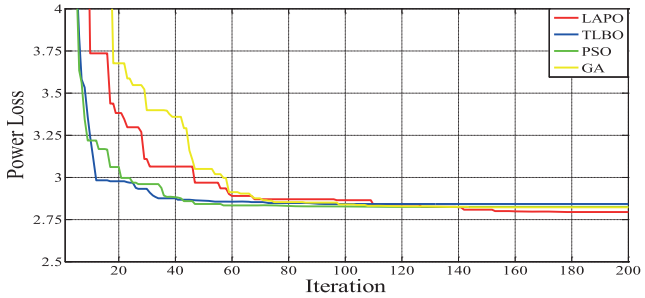


Fig. 9. Convergence characteristics of various optimization methods for power losses with one shunt compensation devices of IEEE 14-bus system.

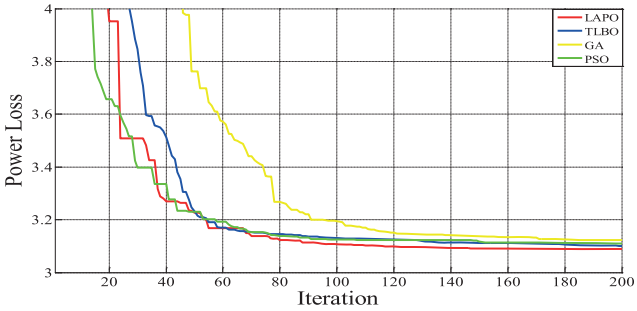


Fig. 10. Convergence characteristics of various optimization methods for power losses without shunt compensation devices of IEEE 30-bus system.

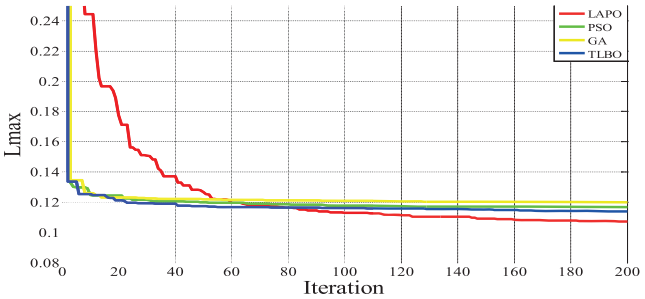


Fig. 11. Convergence characteristics of various optimization methods for VSI with two shunt compensation devices of IEEE 30-bus system.

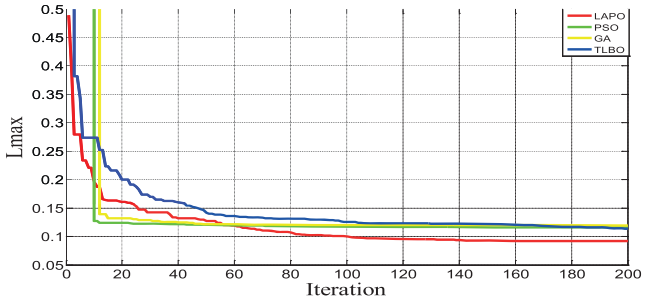


Fig. 12. Convergence characteristics of various optimization methods for VSI with three shunt compensation devices of IEEE 30-bus system.

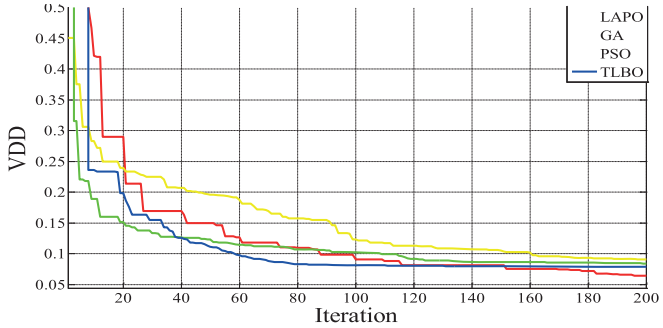


Fig. 13. Convergence characteristics of various optimization methods for VDD with three shunt compensation devices of IEEE 30-bus system.

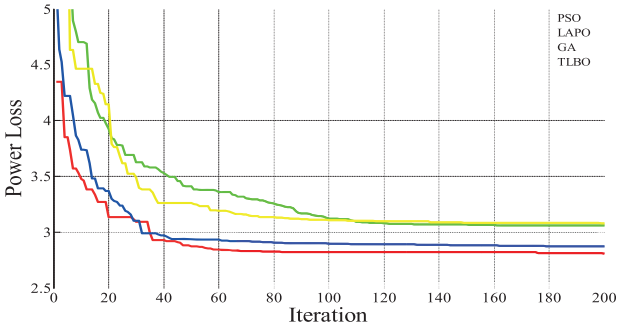


Fig. 14. Convergence characteristics of various optimization methods for power losses with three shunt compensation devices of IEEE 30-bus system.

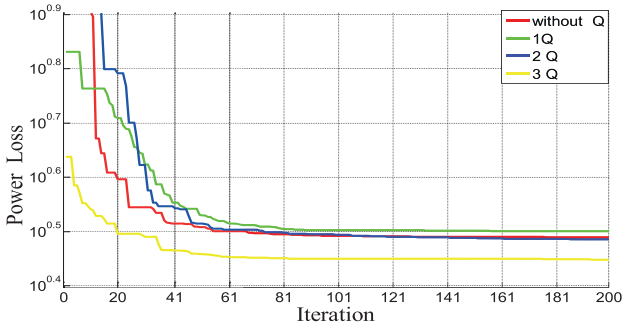


Fig. 15. Convergence characteristics for power losses using LAPO of IEEE 30-bus system.

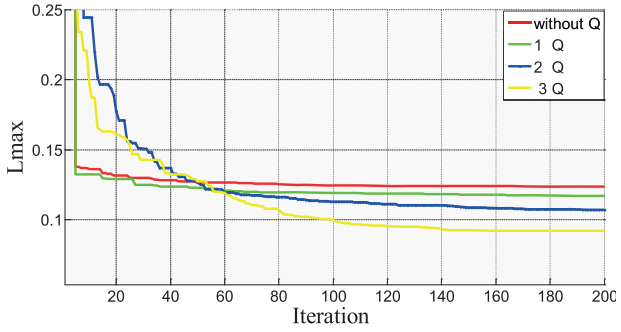


Fig. 16. Convergence characteristics for VSI using LAPO of IEEE 30-bus system.

TABLE IV. VSI, Optimal Placement, Size in MVAR and Suitable Type of Shunt Compensation Devices

		IEEE 14-bus system	IEEE 30-bus system
Without FACTS	LAPO	0.0669	0.1240
	TLBO	0.0712	0.12434
	PSO	0.0713	0.1253
	GA	0.0715	0.1257
Case 1 (value, place, size, and type)	LAPO	0.0656,14,38.3232, STATCOM	,21, 50, STATCOM 0.1173
	TLBO	0.06631, 4, 49.97, STATCOM	0.118, 21, 49.7337, STATCOM
	PSO	0.0703, 5, 33.45, STATCOM	0.125, 7, 22.8768, SVC
	GA	0.0708, 5, 49.8751, STATCOM	0.1216, 21, 23.8204, SVC
Case 2 (value, place, size, and type)	LAPO	0.0648, (13,14), (-50,45.4927), 2 STATCOM	0.1073, (7,27), (9.1307, 44.8712), C, STATCOM
	TLBO	0.0663, (4, 5), (-16.5310, 49.9692), SVC, STATCOM	0.11312, (30, 27), (-16.7354, 50), SVC, STATCOM
	PSO	0.0701, (5, 4), (49.9997, 23.721), STATCOM, SVC	0.1168, (21,15), (42.7866, 38.4338), 2 STATCOM
	GA	0.0664, (4, 5), (27.8177, 28.1106), 2 SVC	0.12001, (7, 17), (-41.2411, 46.7745), 2 STATCOM
Case 3 (value, place, size, and type)	LAPO	0.0645, (4,14,12), (-6.2314,45.0123, -50), C, 2 STATCOM	0.0923, (15,21,30),(35.1034,22.5432,39.1089) 2 STATCOM, SVC
	TLBO	0.0655, (5, 4, 5), (30.1891, -46.0102, 43.9262), 3 STATCOM	0.1112, (17, 21, 15), (26.174, 50.0, 35.377), SVC, 2 STATCOM
	PSO	0.0697, (4, 5, 4), (33.5932, 43.629, 30.7922), 3 STATCOM	0.1161, (21, 7, 3), (33.6635, 49.9984, 34.6893), 3 STATCOM
	GA	0.0662, (4, 14, 4), (-24.9, 30.92, -16.123), 2 SVC, STATCOM	0.1192, (21, 30, 15), (0.1669, 48.2160, 10.9165), C, STATCOM, SVC

TABLE V. Voltage Deviation in PU, Optimal Placement, Size in MVAR and Suitable Type of Shunt Compensation Devices

		IEEE 14-bus system	IEEE 30-bus system
Without FACTS	LAPO	0.0272	0.0866
	TLBO	0.03211	0.088319
	PSO	0.0327	0.0907
	GA	0.03286	0.095745
Case 1 (value, place, size, and type)	LAPO	0.0148, 14, 12.2541, SVC	0.0783,17,4.6421, C
	TLBO	0.0300, 14, -42.3003, STATCOM	0.086812, 15, 17.5913, SVC
	PSO	0.0324, 5, 43.4456, STATCOM	0.089608, 21, 24.3424, SVC
	GA	0.0328, 4, -8.8882, C	0.0943,21, -41.2308, STATCOM
Case 2 (value, place, size, and type)	LAPO	0.0079, (14,12), (12.6500.-1.7732), SVC, C	0.0779, (20,15), (5.6523,14.8365), C, SVC
	TLBO	0.0081, (14, 5), (13.0869, 15.3602), 2 SVC	0.083624, (3, 24), (-14.0, 15.92), 2 SVC
	PSO	0.0083, (14, 4), (13.0757, 15.0135), 2 SVC	0.0848, (21, 15), (4.6344, 16.8076), C, SVC
	GA	0.0183, (12, 14), (-23.9822, 11.6879), 2 SVC	0.092047, (24, 15), (10.8060, 1.2003), SVC, C
Case 3 (value, place, size, and type)	LAPO	0.0068, (4,5,14), (9.4621,7.0088,13.016), 2 C, SVC	0.0643, (15,26,30), (-3.9316,4.8339,5.1494), 3C
	TLBO	0.00794, (5, 14, 12), (16.2095, 13.0638, 22.1965), 3 SVC	0.0813, (4, 24, 15), (14.9389, 8.2276, 26.0513), 2 SVC, C
	PSO	0.0080, (14, 12, 5), (12.9847, 24.5797, 0.6175), 2 SVC, C	0.0843, (21, 7, 15), (7.3001, 19.8402, 12.6049), C, 2 SVC
	GA	0.0111, (14, 5, 4), (12.03, -36.644, 17.7964), 2 SVC, STATCOM	0.091, (30, 24, 7), (30, 4.8605, -49.9498), STATCOM, C

TABLE VI. Power Losses in MW, Optimal Placement, Size in MVAR and Suitable Type of Shunt Compensation Devices

		IEEE 14-bus system	IEEE 30-bus system
Without FACTS	LAPO	2.8178	3.0896
	TLBO	2.8193	3.0997
	PSO	2.8398	3.111
	GA	2.8542	3.1227
Case 1 (value, place, size, and type)	LAPO	2.7922,14,7.7322, C	3.0723,21,13,0046, SVC
	TLBO	2.8182, 14, -8.7037, C	3.0772, 7, 9.7524, C
	PSO	2.8254, 5, -36.9354, STATCOM	3.0905, 30, 44.9121, STATCOM
	GA	2.8204, 4, 6.0963, C	3.1137, 7, 8.6169, C
Case 2 (value, place, size, and type)	LAPO	2.7603, (5,13), (40.5023,0.7516), STATCOM, C	3.0625, (26,7), (2.2619,10.1625), C, SVC
	TLBO	2.7784, (13, 14), (8.6097, 7.1422), 2 C	3.0632, (21, 7), (11.6681, 7.4546), SVC, C
	PSO	2,7942, (5 ,13), (15.9081,9.1433), (SVC, C)	3.0679, (7, 21), (8.8934, 10.6073), C, SVC
	GA	2.812, (14, 5), (7.4191, -10.4339), C, SVC	3.0932, (21, 16), (6.8264, 4.1232), 2 C
Case 3 (value, place, size, and type)	LAPO	2.7571, (12,14,5), (1.7137,5.306, 0.4380), 3 C	2.8087, (4,21,27), (7,92828, -3.4041,29,0705), C, STATCOM, SVC
	TLBO	2.7599, (12, 14, 4), (22.9890, 7.2583, 9.0851), (SVC, 2 C)	2.8712, (5, 7, 21), (45.2345, 34.6556, 23.44323), 2 STATCOM, SVC
	PSO	2.78, (14, 5, 12), (7.0831, 9.4564, 10.3951), 2 C, SVC	3.0581, (30, 21, 7), (2.4394, 11.7325, 8.9302), 2 C, SVC
	GA	2.7837, (4, 14, 13), (-24.4101, 5.5066, 3.4499), SVC, 2 C	3.082, (21, 28, 7), (15.4645, -1.1633, 8.9918), SVC, 2 C

VIII. Conclusion

In this paper, a new hybrid optimization technique based on LAPO and loss sensitivity indices has been proposed to determine the optimal allocation of different shunt compensation devices in power systems. Two LSIs have been developed to determine the candidate locations for the existence of shunt compensation devices in order to decrease the search time in all buses and accelerate the convergence. The proposed optimization technique has been used to achieve different objective functions; voltage stability index, improvement of voltage profile and minimization of total power losses. IEEE 14-bus and IEEE 30-bus test systems have been used to verify the optimization algorithm. The results of the proposed algorithm have been compared with those obtained by other well-known optimization techniques such as TLBO, PSO and GA. The obtained results proved the capability of the proposed algorithm to effectively determine the optimal allocation of such compensation devices and achieve different objective functions during fast computation time. Other compensation devices such as; UPFC, IPFC and CUPFC have not been studied using the developed technique which maybe faces some challenges with them, as they are more complex devices. Hence, the future work will be focused to solve this issue.

Acknowledgment

The authors gratefully acknowledge the contribution of the NSFC (China)-ASRT (Egypt) Joint Research Fund, Project No. 51861145406 for providing partial research funding to the work reported in this research.

References

[1] F. Dincer, "The analysis on wind energy electricity generation status, potential and policies in the world" Renewable and sustainable energy reviews 15.9 (2011): 5135-5142.

[2] P. Pourbeik, P. S. Kundur, and C. W. Taylor, "The anatomy of a power grid blackout-Root causes and dynamics of recent major blackouts" IEEE Power and Energy Magazine 4.5 (2006): 22-29.

[3] A. Sode-Yome, and N. Mithulananthan, "Comparison of shunt capacitor, SVC and STATCOM in static voltage stability margin enhancement" International Journal of Electrical Engineering Education 41.2 (2004): 158-171.

[4] A. Foroughi Nematollahi, A. Rahiminejad, and B. Vahidi, "A novel physical based meta-heuristic optimization method known as Lightning Attachment Procedure Optimization" Applied Soft Computing 59 (2017): 596-621.

[5] A. Youssef, S. Kamel, M. Ebeed, and J. Yu, "Optimal Capacitor Allocation in Radial Distribution Networks Using a Combined Optimization Approach" Electric Power Components and Systems 46.19-20 (2018): 2084-2102.

[6] C. G. Boktor, A. Youssef, A. H. Ali, and S. Kamel, "Optimal Distribution Power Flow Including Shunt Capacitor Allocation Based on Voltage Deviation and Power Loss minimization" IEEE Nineteenth International Middle East Power System conference (MEPCON 2017), Cairo, Egypt, December 19-21, 2017.

[7] T. George, A. Youssef, M. Ebeed, and S. Kamel, "Ant Lion Optimization Technique for Optimal Capacitor Placement Based on Total Cost and Power Loss Minimization" IEEE International Conference on Innovative Trends in Computer Engineering (ITCE 2018), Aswan, Egypt, February 19-21, 2018.

[8] A. A. Mohamed, S. Kamel, and M. M. Aly, "A Simple Analytical Technique for Optimal Capacitor Placement in Radial Distribution Systems" IEEE Nineteenth International Middle East Power System conference (MEPCON 2017), Cairo, Egypt, December 19-21, 2017.

[9] S. Kamel, W. Hamdy, S. Abd-elgwad, A. Selim, and F. Jurado "Development of Probabilistic Power Flow Algorithm for Radial Distribution Systems with Capacitors Using Analytical Approach" IEEE 10th International Renewable Energy Congress (IREC 2019), Sousse-Tunisia, March 26-28, 2019.

[10] S. Kamel, M. Mohamed, A. Selim, L. Nasrat, and F. Jurado "Power System Voltage Stability Based on Optimal Size and Location of Shunt Capacitor Using Analytical Technique" IEEE 10th International Renewable Energy Congress (IREC 2019), Sousse-Tunisia, March 26-28, 2019.

[11] A. Selim, S. Kamel, F. Jurado, and S. Marrouchi "Power and Energy Cost Losses Minimization Using Optimal Capacitor Placement in Distribution Networks" IEEE 10th International Renewable Energy Congress (IREC 2019), Sousse-Tunisia, March 26-28, 2019.

[12] S. Kamel, A. Selim, F. Jurado, J. Yu, K. Xie and T. Wu "Capacitor Allocation in Distribution Systems Using Fuzzy Loss Sensitivity Factor with Sine Cosine Algorithm" 2019 IEEE PES Innovative Smart Grid Technologies Asia (ISGT ASIA 2019), Chengdu, China, May 21-24, 2019.

[13] W. Ahmed, A. Selim, S. Kamel, J. Yu, and F. Jurado, "Probabilistic Load Flow Solution Considering Optimal Allocation of SVC in Radial Distribution System" International Journal of Interactive Multimedia and Artificial Intelligence 5.3 (2018): 152-161.

[14] R. Sirjani, A. Mohamed and H. Shareef, "Optimal allocation of shunt Var compensators in power systems using a novel global harmony search algorithm" International Journal of Electrical Power & Energy Systems 43.1 (2012): 562-572.

[15] S. Kamel, A. Ramadan, M. Ebeed, J. Yu, K. Xie and T. Wu "Assessment Integration of Wind-based DG and DSTATCOM in Egyptian Distribution Grid Considering Load Demand Uncertainty" 2019 IEEE PES Innovative Smart Grid Technologies Asia (ISGT ASIA 2019), Chengdu, China, May 21-24, 2019.

[16] M. Abdel-Nasser, K. Mahmoud, and H. Kashef, "A Novel Smart Grid State Estimation Method Based on Neural Networks" International Journal of Interactive Multimedia and Artificial Intelligence 5.1 (2018): 92-100.

[17] A. Amin, S. Kamel, and J. Yu, "Developed Analytical Technique for Optimal Placement and Sizing of DG and DSTATCOM in Radial Distribution Systems" IEEE Twentieth International Middle East Power System conference (MEPCON 2018), Cairo, Egypt, December 18-20, 2018.

[18] M. Ebeed, S. Kamel, and H. Youssef, "Optimal Setting of STATCOM Based on Voltage Stability Improvement and Power Loss Minimization Using Moth-flame Algorithm" IEEE Eighteenth International Middle East Power Systems Conference (MEPCON 2016), Cairo (Egypt), December 27-29, 2016.

[19] M. A. Taher, S. Kamel, F. Jurado, and A. Ahmed "Optimal Allocation of STATCOM for Voltage Stability and System Loadability" IEEE Twentieth International Middle East Power System conference (MEPCON 2018), Cairo, Egypt, December 18-20, 2018.

[20] A. Selim, S. Kamel, and F. Jurado, "Hybrid Optimization Technique for Optimal Allocation of DG and D-STATCOM in Distribution Networks" IEEE Twentieth International Middle East Power System conference (MEPCON 2018), Cairo, Egypt, December 18-20, 2018.

[21] M. Ebeed, S. Kamel, and A. Youssef, "Optimal Integration of D-STATCOM in RDS by a Novel Optimization Technique" IEEE Twentieth International Middle East Power System conference (MEPCON 2018), Cairo, Egypt, December 18-20, 2018.

[22] B. V. Rao, and G. V. Nagesh Kumar, "Sensitivity Analysis based Optimal Location and Tuning of Static VAR Compensator using Firefly Algorithm" Indian Journal of Science and Technology 7.8 (2014): 1201-10.

[23] A. A. Abou El-Ela, R. A. El-Sehiemy, A. Kinawy and M. T. Mouwafi, "Optimal capacitor placement in distribution systems for power loss reduction and voltage profile improvement" IET Generation, Transmission & Distribution 10.5 (2016): 1209-1221.

[24] Y-H. Song and A. Johns, "Flexible ac transmission systems (FACTS)" IEE power and energy series, London: Institution of Electrical Engineers, 1999.

[25] N. G. Hingorani and L. Gyugyi, "Understanding FACTS: concepts and technology of flexible AC transmission systems" IEEE Electrical Insulation Magazine, Wiley-IEEE press, 2000.

[26] S. Kamel, F. Jurado, and J. P. Lopes, "Comparison of various UPFC models for power flow control" Electric power systems research 121 (2015): 243-251.

[27] S. Kamel, F. Jurado, Z. Chen, M. Abdel-Akher, and M. Ebeed, "Developed generalised unified power flow controller model in the Newton–Raphson power-flow analysis using combined mismatches method" IET Generation,

Transmission & Distribution 10.9 (2016): 2177-2184.

[28] A. Foroughi Nematollahi, A. Rahiminejad, and B. Vahidi, "A novel multi-objective optimization algorithm based on Lightning Attachment Procedure Optimization algorithm" Applied Soft Computing 75 (2019): 404-427.

[29] Power systems test case archive. University of Washington, Seattle, Online: http://www.ee.washington.edu/research/pstca/

[30] R. V. Rao, V. J. Savsan and D. P. Vakharia, "Teaching–learning-based optimization: an optimization method for continuous non-linear large scale problems" Information sciences 183.1 (2012): 1-15.

Salah Kamel

Salah Kamel received the international Ph.D. degree from University of Jaen, Spain (Main) and Aalborg University, Denmark (Host) in Jan. 2014. He is an Assistant Professor in the Electrical Engineering Department, Aswan University. Also, He is a Leader for a power systems research group in the Advanced Power Systems Research Laboratory (APSR Lab), Aswan, Egypt. He is currently a Postdoctoral Research Fellow in State Key Laboratory of Power Transmission Equipment and System Security and New Technology, School of Electrical Engineering, Chongqing University, Chongqing, China. His research activities include power system modeling, analysis and simulation, and applications of power electronics to power systems and power quality.

Heba Youssef

Heba youssef received the B.Eng. from the Faculty of Engineering, South Valley University, Egypt in 2011. She is currently pursuing his MSc degree in Department of Electrical Engineering, Aswan Faculty of Engineering, Aswan University. Her research activities include Power system modeling, analysis, and optimization.

Social Seducement: Empowering Social Economy Entrepreneurship. The Training Approach

Natalia Padilla-Zea, Stefania Aceto, Daniel Burgos *

Universidad Internacional de La Rioja (UNIR) – Research Institute for Innovation & Technology in Education (UNIR iTED), Logroño, La Rioja (Spain)

Received 22 January 2019 | Accepted 1 July 2019 | Published 30 October 2019

Abstract

Long-term unemployment is a persistent problem in Europe, following the economic crisis suffered in 2008. This situation reveals self-employment as a good option for becoming re-involved in the working life. In this context, this paper presents a gamified educational platform to empower social economy entrepreneurship skills in long-term unemployed people. In particular, we present the training approach underpinning the motivational process supported by gamification, which has been developed using the ADDIE model. The learning path is developed according to a story that guides the work throughout the training process. It is based on the premises of alignment with reality, instruction from didactic material and real-life stories, in-game practice, work in groups and assistance from a facilitator. This approach covers the competence needs identified in a previous study and includes gamification techniques to improve motivation and engagement. Therefore, the training program comprises: (1) a set of materials and real social economy enterprise experiences, which are the basis for learning and getting inspiration; (2) activities to develop a business plan based on concepts learned from the learning materials and from real-life stories, as well as the help of a facilitator who walks with trainees during the process; (3) a set of individual and group, mandatory and optional assessment activities to evaluate the learning achieved; and (4) a three-views scoring system that shows learning progress for individuals and groups, and gives players the opportunity to exchange gamification points for benefits in the game. The results presented in this paper are based on a sample of two pilots run in Italy and Spain and involving five facilitators working with around 60 learners. About 60% of participants indicated their intention to apply knowledge obtained in a real-life entrepreneurship initiative.

Keywords

Gamification, Pedagogical Model, Social Economy, Social Seducement, Teamwork.

DOI: 10.9781/ijimai.2019.09.001

I. Introduction

ANTHROPOLOGY combines sociology, psychology, biology, and the humanities to carry out deep analysis of human beings and social groups [1]-[4]. Entrepreneurship uses the same combination of sciences to implement specific actions towards a clear and tangible objective [5]. Games are a key part of this analysis in any type of society or cultural representation [6]. Through play, children develop key abilities, from hunting to counting, through solving problems and by learning other fundamental skills that will support them when they become adults. For adults, games are the natural way to approach situations, such as negotiations and meetings, in which everyone plays a role and must find his or her place and a way of communication [7], [8]. Entrepreneurs can use the key features of games to achieve objectives and acquire specific competences [9], [10]. Nowadays, the social unit consists of distributed contacts and cloud services, which facilitate a 24/7 umbilical connection of the individual to the group [11], [12]. With this background, eGames and gamification techniques look like a natural evolution in achieving real entrepreneurship competences in a digital context, one that is based on interaction and social strategies with social networks, learning management systems, entertainment apps and digital tools that support and extend social contacts. In this setting, learning happens anytime, anywhere. This is the approach of digital anthropology [13], in which groups are analysed as a combination of cultures and contexts.

Furthermore, while working with traditional group skills and interaction, there are a number of features on which researchers focus their attention [14]-[16]: for example, social identity, the user's role in a group, social forms of expression, the group's need for assembly and the development of personal and group skills for improving social reputation. We find similar features in digital entrepreneurship, but in a particular representational medium [17], [18]. Enabled by social networks and Internet apps, individuals and groups can share and express themselves (e.g. via Facebook and Instagram); meet for discussion (e.g. through Google Hangouts and Skype); provide opinions, argue, and call for justice (e.g. on Twitter and WhatsApp); create music (e.g. with Garage Band); and work (e.g. using Outlook, Calendar, and OpenOffice). Social roles within the group are also required for interaction (e.g. adopting the role of a follower or an alpha male). In short, there are many similarities between these and other social structures that are usually the subjects of anthropological studies.

In this context, games play a role in helping individuals achieve social status, develop social skills and personal competences, and adopt roles in group interaction [19], [20]. The positive and negative sides of

* Corresponding author.

E-mail addresses: natalia.padilla@unir.net (N. Padilla-Zea), stefania.aceto@unir.net (S. Aceto), daniel.burgos@unir.net (D. Burgos).

games have been discussed by Prensky, Gee, and others [5], [21].

For example, the dark side of games might be used for promoting effective learning and interaction, including the ability to overlap tasks, leading to skills in multitasking; the manipulation of social identification through color, dress codes, and strange jargon; encouraging social skills for empathy and creating internal relations inside a group; and, finally, for cultural integration. Furthermore, functional diversity based on the inclusion of over-gifted individuals, who are barely integrated into the group but possess special features that make them unique, becomes a one-in-a-million feature that makes a group stand out. All those examples present numerous 'bad habits' related to games, which can be turned upside down in order to support effective and useful learning and interaction. The main objective of our research is to use the best features of games to support entrepreneurs' achievement of specific competences through social interaction and gamification. We discover how to match the social behaviour of individuals and groups through an educational digital game design to spur the process of learning personal skills [22], [23]. Furthermore, we examine those skills derived specifically from recurrent patterns, which can be used for learning and the development of personal and social competences [24], [25]. In doing so, we provide a new approach to the use of eGames, not just as a tool for learning, but as a vehicle for social entrepreneurship and personal behavior analysis, moving towards personalised support offered to every user and group, which is based on individual inputs and pattern identification.

The rest of the paper is organized as follows: in section II, we present a set of related works, including some of the social-entrepreneurship game-based proposals that we have studied; section III gives a general overview of the Social Seducement project, stating its main features; section IV outlines the Social PlaNet gamified learning platform developed as part of the Social Seducement project; section V presents the pedagogical model designed for this eGame; section VI describes the assessment performed on the pedagogical model; and section VII summarises our main conclusions and further works.

II. Related Works

The work presented in this paper has been developed as part of the Erasmus+ Social Seducement project. The research consortium studied different proposals included in the Guide for Developing the Game and its Environment [26] and analysed several internal and external experiences, which have also been summarised in this paper.

Games and gamification took a key role in education innovation some years ago, as we can see, for example, in the NMC Horizon reports from 2013 to 2017 [27]-[29], which state the potential of these approaches to improve students' motivation and engagement. Moreover, there has been an evolution of games used in education, which have been paired with learning pedagogies [30]: a first approach based on behaviorism, where a reward was given for every correct actuation; the next one, based on cognitivism and constructivism, where students learned from past experience and from experience in the virtual environment in collaboration with other partners; and the third one, based on constructionism, where simulations are offered to ease the transfer of knowledge.

Many studies have been carried out in this field, obtaining positive results in several areas that support the theory of game-based learning. A review of the most relevant ones is presented in McClarty et al. [31], which seeks to clarify their strengths and weaknesses. In particular, this review concludes that games have the potential to successfully improve the pedagogical approach's playfulness, personalize learning, improve engagement with the process of learning, teach twenty-first-century skills, and provide sound assessment of learning. The authors highlight the specific features that have allowed games to become a new tool for learning: immediate feedback, failing without consequences in the real world, a low level of frustration compared to that in traditional education, learning from mistakes, improvement of motivation and progressive growth of difficulty across levels. Particularly relevant in this context is games' ability to train learners in problem-solving, innovation, and decision-making, important skills in the current society.

However, there has been one lingering concern about the suitability of using games, posed by Uliscak [32]: '[O]ne of the major concerns identified about using digital games in education is the difficulty in assessing effectiveness at achieving the learning goals. How does a player of Civilization, for example, demonstrate that they know the seven wonders of the ancient world rather than merely acquiring stars?' As Jeffery Chin and colleagues state [33]: 'Designing ways to collect data on student learning in simulation and gaming is particularly difficult because of the open-ended nature of these activities.' This issue can be more easily addressed by teachers, who can act as facilitators or mediators of the process of learning. The teacher can lead, stimulate, and make explicit what the student learns.

Malone [34], cited in [35], defines some key elements of the design of an educational game:

- The activity should be structured in such a way that players can increase or decrease the difficulty of challenges faced, in order to match exactly personal skills with the requirements for action;
- It should be easy to isolate the activity, at least at the perceptual level, from other stimuli, external or internal, which might interfere with involvement in it;
- There should be clear criteria for performance; a player should be able to evaluate how well or how poorly (s)he is doing at any time;
- The activity should provide concrete feedback to the player, so that (s)he can tell how well (s)he is meeting the criteria of performance;
- The activity ought to have a broad range of challenges, and possibly several qualitatively different ranges of challenge, so that the player may obtain increasingly complex information about different aspects of her/himself.

An additional element to be considered is that, in our case, the target group is made up of long-term unemployed people, the majority of whom are at risk of social exclusion. The number of proposals about the corporate sector is significantly higher than the number of proposals in formal education, although there is still a lack of evidence about the effectiveness of eGames in social inclusion. This subject has been studied by Stewart et al. [36], who conclude that digital game-based approaches provide adaptable, motivating, and engaging techniques that can be used to empower individuals and enhance social inclusion. More concretely, they state that 'the use of digital games and gaming is starting to show potential in addressing issues of policy concern including wellness and ageing, education and employability of poor learners, improved quality of training, skills development and civic participation'. Moreover, in this study, the role of 'intermediaries and professionals working in the field of social inclusion' is stated as a key enabler of inclusion. We could assume the same role for teachers, trainers, and learning facilitators as mediators in the learning process via educational games.

Following the review of the trends and challenges of using online games for learning, a process of research on relevant practices was developed [26], whose relevance in the scientific community is currently growing [38]. In Table I, we have included several games used in the context of learning and training for vocational jobs, social entrepreneurship skills, and any other feature to improve learners' capabilities to run social enterprises.

From this review, and after a process of analysis on several criteria (game objectives, tools used to achieve the project goals, target groups, methodologies applied or underpinning the learning

model, assessment and artwork), a reduced set of key practices was selected, as shown in Table II. These helped the consortium to get inspiration and to decide some of the features of the gamified training solution to be developed.

Thus, from the analysis of the games and projects presented in this section and according to the goals and target group of Social Seducement, we conclude that the pedagogical approach needs to be linked to the social reality of participants to provide them with a situated and experiental learning experience without losing motivation. In addition, a clear mechanism to assess progress is needed, and immediate feedback needs to be provided. Moreover, supporting disadvantaged learners is a key point in Social Seducement, as doing so promotes civic participation and social skills. For that reason, it was judged necessary to include the role of the facilitator (a sort of game master who will facilitate the learning of game users throughout the

TABLE I Sample of Games to Learn or Train Social Entrepreneurial Skills (Adapted from [26])

Reference	Games and projects scanned
[39]	• **EVE Online**. Players practise running the most powerful company in the world to be able to sharpen their management skills in the real world. • **Informatists** helps players learn and polish business skills. This game helps players practise strategies and learn to succeed without a high risk of failure. • **Gazillionaire Deluxe**. Used in colleges to teach business, math and economics, this game teaches about supply and demand. • **Ports of Call**. This game helps players build their shipping fleets wisely and learn plenty about economics while playing. • **Fistful of Dollars**. Set against a space backdrop, this game teaches players about managing working capital. • **Robo Rush**. Players start out selling their robots door-to-door, then upgrade to a storefront and a factory. • **The EIS Simulation**. Designed at the Center for Advanced Learning Technologies, this game challenges players to become part of a team that will introduce an innovation and convince 22 members of the management team to accept the innovation. • **Innov8**. This simulation game from IBM provides players with a chance to sharpen their business acumen with three different scenarios focusing on smarter traffic, smarter customer service, and smarter supply chains. • **CyberCIEGE**. Designed by the US Navy, this game teaches about network security. Players purchase and configure workstations, servers, operating systems, and more, then try to keep them all secure while also balancing budget and productivity factors. • **A Force More Powerful**. This is a nonviolent strategy game that teaches players how to use nonviolent tactics to disarm potentially violent situations, possible even useful as conflict resolution at online colleges for social work. • **Karma Tycoon**. This is a nonviolent strategy game whose goal is teach about business and community in a nonprofit environment. • **A Tale in the Desert 4**. Set in ancient Egypt, this massively multiplayer online role-playing game has a focus on economic and community development. This game is one of the few MMORPGs that includes no combat at all. • **Democracy 2**. With this game, players can learn to make decisions, such as implementing green policies, raising or lowering taxes, creating harmonious foreign relations, and more, from the politician's perspective. • **Ars Regendi**. The aim of this game is becoming the president of a state and applying political and economic strategies to form alliances with other countries. • **Peacemaker**. This game teaches about the balance of power. If the player chooses the right path, peace is obtained; if the player makes wrong choices, a violent disaster will occur.
[40]	• **The New Heroes** tells the dramatic stories of 14 daring people from all corners of the globe who, against all odds, successfully alleviate poverty and illness, combat unemployment and violence, and bring education, light, opportunity, and freedom to poor and marginalised people around the world.
[32]	• **SimVenture** is a business simulation game designed for people 14–30 years old who are learning about the realities of setting up a business. • **Climate challenge game** is an online sandbox-style strategy game developed by Red Redemption Ltd. Its aim is engaging 20- to 35-year-old professionals in the realities of climate change and measures that can be taken to decrease carbon dioxide emissions.
[41]	• **White Card Game.** This game is intended to promote safety practices in carpentry workshops.
[36]	• **In-living**. To help young people become good tenants. • **Rock'n'high roller.** For financial planning, addressing people aged 18 to 24. • **Footfall**. To help young people learn financial responsibility for small businesses. • **TARDIS project**. To improve interview skills of young people with low employability. • **No credit, Game over**. To train young people in debt management. • **Thuis in the Netherlands**. Supporting migrants in the preparation of the naturalisation exam. • **Mixopolis**. Vocational orientation and participation for young migrants.
[42]	• **Hot Shot Businesses.** A serious Flash game in which users have to decide how to find capital, rent tools and material, decide about advertising, etc. in order to be introduced to the rules of entrepreneurship. • **Industry player.** A massive online adventure game in which players fight to become the leader of the market. They can choose between around 250 types of enterprise to build a five-company holding, by developing which they train their business skills. • **INNOV8.** This game has a 3D environment where players get familiar with the IBP Business Process (BPM). By playing this game, players can work together and try different scenarios in order to discover how this business model can improve the world. • **SimVenture.** A business simulation game about computer sales and repair in which players work in monthly matches. Over the course of one month, they can see real-life examples as well as a tutorial to guide their decisions. After each month, they receive feedback about results and how their decisions affect the business. • **Small Business.** Like SimVenture, but more focused on game-related aspects. Players have to choose between a sandwich shop, a café, a music store, and an ice-cream store in a time-lapse simulation taking several days, which also includes such real-life factors as economic conditions, changes in the market, and time dedicated to the family. • **Market-place.** More intended for professional training, this game is divided into several modules that highlight specific aspects of a business, including links to texts that can help in running an enterprise. This game is more challenging than the ones previously mentioned. • **MetaVals.** This game is focused on lifelong learners in business schools and presents a quiz manager to check how much players have learned in different areas. It includes collaboration and competition. • **Team Up.** This game is played in groups and is intended to train them in the social aspects of entrepreneurship. The scenario is a ruined island they have to leave by solving a set of puzzles. • **The balance sheet.** This very practical game is intended to train players in the balance sheet, using real-life examples. • **The Enterprise Game.** This game allows the player to learn about cash flow, market needs, profits, etc. while playing. • **Virtonomics.** In this game, players have to create and manage a company with different subdivisions. Players receive feedback during the game.

Reference	Games and projects scanned
[43]	• **Go Venture Any Business.** Creates a business simulation for ANY product, ANY industry, and ANY market.
[44]	• **Kenteq Craft.** This game presents players with challenges to learn to use the machines that they will use afterwards in their jobs, as well as promoting motivation and social learning.
[45]	• **The Start Up Game.** In this simulation, students play the role of different stakeholders in a market full of early-stage startups. They gain an understanding of decision-making under uncertainty, the variations in strategy among individuals, and experience with common trade-offs resulting from different strategies and decisions.
[46]	• **Safety Training Cross Cut Sew.** This goal of this game is reducing accidents in furniture manufacturing during regulated instruction.
[47]	• **Everest V2**. Team-based simulation using the dramatic context of a Mount Everest climb to teach group dynamics and leadership to novice entrepreneurs.
[48]	• **Entre Explorer – Serious Game for Immersive Entrepreneurs.** This game provides players with contents which help them in acquiring knowledge. The game takes place in a building that players have to go through. On each of the floors, players learn something about enterprise to help them build their business plans incrementally.
Other proposals found by the consortium	• **YENTELS.** Young European Entrepreneurs E-Learning Suite. The main objective of the project is to encourage entrepreneurship in young people to improve employability, competitiveness, and innovation through a simulation/computer gaming–based e-learning suite of training materials that will assist them in acquiring the entrepreneurial skills necessary to set up and run their own businesses (http://yentels.virtech.bg/staticsections/view/game). • **SPENT.** An online game about surviving poverty and homelessness, whose mission 'is to provide food, clothing, shelter and supportive services to neighbors in need'. • **Games for change.** Founded in 2004, Games for Change facilitates the creation and distribution of social impact games that serve as critical tools in humanitarian and educational efforts. • **Business Simulations for Executive Education and Business Learning.** A simulation putting the students in front of situations related to the topics they need to train in to enhance their decision-making process (http://www.industrymasters.com/). • **Green community.** An establishment in the province of Sassari (Italy) of a technological instrument (available via web and mobile) that supports knowledge of the 'green economy' among young people between 14 and 20 years old.

TABLE II. Selected Practices for Social Seducement (Extracted from [26])

Name of the game/project	Reasons for relevance
KARMA TYCOON	As in the case of Social Seducement (inspired by the ILO board game Co-opolis), Karma Tycoon is the online conversion of a board game focused on social entrepreneurship.
THE NEW HEROES	The game deals with social entrepreneurship in third-world countries and is interesting for its effective use of real stories for inspiration and learning. Storytelling and transposition of real-life experiences could be used in the Social Seducement game.
EVEREST V2	The game is a team-based simulation about climbing Mount Everest for team-building and leadership training. It is interesting and relevant, as it uses simulation and F2F virtual discussion among a group. It also uses incremental (learning and playing) steps based on decisions taken jointly and leading to success (or not). This strategy could be adopted by Social Seducement for collaborative, incremental learning through the game.
THE START UP GAME	The game is relevant for the topic, the group discussion, and the cooperation required; the clear set of roles foreseen; and the short duration.
SPENT	An online game about surviving poverty, this is relevant because of the possibility of asking friends for support and for the advice messages when wrong turns are taken.
SIMVENTURE	This simulation of a business start-up is relevant because of the availability of individual and group play, the inclusion of a virtual tutor and for its presentation of videos. The game includes ready-made scenarios and a save and load function. It can be used to run active case studies in which people face the consequences of the business decisions they make.
GO VENTURE ANY BUSINESS.	This game is relevant for the presence of the instructor (some sort of facilitation will be needed in Social Seducement), the possibility for team evaluation, and its competency quizzes, which are personalised for each student. It has an interesting infographic, it is possible to create new case-studies and for individuals and teams to play at the same time, synchronous and asynchronous messaging are available, and there are coaching opportunities.
THE GREEN COMMUNITY	This game is relevant because of the presence of virtual facilitators, the possibility of monitoring players' progress, and the different elements built in to remind participants of the 'game character' of the initiative and to let them have fun in learning.

game, supporting teamwork, animating the groups of learners, and supporting project partners in the assessment of the game experience) in order for learners to be accompanied in their learning journey through the gameplay, and therefore in the acquisition of skills.

III. Social Seducement Project: An Overview

The Social Seducement project aimed to foster self-empowerment and entrepreneurship attitudes within disadvantaged groups in Europe, with a focus on unemployed adults aged 18+. It did so by engaging such groups in a gamified training solution, of type Online Role-Playing Game (ORPG), that introduced them to a world of social collective cooperatives and guided them through the principles, concepts, and practices necessary to be able to set up and successfully manage a social enterprise, with specific reference to social cooperatives.

The project was developed thanks to the support of the European

Commission, in the frame of the Erasmus Plus Program (Strategic Partnerships in Adult Education). It gathered a group of partners located across the UK (Tavistock Institute of Human Relations, coordinator), Spain (Universidad Internacional de la Rioja), Italy (Ecobyte and Le Mat), Belgium (REVES – European Network of Cities and Regions in the social economy), and Sweden (Coompanion). The partners all had different backgrounds and expertise, which was a strategic design choice made to guarantee the availability of all the necessary competences an innovative project like this one needs to be successful. In particular, it leaned on the experience of the coordinator in the field of social sciences, which was necessary to design the methodological approach taken in the project; on the technical experience of UNIR in the field of game architecture design; and on the experience and networking capacity of social partners (Le Mat, Coompanion, and REVES) in adapting the game to the needs of the target group.

The project included the following main phases:

- Target needs analysis and game design: Based on desk and field research, the project mapped similar experiences in the field (i.e. online games about social economy), identified the needs of the target group (unemployed) in terms of training to become social entrepreneurs and elaborated a possible architecture for the game.
- Game development: Based on the outcomes of the previous phase, the project team developed the online game and gathered the necessary training material to support learners in their learning process while playing the game.
- Facilitator training: The facilitator was introduced as a sort of game master facilitating the learning of users throughout the game, supporting teamwork, animating the groups of learners and supporting project partners in the assessment of the game experience.
- Piloting: Five pilot countries were involved (UK, Italy, Spain, Sweden, and Belgium). In each country, the game was tested with 36 learners (divided into small groups), who played the game with the support of a facilitator. There were three facilitators per country, with each facilitator supporting at least two groups of learners.
- Validation: Based on the pilot results, the game was finalized and made available to the public (with special reference to employment and training centres, and all those entities supporting employment, employability, and social inclusion).

Social Seducement was an innovative project in that it developed an ORPG in the social economy field. The mapping carried out in the first phase of the project revealed a substantial lack of similar experiences in the field. Addressing disadvantaged groups (the unemployed, who are in a fragile situation in terms of self-esteem and self-confidence most of the time) and training them in the creation and management of social collective cooperatives (involving teamwork, respect for others, collective decisions, and collective responsibilities) poses a lot of challenges in the design of an online game, which has its own technical rules and requirements.

IV. Social Seducement Project: The Gamified Learning Experience

As result of the work in the Social Seducement project, the gamified learning platform Social PlaNet (http://ss.epscms.com) was developed. In this learning experience, groups of students learn about the Business Model Canvas (BMC) to be applied to a social economy enterprise idea, which will be developed while they are working on the platform.

The story underpinning the learning experience is about a group of neighbours who meet in a cafe in a European city and are recruited by a magical person to run a social economy enterprise (Fig. 1) [50].

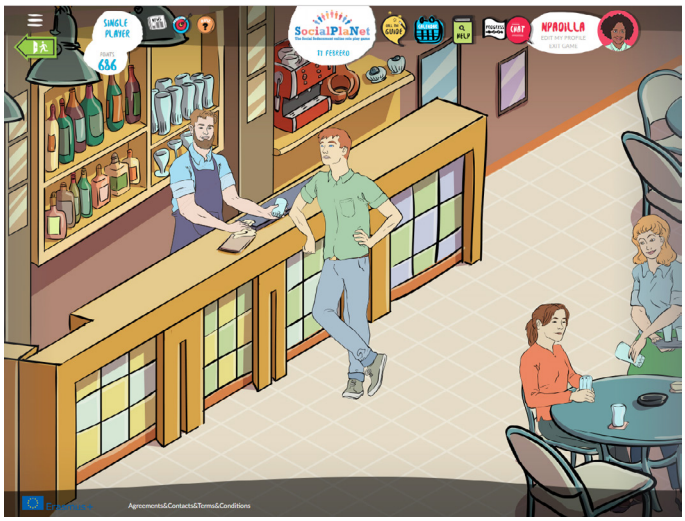


Fig. 1. Cafe in the Social PlaNet platform.

The group of potential entrepreneurs has to start the process of developing the idea for their enterprise. The first step is looking for an office where they can work and going to the library to get some information about social entrepreneurship. The BMC arises as an ideal tool to find and study the different elements involved in the development of the enterprise idea, and learners need to think about each of them: key partners, key activities, key resources, value proposition, customer relationships, channels, customer segments, cost structure and revenue streams. These elements are structured in three chunks, as shown in Fig. 2, which will be accessible from the office as well as from the library (Fig. 3).

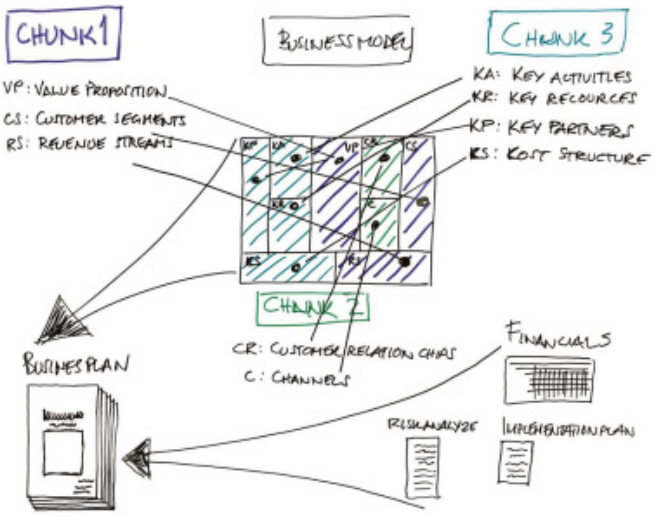


Fig. 2. BMC chunks assignments in Social PlaNet [extracted from 50].

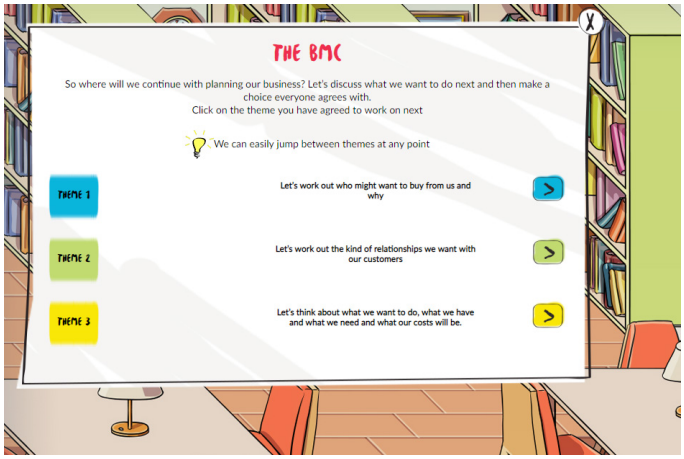


Fig. 3. BMC access.

The learning experience implemented in Social PlaNet has been divided into eight levels (Table IV), from user registration to building an optional business plan. The gamified strategy encourages learners to progress. The main idea of this gamification approach is to reward group activities more than individual ones and use points obtained to allow learners to furnish their office. That way, each piece of furniture has a cost in points, and the group can 'buy' each of them when they obtain the necessary number of points.

During this learning process, students have to develop the BMC in groups. To set up the group session, they have to arrange a meeting and connect to the platform at the date and hour previously agreed in the platform schedule. However, a single user can connect to the platform in an individual session for a deeper study or a wider read of the materials. A learner can start an individual session whenever a group session is not fixed (see Fig. 1, upper left corner: 'single player'), but (s)he will be not allowed to further develop the BMC.

A role-based approach is included in order to develop students' soft skills. A group member is randomly selected to be the task leader in each of the group sessions. The task leader is responsible for organizing the group discussion, extracting the main ideas and promote agreement between group members in order to register the agreement in each phase of the process. This role can be rejected and the former learner can select another partner to play that role. Only the task leader is allowed to fill in the blanks with the information agreed to by the group. For example, in Fig. 4, the group has agreed how to transport its products.

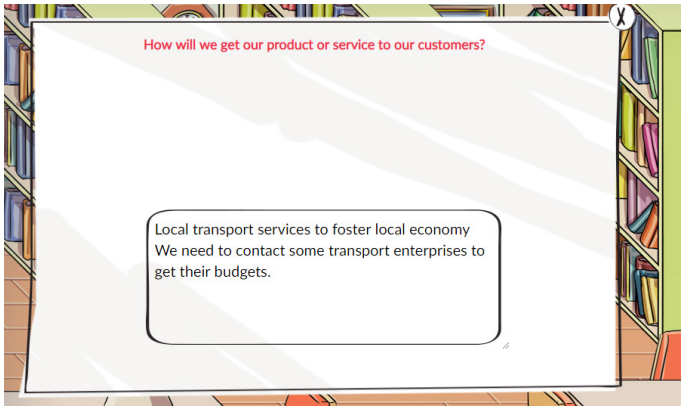


Fig. 4. Group agreement about product transportation.

During this training course, a facilitator helps the group to advance in its BMC development. This facilitator reviews progress and is available to solve any doubts arising in the group. To contact him or her, both individual and group sessions have an icon participants can click to chat with the facilitator (CALL THE GUIDE), in addition to the one they have for internal discussions (CHAT) (Fig. 5).

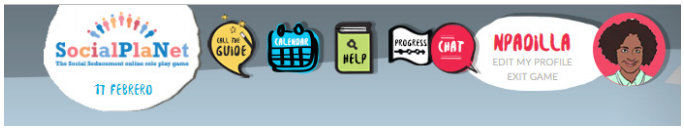


Fig. 5. Icons in Social PlaNet: call the guide, calendar, help, progress and chat.

Moreover, a selection of real social economy enterprise stories is available on the platform to foster brainstorming in each of the levels. As the group goes forward, the stories are unlocked and students can read about the specific item on which they are working. These stories are based on real cases the project partners collected by contacting national social economy enterprises.

To succeed in this learning experience, the group has to finish the BMC for the idea initially posed. At the of the process, when they have filled in the BMC, they can spend the remaining points in their office or continue developing a final business plan, which will be printed as a pdf file with all the information developed by the group.

V. Social Seducement Project: The Training Approach

As previously stated, Social Seducement is a gamified learning platform intended to train long-term unemployed people in social entrepreneurship skills. By participating in the pilot, groups of learners came up with their business ideas using the Business Model Canvas (BMC). In this section, we focus on the pedagogical approach underpinning the learning process.

A. Foundations

The pedagogical model introduced in Social Seducement has been developed on the basis of three main pillars [51]: (1) the ADDIE instructional design model; (2) the European Qualification Framework, to split the assessment model into knowledge (to know), skills (to be able to) and attitudes (to behave with); and (3) a comprehensive competences framework for social economy entrepreneurs.

ADDIE (Analysis, Design, Development, Implementation, and Evaluation) [52] is a successful and widely used instructional design model comprised of five phases. One of the most important characteristics of this model is its ability to be applied both iteratively and sequentially, which offers the instructional designer some freedom in developing the instruction. Next, we present a brief explanation of each of its phases, including how they have been implemented in our project:

1. **A**nalysis, in which the scenario has to be defined: students and their characteristics, context of use, available resources, etc. In Social Seducement, this phase was performed by delivering a survey to a wide range of key actors in social economy entrepreneurship in order to find out about the needs associated with the creation of a cooperative social enterprise. Additionally, conclusions from this study were compared to a literature review, which allowed for adjustment of the conclusions obtained.
2. **D**esign: Tasks to be performed in this phase are related to learning itself. Thus, the teacher or instructor has to describe the goals, determine how they are going to be assessed, choose the way to provide students with materials, define the general didactic model, plan instruction by deciding the elements and order of the content, design activities for students and identify resources they need. In Social Seducement, this phase has been developed by structuring contents into levels in the game and by managing progress through the storytelling that guides the game.

3. **D**evelopment: This phase is intended to generate the educational content designed in the previous phases. In Social Seducement, it was developed by proposing a set of materials, which were divided into external and internal materials. The external material is a set of links to resources (texts, videos, wikis, etc.) on the web, including concepts about social economy enterprises and how to manage every step in their development and operation; the internal material is a set of real-life stories intended to give inspiration to learners while they are playing. Partners of the consortium gathered these stories from enterprises in every country.
4. **I**mplementation, in which the instruction takes place: In this phase, the instructor has to promote students' understanding of the content, support their learning and follow up on their progress. In Social Seducement, this fourth activity was developed in a pilot process in five countries: Italy, Spain, Belgium, Sweden and the United Kingdom. Additionally, before testing the game, we tested the instructional model itself, in order to avoid pitfalls once that model is included in the platform.
5. **E**valuation: Given the possibility of applying this model both iteratively and sequentially, evaluation could be formative or summative. The complete game Social Seducement will be tested and results in this phase will be used to refine and validate it. Results of the pedagogical model test will also be evaluated to improve the final version to be included in the game.

The pedagogical model of Social Seducement is intended to train people in competences identified in the European Qualification Framework, which separates them into knowledge, skills, and attitudes [53]. For that reason, our assessment model is also organized into these three parts, each of them involving specific activities. However, since competences are difficult to assess, it was decided to evaluate learners' achievements based on learning results. A complete list of learning results for each competence was developed [54], without detriment to a learning result that could promote the acquisition of more than one competence at the same time.

Finally, a comprehensive portfolio of competences was developed based on a revision of the relevant literature on social economy enterprises and a survey of local actors in the different of consortium members' countries [54]. This portfolio, which falls outside the scope of this paper, covers a complete set of competences needed for prospective and struggling social economy entrepreneurs, grouped into (1) organization of social economy enterprises; (2) knowledge of the environment of the social economy; and (3) individual competences that smooth the running of a social enterprise.

B. Learning Structure

Using the competence portfolio, consortium members developed a high-level view of the learning results that social economy entrepreneurs are intended to achieve. However, in order to be able to assess what they are actually learning during the training process, the attention was focused on learning results, which can be analysed at a lower level. A set of competences that social economy entrepreneurs need to run their social enterprises was obtained from the prior knowledge of partners in the consortium, the literature review and the survey findings. That way, the complete and complex portfolio was easier to manage and understand. Below, we present the set of competences packages obtained by partners of Social Seducement from that process [50]:

- Understanding the concept of social economy. Social economy enterprises have several features that differentiate them from traditional enterprises, such as governance type, use of benefits, and rules and regulations related to tax. Therefore, learners need to learn these principles to run the idea properly.
- Conception of the idea by a group of persons. Since the game is played in groups, learners need to develop their skills to work collaboratively to conceive and develop their business idea. In this package, players will train empathy, listening capacity, mediation capabilities, etc.
- Establishing a business plan. To achieve a proper business plan, several competences need to be trained, such as research and analytical capabilities, development of a market study, and using financial instruments. Additionally, editorial skills are needed to write the business plan and present the idea and proposals to stakeholders. The capacity to learn from previous experiences to adapt the business plan to changes in the local environment is also necessary.
- Management, including communication and marketing. In this package, skills related to bookkeeping, the legal environment, administrative requirements, etc. are trained for. Learners will also need to be able to organize meetings, plan, make decisions, use IT instruments, etc.
- Creation of a conductive stakeholder environment. Knowledge of and commitment to the local area is essential for social economy enterprises, as is direct contact with local authorities, other social economy or traditional enterprises, universities, citizens, etc. These relationships allow entrepreneurs to identify local needs, define and develop proper activities to solve these needs, and adapt to local evolution.
- Evaluation and impact assessment. Social economy enterprises also need to evaluate their activities and revise strategies on the basis of the results obtained or to attract new investors according to the impact achieved.
- Operations. Beyond development of the idea or planning strategies to be developed in the local environment, a social economy enterprise needs to be managed daily. For example, deliveries to customers, meetings with local actors, and negotiations with banks need to be maintained and studied in order to properly assess the implications of these actions and how they will (or won't) help the business.

Until now, we have three elements to be combined in the training process: (1) the competences portfolio; (2) learning results; and (3) competences packages. In order to make explicit the relationships among them, we will use the term 'learning outcome' to refer to learning results related to several competences in the competences portfolio. Thus, in Table III, we show how they are related [50].

C. Learning Materials

In Social PlaNet, learners are provided with two kinds of material [50]: the first type, *external material*, is a set of documents, videos, references, etc. where they can find conceptual knowledge about social economy, enterprises, social economy enterprises, etc. The second type, *internal material*, is a set of stories of real social economy enterprises (see Appendix I for an example) running in the different countries of the consortium, which are intended to give inspiration to learners in the process of developing their ideas to build their own social economy enterprises.

In the first group, contents have been divided according to learning outcomes, and subsections have been formed to encourage learners to focus more precisely on what the reference is about. Throughout the training process, these concepts will be shown to learners when they need to know them to face or be prepared for the activities in every level. Each country has a different set of links available because legislation concerning social economy enterprises varies nationally. External learning material for Spain is included in Appendix II.

On the other hand, we have internal materials. A set of 20 stories are included on the Social PlaNet platform. These have been split in several

TABLE III. Relationships between Learning Outcomes and Competences in the Portfolio (Extracted from [50])

Nº	Learning outcomes	Connection with the competence portfolio for social economy entrepreneurs
1	Understanding the concept of the social economy	Organization of social economy enterprises
2	Conception of the idea by a group of persons	Individual competences to smooth the running of a social enterprise
3	Establishing a business plan	Organization of social economy enterprises Knowledge of the environment of the social economy
4	Management (including communication and marketing)	Organization of social economy enterprises Individual competences to smooth the running of a social enterprise
5	Creation of a conductive stakeholder environment	Knowledge of the environment of the social economy
6	Evaluation and impact assessment	Organization of social economy enterprises Knowledge of the environment of the social economy Individual competences to smooth the running of a social enterprise
7	Operations	Organization of social economy enterprises Knowledge of the environment of the social economy Individual competences to smooth the running of a social enterprise

parts according to the contents that learners need in every scene in the game. Thus, each part begins with an introduction of the enterprise, in order to contextualise the story. If learners have already read any other part, they can skip this introduction and focus on the specific part that they need. After that, according to the scene in the platform where learners are, they can read about how to manage concrete situations in the starting up or running of one of these enterprises, some questions for them to think about and the solution that enterprise gave, as well as its implications.

D. Learning Assessment

In order to continue advancing in the platform, learners have to pass several quizzes based on the content they have previously covered. If learners obtain a score of more than 60% on quizzes, they will continue to the next level. If not, they will be encouraged to review some specific internal and external material to learn more about these concepts. Materials that will be presented will be the ones they have not yet read or, if all material has already been read, the ones most closely related to the main concepts in that level.

The first step to properly situate these quizzes is to allocate content and quizzes along the levels (Table IV). To present this distribution, we show where the learning outcomes are trained for in the game and mark the assessment points, including a description of what we assess in each of them [50].

The program-level learning outcome assessment works on the three defined levels of knowledge, skills and attitudes. In particular:

- Knowledge assessment performed throughout the game is structured around several kinds of quizzes, which are to be solved individually or in groups, in order to promote individual commitment, learning by doing and collaborative learning. The correctness of the solutions is checked automatically.
- Skills assessment (optionally performed at the individual level) will occur (1) through self-evaluation of learners at the beginning and at the end of the game (learners will complete a questionnaire at the beginning and at the end of the game to be aware of their self-perception and progress thanks to the game); (2) through automatic tests while the game is played; and (3) through qualitative evaluation by the facilitator.
- Attitude (behaviour) assessment is mainly provided by the facilitator (both at the individual and at the group level), being promoted by a gamified approach intended to improve motivation [56].

1. Assessment Milestones

In order to avoid overloading learners, we have identified some moments in the game where assessment plays a key role to verify the acquisition of knowledge, skills, and behaviors. These key moments are:

- Level 3 (Entering headquarters, defining roles), where a first assessment of skills happens.
- Level 5 (Look at your group: do you have basic features to realise the idea?), where, having presented some examples of social economy enterprises, we assess if learners understand what a social economy enterprise is and is not.
- Level 7.1.1 (Value proposition), where we assess the learners' knowledge of the meaning of value proposition in social entrepreneurial thinking.
- Level 7.1.2 (Customer segments), where, once customer segments can be managed, we assess the capacity to define the different customer segments.
- Level 7.1.3 (Revenue stream), where we determine the capacity to distinguish how important the different customer segments are in relation to the whole value proposition of the enterprise.
- Level 7.2.1 (Customer relationship), where, once information about customer relationships is presented, we assess whether learners know how to find/select different customers and the best customers for their social enterprise and products.
- Level 7.2.2 (Channels to distribute products), where we check knowledge about how products reach customers, how difficult this process is and how many resources are needed.
- Level 7.3.1 (Key activities), where the capacity to describe the whole production process of the social enterprise in terms of activities to be performed is evaluated.
- Level 7.3.2 (Key resources), where we check whether learners have acquired knowledge about all the resources necessary to be able to realise the productive process.
- Level 7.3.3 (Key partners), where the goal is the knowledge acquired about the partners that the group would need to develop its business and social goals.
- Level 7.4 (Costs), where we evaluate if learners are able to recognize costs and their features (fixed, variables, initial capital).
- Level 8 (business plan), which has the aim of checking teamwork, social competencies and planning competencies.
- Final scene, when, having played the whole game and acquired the necessary knowledge, skills and attitudes, learners have to present their finalised business plan.

2. Type of Quizzes

In line with the overall objective to promote motivation and foster engagement, quizzes include a variety of exercises:

TABLE IV. GAME LEVELS AND SUB-LEVELS [50]

Game levels	Game sub-levels		
Level 0 – Registration			
Level 1 – Participant profiles			
Level 2 – Fixing the date for the first meeting			
Level 3 – Entering headquarters, defining roles			
Level 4 – Building an idea			
Level 5 – Look at your group: do you have basic features to realise the idea?			
Level 6 – Make your headquarters a place you feel good about			
Level 7 – Develop your business plan	7.1 – Goal	**7.1.1** – Value proposition segments	
		7.1.2 – Customer	
		7.1.3 – Revenue stream	**7.1.3.1** – Definition of price **7.1.3.2** – Mode of payment **7.1.3.3** – Segmentation of revenues **7.1.3.4** – Initial capital/funding resources
	7.2 – Customer relationships	**7.2.1** – Kinds of customer relationships	**7.2.1.1** – Showing our services **7.2.1.2** – Selling our services **7.2.1.3** – Explaining social impact **7.2.1.4** – Loyalty policies
		7.2.2 – Channels	**7.2.2.1** – Delivering products or services **7.2.2.2** – Kinds of distribution
	7.3 – Activities and resources	**7.3.1** – Key activities	**7.3.1.1** – Activities to be undertaken **7.3.1.2** – Making the product or service available **7.3.1.3** – Widening social impact **7.3.1.4** – Complementary activities
		7.3.2 – Key resources	**7.3.2.1** – Resources for goal, customers, and activities **7.3.2.2** – Resources for social impact **7.3.2.3** – Resources inside and outside our social enterprise **7.3.2.1** – Finding resources
		7.3.3 – Key partners	**7.3.3.1** – Partners to achieve commercial goals **7.3.3.2** – Partners to achieve social impact **7.3.3.3** – Appearance of your social enterprise **7.3.3.2** – Relationship between partners
	7.4 – Costs		
Level 8 – Business plan	**8.1** – Budgets **8.2** – Risk analysis **8.3** – Implementation plan		

- Multiple choice: Given a question, definition, or term, choose the correct answer from the list offered.
- Matching concepts: Given a list of concepts and a list of definition or applications, match each element in the first list to the correct one in the second list.
- Given a set of features, provide a specific subset that fits with a given concept.
- Complete the sentence to compose definitions or describe how to run a task in practice in relation to items/stories presented as training material.
- Build the sentence: Given a set of words, build a sentence with them according to the request.
- Order concepts or features according to a set of defined criteria.
- Puzzles: in order to motivate effort from learners, fragmented images are linked to concepts and definitions. If concepts and definitions are linked correctly, an image appears.
- Divide the set of answers among group members: As a strategy of shared learning, every group member has to learn about concepts presented to contribute to the final common solution.
- True/false questions.
- Classification: Given a list of elements, choose the correct group in which they fit.
- Against the clock: Quizzes where the score obtained depends on how quickly the correct answer is provided.

Learners are allowed to revisit their decisions and answers as a group: given the opportunity to check previous work, the group can refine its idea and its evolution, as well as realise the mistakes made, discover areas to be improved, etc. The possibility of going back in the work is intended to promote dialogue and knowledge-sharing among learners in the group, which is a beneficial way of learning collaboratively.

In addition, groups could also be provided with pre-defined solutions to the challenge that they are facing. Such solutions could come from pre-solved situations included in the game or from other groups' solutions. With this approach, we allow learners to see other ways to use concepts that they learn from training material, recognise good ideas to be included in their social enterprise and detect bad decisions that jeopardise the success of their entrepreneurial adventure.

3. Skills Assessment

Skills assessment happens by means of direct and indirect attitude feedback and by assigning scores to pre-determined positive behaviours.

In the case of implementation with feedback, the facilitator or another group will send comments; in the case of non-feedback implementation, only internal discussion is performed. For every activity in the BMC that coincides with a level in the platform, this assessment would consist of taking pieces of the work from other groups (anonymised) or pre-defined problem/solution sets and giving this group the opportunity to learn by (1) highlighting good decisions, ideas, resources, etc.; (2) incorporating elements into their own project; and/or (3) proposing improvements for the other group (this feedback could be sent to the other group or not, depending on the option selected). This assessment can be included in every learning outcome without repeating the same activity or level.

The gamification approach is used in the attitude evaluation, and it involves gaining points for the following positive behaviours:

- Accepting the role of leader: 50 individual points and 50 points for the group if the first candidate for being task leader accepts.
- Participating in the booking of the next meeting: 10 individual points + 100 group points if everyone participates.
- Assisting with a meeting: 10 points + 100 group points if everyone assists.
- Once a piece of internal material is read, 10 individual points for every participant who contributes in the chat + 100 group points if everyone participates.
- The top contributor in the chat during the discussions in every scene obtains 10 extra points, and the group obtains 100 extra points if all team members participate in the discussion.
- For optional individual activities, 25 extra points will be given, + 50 extra group points if every group member performs the activity.
- For compulsory group activities, 10 individual points will be given to group members who participate in the activity, + 50 group points if all group members participate.
- For optional group activities, 10 individual points will be given to every group member who participates in the activity, + 100 group points if all group members participate.

Additionally, we have two possibilities for implementation: with or without feedback.

4. Assessing Learning Outcomes

In the following, we describe the assessment included in the proposal for knowledge, skills and attitudes related to every learning outcome.

For learning outcome (LO) 1, *understanding the concept of social* economy, we include the following assessment activities, which will be presented at the proper level throughout the game, according to the content taught at each moment. This first set of activities is related to knowledge (Table V).

To assess skills, we proposed the next optional individual activity: Given the work of another group (anonymous) or pre-defined problems and solutions, ask for one of the following, maybe repeated at several

TABLE V. Learning Outcome 1 Assessment

LO1 - Understanding the concept of social economy			
Compulsory		**Optional**	
Individual	*Group*	*Individual*	*Group*
• Multiple – answer questions to select the features of SEE. • Select a number of values applicable to the particular sector in which the enterprise is allocated. • Select the payment method chosen for their idea. • One question to select features of every customer segment.	• One question to match every possible form of governance with its main features. The group has to discuss and to send a common answer. The test will be passed when every governance form is correctly assigned to its definition and features. If more than three tries are incorrect, group members will have to review additional documentation to better understand the concepts. • From a list, select the two values that make their enterprise different (related to the idea, sector, etc.). This question would need to extract responses from previous work. • Present two questions like this: For a SEE in the X sector key resources for commercial aims could be: (select two from a list of five). • "Complete the sentences" activity to link resources to every customer segment. • Link five resources to specific social needs. More than one link could be possible.	• Present the ideas of enterprises from different sectors to decide if they are SEE or not. This is an alternative to the first compulsory individual assessment. Asking for help from other group members could be allowed. A second round for wrong answers is allowed. • Gamification points: 10 points for revisiting or modifying the idea (needs facilitator validation). • Select three local needs that will be met by the SEE. This question needs a pre-established list of needs in every local context, from which the learner can select the most important ones.	• Discussing wrong answers from everyone and giving a common answer. If 60% is obtained, the test is individually passed. Learners not passing in the first-round pass with the obtained score.

points: (1) choosing the best points; (2) presenting improvements; and (3) checking whether lessons learned during the process were incorporated in the proposal.

As previously explained, we include the assessment for the second learning outcome (LO 2, conception of the idea by a set of persons) in the knowledge category (Table VI).

The following activities are used to assess learning outcome 3 (*establishing a business plan*), in the knowledge category (Table VII).

In learning outcome 4, in which competences related to management, including communication and marketing, are taught, the design of the assessment for knowledge is as shown in Table VIII.

In this package, there is an initial skills self-evaluation, which is compulsory. Here, different aspects will be analysed, such as respecting opinions, participating in debates or discussions, confidence, detecting and solving errors, sharing responsibilities, technological skills, game usage and e-learning. Next, given the set of roles arisen from the test of group members, the learner has to match each of these roles with activities that fit with them.

In addition, an optional task can be performed in order to check roles in the SEE that will be developed. Thus, from the roles selected in the compulsory knowledge assessment group activity, the learner would have to select the one that best fits each of the group members.

The next learning outcome, number 5, is about the creation of a conductive stakeholder environment. The assessment of knowledge related to this outcome is designed as shown in Table IX.

For skills, we here propose some optional individual activities:

- Given a set of characters, classify them into the correct segment of a given set.
- Given a set of products to be delivered to specific customers (considering segments) in specific locations, decide the best way to deliver them.

The sixth learning outcome is the one about evaluation and impact assessment, in which the SEE revises results obtained and impact achieved and checks if new investors need to be attracted. We have included the assessment activities for this LO in Table X.

The last learning outcome is related to operations. For this LO, we propose the set of assessment activities about knowledge included in Table XI.

For skills evaluation, we propose an optional group activity: The group is given a budget for some issue, but a problem arises and some money is needed to solve it. The group has to decide how to face this problem and where to find the needed money in its budget.

TABLE VI. Learning Outcome 2 Assessment

LO2 - Conception of the idea by a set of persons			
Compulsory		**Optional**	
Individual	*Group*	*Individual*	*Group*
Given a list of activities, select the ones that are more related to their idea (value proposition). Once every group member has a proposal, share opinions and give a final solution for each particular idea.	Given a list of n possible ideas, order them in descending order of social commitment. Several ideas could have the same level of social commitment. Thus, different partial orders would be correct. The group has to order the list correctly in three attempts. Otherwise, the application will explain why it is not correct.	The game will select and present again the individual and group activities with the lowest scores in order to encourage revisiting material and obtain better performance.	

TABLE VII. Learning Outcome 3 Assessment

LO3 - Establishing a business plan			
Compulsory		**Optional**	
Individual	*Group*	*Individual*	*Group*
• Three of the concepts in LO3 are randomly selected and a puzzle is presented for them each of them in such a way that, once the solution is correctly built, an image is formed. • Simulating staff recruitment, a set of questions is presented and different characters in the game answer the questions. The learner has to select the one who says the correct answers to the questions posted. • Against the clock: every group member has to answer a set of questions, with scoring rated depending on how quickly the group finishes. • Given a list of activities in the enterprise and a list of costs, match them. • If the group needs initial capital, present a 'complete the sentence' activity in which members have to select a word or phrase to complete a sentence about different ways to find start-up capital.		After against the clock: Correct and incorrect answers are discussed in the group. After that, each group member is provided with the answers from a partner in order to check their correctness. If a group member corrects the partner's answers, his scoring increases 10%.	Simulating staff recruitment: Once each of the group members has chosen a candidate for each question, they have to discuss the set of candidates selected and choose the most appropriate one.

TABLE VIII. Learning Outcome 4 Assessment

LO4 - management			
Compulsory		**Optional**	
Individual	*Group*	*Individual*	*Group*
• Given a set of customers, each of them with different features, choose the best way to distribute goods. • - Match a list of possibilities for advertising to different kind of customers/customer segments in terms of effectiveness. More than one selection could be allowed.	Given a description of a SEE and its members, select the roles they fit and describe the ones still needed for success; from a list, select the features that they would need.	Given an example of a SEE and a set of resources, select the ones with wider social impact.	For the same SEE, select the resources that have to be found outside.

TABLE IX. Learning Outcome 5 Assessment

LO5 - Creating a conductive stakeholder environment			
Compulsory		**Optional**	
Individual	*Group*	*Individual*	*Group*
• Ask one question to select features of every customer segment. • - Given a set of situations and idea of SEE, choose the value proposition for each one.	• Present the set of roles derived from the strengths obtained for each group member. Starting from this set of roles, relate every role to activities that could appropriately be performed by each of them. • Build a sentence defining what value proposition is. Two or three correct options are allowed, each of them with different scoring according to the completeness of the definition.	Each of the group members has to choose one possible value proposition for a concrete idea in a concrete local situation.	From the set of customer segments, and given one enterprise sector, choose two customer segments that better fit with that enterprise.

TABLE X. Learning Outcome 6 Assessment

LO6 - Evaluation and impact assessment			
Compulsory		**Optional**	
Individual	*Group*	*Individual*	*Group*
	• A set of n questions has to be answered, but each one will be presented to a group member. That way, all of them need to know about budget.	A set of true/false questions in which several rules about implementation are presented.	(Assuming different risks apply to different sectors or kinds of enterprise) given an enterprise (real or ideal), select the three most important risks it assumes.

TABLE XI. Learning Outcome 6 Assessment

LO6 - Evaluation and impact assessment			
Compulsory		**Optional**	
Individual	*Group*	*Individual*	*Group*
	• From a list of activities, select the ones related to social impact. • The same template is used in different points by presenting different sets of activities every time (different correct answers will be needed, of course). • Given a list of elements, classify their costs into fixed, variables, mandatory, optional, remains, etc.	If a bad score is obtained on 7.2.1's first question, the question could be presented again here in order to improve scoring and to fix concepts.	(Assuming different risks apply to different sectors or kinds of enterprise) given an enterprise (real or ideal), select the three most important risks it assumes.

VI. Case Study

This pedagogical approach has been assessed with the support of five of the facilitators who took part in the piloting. They have been provided with the pedagogical description, and they have tested the platform, Social PlaNet, in which this pedagogical approach has been implemented. In this section, we outline their opinions [55]. The goal of this assessment was testing the effectiveness of the proposal as well as finding elements to be improved.

A. Pilot Information

Five facilitators have participated in this assessment: three from Spain and two from Italy. Their assessment refers to the piloting activities carried out with 63 learners in total (36 in Spain and 27 in Italy), divided into 14 groups (seven in each country). Out of these 14 groups, two were run testing a completely online training/gaming model, whereas the others were run with a blended model. Learner demographics were quite varied, ranging from a group of African refugees in one of the Italian pilots to unemployed but qualified people in some of the Spanish pilots. Pilots were run from May till June 2017, with the work of each group lasting two to three weeks on average. There were two to three meetings per week, with each lasting two-and-a-half to three hours.

B. Results

Facilitators were asked about the benefits for learners that they found in taking part in the pilots, with specific reference to the learning experience and outcomes.

A first general finding was that the pedagogical approach was effective in relation to the expected learning outcomes, since it enhanced the development of the expected knowledge and skills in the terms defined in Table III. The group composition of the pilots had a great influence on the type of most-enhanced skills, since facilitators emphasised different skills in each of them. That way, groups of low-skilled and low-literacy learners were encouraged to develop soft skills (self-esteem, empowerment, capacity to work in groups) rather than specific knowledge and competences. This was the case, for instance, with the Italian pilot addressing African refugees, where the facilitator stated that 'the biggest learning outcome – considering the specificity of the target group (African refugees) was a cultural one: the hardest part was to convince them they could become entrepreneurs (and not just execute the orders of someone else), adding value to the community rather than thinking only of profit for personal benefit (as social entrepreneurs)'. However, in groups involving already-experienced social entrepreneurs and unemployed people with a higher education profile, the main learning outcomes achieved were the following, according to the facilitators interviewed:

- Improvement of skills and knowledge in relation to the creation and management of social economy enterprises, with specific reference to the use of the BMC and marketing techniques.
- Improvement of knowledge about social entrepreneurship.
- Improvement of social awareness.
- Improvement of tolerance for frustration.
- Improvement of collaborative skills and teamwork (working together on a common project).

As anticipated in section V.A, two out of the 14 groups analysed in this paper tested the game completely online, in line with the initial idea of the project to deliver an online game on social entrepreneurship. The results of the online testing were disappointing, since the technical challenges met by learners in proceeding with the game ended up in killing motivation to play, and the two online groups did not complete the testing phase. On the other hand, the remaining 12 groups, which adopted an in-person approach (all learners connected to a PC, but in the same room and with the in-person support of the facilitator), successfully completed the testing, developing business ideas and related business plans.

Once the training program was finalized, about 60% of participants were intended to became entrepreneurs in the real life [55].

C. Discussion

Based on the above results, we can conclude that the pedagogical model elaborated as a support for the Social PlaNet game was partly successful, since it worked with a stronger than foreseen role of the facilitators in guiding the groups and as long as the testing was run in person, or with a blended approach. Technical difficulties related to the gameplay concern both the game development (with specific feedback provided by facilitators on the changes needed to make the game more user-friendly) and the IT skills of the targeted participants. Based on the pilot results, the team is now working to improve the game, making it more user-friendly and granting more power to facilitators to intervene on the online game platform with ad-hoc support material for the training. The recommendations that will be produced for future facilitators willing to use the game will also refer to the need for an introductory session on basic IT skills to allow for more fluent and relaxed gameplay.

VII. Conclusions and Further Works

In this paper, we have presented the pedagogical approach developed in the Social Seducement European project, which has been included in the platform Social PlaNet. This pedagogical approach is based on three main principles: the use of the instructional design model ADDIE; the European Qualification Framework (EQF); and the particularities of the social economy. Based on these foundations, we have presented the set of learning outcomes that are taught in the platform, obtained from a deep study of the literature as well as a wide set of interviews with main actors in this field.

The pedagogical model works on a set of knowledge, skills, and attitudes learned through a gamified e-learning tool (Social PlaNet). This game encourages learners to become social entrepreneurs that develop a business plan for their own social economy enterprises. To support this learning process, Social PlaNet includes a set of levels that help learners to advance from the initial idea to the final business plan by using real-life stories as means to inspire them. Additionally, the game includes learning materials, distributed by country, in order to provide native resources for every learner in every language. This specific feature focused on language-based resources provides the learner with a local library of resources, along with the international scope of the game. Finally, this process is made more dynamic by the inclusion of gamification points, which learners use to furnish their offices, so that the space accommodates their social entrepreneurial activities. This contextualisation of the personal office facilitates a deeper identification of the learner with the role of entrepreneur.

This pedagogical approach has been tested in two pilots developed in Italy and Spain by means of the Social PlaNet platform. These pilots show that results differ depending on the learners' features: the lower the level of literacy, the higher the impact in terms of soft skills; the higher the level of literacy, the higher the level of specific skills, knowledge and attitudes acquired. Further, this game can address various learner profiles that will get particular benefits based on their features. This multi-faceted use of the game allows for multiple learner targets and competences to be addressed using the pedagogical model by Social PlaNet.

Future work will combine a) additional pilots, in order to widen the scope of these results; b) additional facilitators that will support the development of Social PlaNet; and c) a facilitators' network that is fostered by the project partners group, with the aim of giving support to incoming facilitators that start the training of new learner groups.

Social PlaNet is available to be used at the following URL: http://ss.epscms.com. It is open to be used by anyone who wants to use this learning platform for social entrepreneurship training. The website and guidelines are available in several languages, and can be downloaded at https://www.socialseducement.net/game-and-resources .

Appendices

A. Appendix I

Story title: 'Co-op57: From company to worker's cooperative'

Context

Co-op57 is a financial services cooperative that was created in 1995 as a consequence of long years of workers' struggle: When the publisher Bruguera went bankrupt, some of the fired employees put part of their compensation into a common fund to promote social economy projects. They chose the cooperative model because of their commitment to social change.

Problem

Many times, banks ask for very high guarantees before granting a loan, but loan-seekers have difficulty meeting these requirements. Some fired workers have business ideas and need initial support to fund their projects.

The financial crisis made this problem still bigger because public funding comes late or never. In this situation, many enterprises have difficulty providing their services and products.

Solution

Due to the circumstances in which it arose, Co-op57 is highly engaged to change the economy's rules. Therefore, it offers financial services (loans) for auto-managed projects for social change and workers' fight to contribute to change the economy's rules and help create new ways of work.

The most important thing in Co-op57 is confidence. Thus, every client in the co-op becomes a partner, having the power to give opinions and participate in the decision-making process. This makes every client feel like an owner.

Transparency is very important for confidence. Social impact is measured by seeing how projects advance and new partners arrive, but there is no index by which to measure it. Co-op57 has a newsletter every six months provide information about new partners and projects to which it has given loans, along with the amounts of those loans. At present, it has 700 entities as partners, which it considers as having a high degree of social impact.

Projects are evaluated according to their social repercussions and whether they contribute to new types of economy and employment. Evaluating the risks of proposals that the co-op receives is necessary so it can be responsible with partners. However, metrics are different because the viability of the project is very important. Co-op57 usually contributes to the improvement of proposals in order to make the initial projects viable and, hence, able to access loans.

Locally, the co-op promotes new social and caring projects, although it also works with national foundations and organizations focused on changing the economy (Alternative Economy Network, Cooperatives Federation) as well as building a person-based economy.

Its governance structure is based on consensus by a horizontal organization and on territorial semi-autonomous management. Decisions are sent to a central board, which makes the process slow but democratic. However, this auto-management and need for coordination is highly challenging.

Benefits obtained from loans are partially saved. The rest of the benefits are invested annually as partners decide. So far, they have always decided to retain them in Co-op57 to face possible unexpected events.

B. Appendix II

LO 1 – Understanding the concept of social economy	
General description of social economy	• https://es.wikipedia.org/wiki/Econom%C3%ADa_social_en_España • http://www.cepes.es/social/econ_social_que_es
Typology of social economy enterprises	• Cooperatives: http://www.cepes.es/social/entidades_cooperativas • Labor society: http://www.cepes.es/social/entidades_sociedades_laborables • Mutuals: http://www.cepes.es/social/entidades_mutualidades • Insertion companies: http://www.cepes.es/social/entidades_empresas_insercion • Fishermen's associations: http://www.cepes.es/social/entidades_cofradias_pescadores • Special Employment Centres: http://www.cepes.es/social/entidades_centros_empleo
Legislation	• http://www.cepes.es/documentacion/426
LO 2 – Conception of the idea by a set of persons	
Guide for entrepreneurs on the constitution of social economy enterprises	• http://www.cepes.es/index.php?action=carga&a=archivo_6f6cc27_22-10-13_pdf.pdf
How to create a cooperative	• http://www.cepes.es/index.php?action=carga&a=archivo_5f6cc27_22-10-13_pdf.pdf • http://www.cepes.es/index.php?action=carga&a=archivo_2f6cc27_22-10-13_pdf.pdf
LO 3 – Establishing a business plan	
Guide for drafting of the business plan by COCETA	• http://www.cepes.es/index.php?action=carga&a=archivo_2f6cc27_22-10-13_pdf.pdf
Example of Social Business Plan	• https://www.entrepreneur.com/article/268780
Business Model Canvas	• http://www.emprendedores.es/gestion/modelo-3
LO 4 – Management, including communication and marketing	
Concept and importance	• http://www.cepes.es/index.php?action=carga&a=archivo_2f6cc27_22-10-13_pdf.pdf • https://dialnet.unirioja.es/descarga/articulo/2649005.pdf
LO 5 – Creation of conductive stakeholders' environment	
Concept and importance	• http://www.anel.es/la-creacion-de-redes-principal-reto-para-el-desarrollo-de-la-economia-social/
LO 6 – Evaluation and impact assessment	
Concept and importance	• http://www.economiasolidaria.org/documentos/medicion_del_valor_social_y_el_impacto
LO 7 – Operations	
Important skills in entrepreneurship	• http://capsulasdeemprendimiento.blogspot.com.es/2010/10/las-7-habilidades-sociales-de-un.html • http://www.injuve.es/sites/default/files/2014/12/publicaciones/guia_jovenes_talento_perfilemprendedor_3.pdf

Acknowledgment

This work is partially supported by the project Social Seducement (contract number: 2014-1-UK01-KA200-001830) funded by the European Commission under the Erasmus+ Programme, and the Research Institute for Innovation & Technology in Education (UNIR iTED) (http://ited.unir.net) at Universidad Internacional de La Rioja (UNIR).

References

[1] Waitz, T. (1863). *Introduction to anthropology* (Vol. 1). Anthropological Society.

[2] Cohen, R. (1978). Ethnicity: Problem and focus in anthropology. *Annual review of anthropology*, 379–403.

[3] Aarseth, E. (2001). Computer game studies, year one. *Game studies*, 1(1), 1–15.

[4] Bernard, H. R. (2011). *Research methods in anthropology: Qualitative and quantitative approaches*. Rowman Altamira.

[5] Hindle, K. (2002). A grounded theory for teaching entrepreneurship using simulation games. *Simulation & Gaming*, 33(2), 236–241.

[6] Schwartzman, H. (Ed.). (2012). *Transformations: The anthropology of children's play*. Springer Science & Business Media.

[7] Prensky, M. (2005). Computer games and learning: Digital game-based learning. *Handbook of computer game studies*, 18, 97–122.

[8] Prensky, M. (2009). H. sapiens digital: From digital immigrants and digital natives to digital wisdom. *Innovate: journal of online education*, 5(3), 1.

[9] Steyaert, C., and Hjorth, D. (Eds.). (2008). *Entrepreneurship as social change: A third new movements in entrepreneurship book* (Vol. 3). Edward Elgar Publishing.

[10] Mishra, P., and Shrawankar, U. (2016) Comparison between Famous Game Engines and Eminent Games. *International Journal of Interactive Multimedia and Artificial Intelligence* 4(1), 69–77.

[11] Trouillot, M. R. (2016). *Global transformations: Anthropology and the modern world*. Springer.

[12] Scupin, R. (2015). Cultural anthropology a global perspective. Pearson.

[13] Horst, H. A., & Miller, D. (Eds.). (2013). *Digital anthropology*. A&C Black.

[14] Malinowski, B. (1929). Practical anthropology. *Africa*, 2(01), 22–38.

[15] Fried, M. H. (1975). *The notion of tribe*. Benjamin-Cummings Publishing Company.

[16] Keesing, R. M. (1981). Cultural anthropology: A contemporary perspective. Holt McDougal.

[17] Pfaffenberger, B. (1992). Social anthropology of technology. *Annual review of Anthropology*, 491–516.

[18] Guimaraes, M. J. (2005). Doing anthropology in cyberspace: Fieldwork boundaries and social environments. *Virtual methods: Issues in social research on the Internet*, 141-56.

[19] Piaget, J. (1962). Play, dreams, and imitation in childhood. New York: Norton Books.

[20] Gross, R. (2010). *Psychology: the science of mind and behavior*. Hodder Education: London.

[21] Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment* (CIE), 1(1), 20.

[22] Malaby, T. M. (2007). Beyond play a new approach to games. *Games and culture*, 2(2), 95–113.

[23] Malaby, T. M. (2009). Anthropology and play: The contours of playful experience. *New Literary History*, *40*(1), 205-218.

[24] Kline, S., Dyer-Witheford, N., & De Peuter, G. (2003). *Digital play: The interaction of technology, culture, and marketing*. McGill-Queen's Press-MQUP.

[25] Raessens, J. (2006). Playful identities, or the ludification of culture. *Games and Culture*, 1(1), 52–57.

[26] S. Aceto, D. Burgos, D. Parente. (2015) Guide for developing the game and its environment. IO1 report, Social Seducement Project (2014-1-UK01-KA200-001830)

[27] New Media Consortium (2013). NMC Horizon Report 2013 Higher Education Edition, retrieved from https://www.nmc.org/pdf/2013-horizon-report-HE.pdf

[28] New Media Consortium (2015). NMC Horizon Report 2015 Higher Education Edition, retrieved from https://www.nmc.org/publication/nmc-horizon-report-2015-higher-education-edition/

[29] New Media Consortium (2017). NMC Horizon Report 2017 Higher Education Edition, retrieved from http://cdn.nmc.org/media/2017-nmc-horizon-report-he-EN.pdf

[30] BinSubaih, A., Maddock, S., & Romano, D. (2009). Serious games for the police: Opportunities and challenges. *Special Reports & Studies Series at the Research & Studies Center (Dubai Police Academy)*.

[31] McClarty, K. L., Orr, A., Frey, P. M., Dolan, R. P., Vassileva, V., & McVay, A. (2012). A literature review of gaming in education. *Gaming in education*, 1-35.

[32] Ulicsak, M. (2010). *Games in education: serious games*. Bristol: Futurelab.

[33] Chin, J., Dukes, R., Gamson, W. (2009). Assessment in Simulation and Gaming: A review of the last 40 years. *Simulation and Gaming* 40(4), 553-568.

[34] Malone, T (1980). What Makes Things Fun to Learn? A Study of Intrinsically Motivating Computer Games. Palo Alto: Xerox

[35] Kirriemuir, J., McFarlane, A. (n.d.) Literature Review in Games and Learning. FutureLab Series, report 8.

[36] Stewart, J., Bleumers, L., Van Looy, J., Mariën, I., All, A., Schurmans, D., Willaert, K., De Grove, F., Jacobs, A., and Misuraca, G. (2013). *The Potential of Digital Games for Empowerment and Social Inclusion of Groups at Risk of Social and Economic Exclusion: Evidence and Opportunity for Policy.* Joint Research Center, Institute for Prospective Technological Studies (JRC-IPTS).

[37] Stuart, A. (2014). A blended learning approach to safety training: Student experiences of safe work practices and safety culture. *Safety Science* 62, 409–417.

[38] Antonaci, A., Dagnino, F. M., Ott, M., Belloti, F., Berta, R., De Gloria, A., Lavagnino, E., Romero, M., Usart, M., and Mayer, I. (2015). A gamified collaborative course in entrepreneurship: focus on objectives and tools. *Computers in Human Behaviours* 52, 1276–1283.

[39] Jensen, R. 50 great sites for serious educational games, available online at: http://www.onlinecolleges.net/50-great-sites-for-serious-educational-games/

[40] Kuratko, D. F. (2016). Entrepreneurship: Theory, process, and practice. Cengage Learning.

[41] O'Rourke, M. (2014). Increasing engagement with vocational education and training: a case study of computer games-based safety training. *2013 postgraduate research papers: a compendium*. Adelaide, Australia: National Centre for vocational education research, 113–133.

[42] Belloti, F., Berta, R., de Gloria, A., Lavagnino, E., Antonaci, A., Dagnino, F., Ott, M., Romero, M., Usart, M., and Mayer. I. S. (2014). Serious games and the development of an entrepreneurial mindset in higher education engineering students. *Entertainment Computing*, 357–366.

[43] Lim, T., Loucher, S., Hauge, J. B., Stanesco, I. A., Ortiz, I. M., Moreno-Ger, P., Belloti, F., Carvalho, M. B., Earp, J., Ott, M., Arnab, S., and Berta, R. (2014). Narrative Serious Game Mechanics (NSGM) – Insights into the Narrative-Pedagogical Mechanism. *LNCS 8395*, 23–34.

[44] Lukosch, H., van Bussel, R., and Meijer, S. A. (2013). A serious game design combining simulation and sandbox approaches. *LNCS 8264*, 52–59.

[45] Mollick, E. (2014). Entrepreneur simulation: the startup game. Harvard Business Publishing for educators.

[46] Stuart, A. (2014). A blended learning approach to safety training: Student experiences of safe work practices and safety culture. Safety Science 62, 409–417.

[47] Nichols, E., and Wright, A. L. (2015) Using the Everest Team Simulation to Teach Threshold Concepts. Journal of Management Education 39(4), 531–537.

[48] Almeida, F. L. F. (2017). Learning entrepreneurship with serious games – a classroom approach. *Information Technology* 2(1), 1–4.

[49] EUROSTAT. v3.4.1-20170407-5840-PROD_EUROBASEDATA-EXPLORER_PRODmanaged12. Retrieved from http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=lfsi_sup_a&lang=en on 2018, Jan 23th.

[50] S. Aceto, D. Burgos, S. Francario, R. Goergen K. Junge, N. Padilla, D. Parente. (2016) Guide on the Social Seducement RPG learning model. IO3 report, Social Seducement Project (2014-1-UK01-KA200-001830)

[51] N. Padilla-Zea, S. Aceto and D. Burgos, "Social Seducement: Towards the

foundations of a pedagogical model". Ingeniería Solidaria, vol. 13(21), January 2017.
[52] Branch, R. (2009). *Instructional Design: The ADDIE approach*. Springer: New York.
[53] EQF, European Commission, https://ec.europa.eu/ploteus/content/descriptors-page. Last access on March, 8th
[54] K. Junge, C. Castellanos, S. Francario, E. Klaer, R. Goergen and J. Keim. (2015) Target needs' analisys report. IO2 report. Social Seducement project (2014-1-UK01-KA200-001830).
[55] Cullen, J., Drabble, D. (2017). Social Seducement. RPG model validation report. IO6 report. Social Seducement project (2014-1-UK01-KA200-001830).
[56] González-González, C.S., del Río, N.G., Navarro-Adelantado, V. (2018) Exploring the Benefits of Using Gamification and Videogames for Physical Exercise: a Review of State of Art. International Journal of Interactive Multimedia and Artificial Intelligence, 5(2):46-52.

Natalia Padilla-Zea

Dra. Natalia Padilla-Zea is Assisstant Professor at the Universidad Internacional de La Rioja (UNIR). She received her PhD from the University of Granada. She is researcher at the Research Institute for Innovation & Technology in Education (UNIR iTED, http://ited.unir.net), as part of Vice-chancellor for Knowledge Transfer and Technology (http://transfer.unir.net). Her main line of research is focused on e-learning, gamification and game-based learning. She develops her teaching activity at the Engineering and Technology School of UNIR.

Stefania Aceto

Stefania Aceto, born in Bologna (Italy) on 21.12.1970, graduated in Political Sciences and Economics. Since 1998 she is active in research and policy advisory activities concerning innovation in learning; technology enhanced learning and internationalisation of Higher Education. From 2001 to 2013 she was Head of the Observatory Unit on Learning Technologies at Scienter, Italy, conducting research on policy, market and innovative practices in the field of e-learning across all sectors of Education and Training. From 2013 to 2014 she collaborated with the Menon Network EEIG coordinating FP7 and Lifelong learning projects in the area of learning, technology, creativity and innovation. Since 2015 she is a senior researcher and Project manager at the Universidad Internacional de La Rioja (UNIR), working on projects related to quality assurance, game-based learning, social inclusion, digital competences and Open Education.

Daniel Burgos

Prof. Dr. Daniel Burgos works as Vice-rector for Knowledge Transfer & Technology (http://transfer.unir.net), UNESCO Chair on eLearning and ICDE Chair in OER (http://research.unir.net/unesco) at Universidad Internacional de La Rioja (UNIR, http://www.unir.net). He is also Director of the Research Institute for Innovation & Technology in Education (UNIR iTED, http://ited.unir.net). He contributes to research and implementation projects about, i.e. social entrepreneurship (Social Seducement), regional Open Education (OpenMed), clustered Open Education (Open Educators Factory), eLearning (Intuitel), STEM (Inspiring Science Education), etc., funded by the European Commission and Worldwide. His interests are mainly focused on Educational Technology and Innovation: Adaptive/Personalised and Informal eLearning, Open Education, Learning Analytics, Social Networks, eGames, and eLearning Specifications. He has published over 100 scientific papers and 15 books on these matters. He has been the jury chair of the 2016 and 2018 UNESCO Prize for the Use of ICTs in Education. He holds degrees in Communication (PhD), Computer Science (Dr. Ing), Education (PhD), Anthropology (PhD) and Business Administration (DBA).

Image Classification Methods Applied in Immersive Environments for Fine Motor Skills Training in Early Education

Andrés Ovidio Restrepo Rodríguez[1], Maddyzeth Ariza Riaño[1], Paulo Alonso Gaona-García[1], Carlos Enrique Montenegro-Marín[1], Íñigo Sarría[2] *

[1] Universidad Distrital Francisco José de Caldas, Bogotá D.C., Colombia
[2] Universidad Internacional de La Rioja, Logroño (Spain)

Received 15 September 2019 | Accepted 16 October 2019 | Published 30 October 2019

Abstract

Fine motor skills allow to carry out the execution of crucial tasks in people's daily lives, increasing their independence and self-esteem. Among the alternatives for working these skills, immersive environments are found providing a set of elements arranged to have a haptic experience through gestural control devices. However, generally, these environments do not have a mechanism for evaluation and feedback of the exercise performed, which does not easily identify the objective's fulfillment. For this reason, this study aims to carry out a comparison of image recognition methods such as Convolutional Neural Network (CNN), K-Nearest Neighbor (K-NN), Support Vector Machine (SVM) and Decision Tree (DT), for the purpose of performing an evaluation and feedback of exercises. The assessment of the techniques is carried out using images captured from an immersive environment, calculating metrics such as confusion matrix, cross validation and classification report. As a result of this process, it was obtained that the CNN model has a better supported performance in 82.5% accuracy, showing an increase of 23.5% compared to SVM, 30% compared to K-NN and 25% compared to DT. Finally, it is concluded that in order to implement a method of evaluation and feedback in an immersive environment for academic training in the first school years, a low margin of error must be taken in the percentage of successes of the image recognition technique implemented, to ensure the proper development of these skills considering their great importance in childhood.

Keywords

Immersive Environment, Augmented Reality, Image Recognition Techniques, Support Vector Machine, K-nearest Neighbor, Decision Tree, Convolutional Neural Network.

DOI: 10.9781/ijimai.2019.10.004

I. Introduction

Much of the cognitive activities, such as writing, reading or speaking, also include the use of fine motor skills, involving hand movement, hand-eye coordination, sound production control skills, etc.; for this reason, it can be said that there is a relationship between cognitive development and the fine motor [1]. In older children and adults, the development of fine motor skills implies learning the set of complex and specific sequences of movement necessary to perform the activities of daily life, traditional writing and using the keyboard, leisure activities and vocational tasks [2]. In studies such as [1], it is determined that attention, fine motor skills and general knowledge help to predict the percentages that students will subsequently obtain in math, reading and science, so these indicators are essential for preparation in early education. In addition, and according to the above, in [3] the authors perform the identification of brain wave patterns of people with dyslexia during the writing process, stating that writing skills are a relevant academic indicator. Therefore, since there is a need for the use of fine motor skills in cognitive activities, the constant motivation in school classrooms of fine motor practice is important.

Technology represents a great importance nowadays and there has been a great growth in the implementation of new mechanisms and means that allow to involve technology in education, as a way to increase motivation in learning. This is how Augmented Reality (AR) permits the creation of different immersive scenarios through the use of virtual elements on a real environment, facilitating the acquisition of knowledge, meaningful learning and user interest [4]-[5].

There are numerous references that have focused on the creation of AR environments in education and their influence on motivation, as well as promoting motor activities; studies such as [6]-[7] present a 3D navigation structure through the use of mobile devices, based on a Knowledge Organization System (KOS) using school subjects about animals, in a population of children between 10 and 12 years, in order to analyze aspects of learning and usability, through a set of questions related to the system; the results show that, in the test carried out, most participants successfully identified the animals, their relationship with the habitat and the category of the animal, according to the classification defined in the

* Corresponding author.
E-mail addresses: A. O. Restrepo Rodríguez (aorestrepor@correo.udistrital.edu.co), M. Ariza Riaño (marizar@correo.udistrital.edu.co), Paulo Alonso Gaona-García (pagaonag@udistrital.edu.co), Carlos Enrique Montenegro-Marín, (cemontenegrom@udistrital.edu.co), Íñigo Sarría (inigo.sarria@unir.net).

KOS. Authors such as [8] propose a virtual environment based on 3D navigation for gestures with two hands, offering control on the speed of navigation, being easy and intuitive; when evaluating the technique with a set of participants, the study finds its feasibility making it easy to learn and use, in addition to reaching high rates of accuracy, performance and prediction speed. In research such as [9], the planning, assignment and use of digital learning objects with augmented reality is proposed to support the teaching-learning process of higher education students, taking into account the students' learning style, resulting in a better academic performance, in addition to making the learning process more dynamic and attractive to students. For their part, the authors of [10] develop an augmented reality platform to simulate a 3D book versus a traditional book for chemistry learning in high school students, given the low popularity of the subject, resulting in those students who used the platform had better results in their semiannual exams, compared to those who used the book traditionally. In [11] a method of learning in Augmented Reality was carried out through the use of mobile devices, using multidimensional concept maps, and having a standard group and an experimental group for testing; as a result, it is evident that the experimental group performed significantly better than the other group and a greater motivation is concluded during the development of the activity. In a study carried out by [12], a mobile platform is exposed through the use of augmented reality, for descriptive geometry learning; results show a great increase in positive feelings on the part of the students, through the proposed method that is characterized by the use of virtual models of the figures.

Likewise, there are studies related to AR in early education, as well as the influence of technology in increasing motivation and interest during the learning process in children. For their part in [13] they develop an interactive-educational game for learning concepts such as colors, geometric shapes and mathematics through augmented reality, for children between 4 and 7 years old, obtaining as a result a pleasant and fun teaching environment, as well as the increased interest and commitment on the part of the students, who were also evaluated by their parents. In [14] content on animals is implemented for children aged 4 to 5 years, based on augmented reality; as a result, the authors present an increase in the effectiveness of learning and an increase in student activity, as augmented reality is the promoter of active behavior and development of communication skills in children.

Images classification methods in artificial intelligence aim to improve the feature and pattern extraction learning, and their application can be seen in a number of contexts, including education [15]. Methods such as Support Vector Machine (SVM) that is based on the principle of margin maximization [16], K-Nearest Neighboor (K-NN) whose purpose is to find the nearest neighbor between nodes [17], Convolutional Neural Network (CNN) that it is a neural network characterized by using the convolution operation [18] and Decision Tree (DT) defined as a simple representation of a finite set of classes [19], are considered the most commonly used and known traditional approaches [20] [21]. Studies like [22] present the application of a convolutional neural network model for plants recognition as of an images dataset; as a result, 86.2% effectiveness was seen with the CNN model implemented. In the field of education, references such as [15] worked with a classification method and the use of augmented reality by presenting a CNN model for geometric figures classification captured from an AR-Sandbox, with the aim of supporting early education and fine-motor therapy in children. The authors found that the proposed method shows a decrease of 39.45% based on loss and an increase of 14.83% of correct answers, concluding that there is an increase in performance with the selected CNN model.

Taking into account the above, this study aims to apply the classification methods SVM, CNN, K-NN and DT with images captured from the immersive environment presented in [23], a study that addresses an environment based on augmented reality in support of training in children from 4 to 6 years old, through 3D scenarios where children reinforce the learning of vowels and their training. The application of the methods is done in order to provide feedback on the activities carried out by the children and these will be compared using a confusion matrix, classification report and cross validation as metrics to measure their performance, given that, traditionally, machine learning-based approaches are assessed in a cross validation scenario that validates the classification model by assessing how the result will generalize to an independent dataset [24]. Additionally, the representation of a confusion matrix helps in analyzing the belongingness of each class and makes a clear view of the classification accuracy of each class showing the correct classification and misclassification of each class against the belongingness of each target class [25], in addition, usually classification report is used to check the quality of classification algorithm predictions. The rest of this article is organized as follows: in Section II there are Materials and methods where the general scenario is presented, which is divided into the immersive environment and the development of activities feedback with the application of images classification methods; Section III presents results and discussion; finally, conclusions and future work are presented.

II. Materials and Methods

Fig. 1 shows the general scenario with which it will work, structured with two main components: *Immersive environment* and *Activities feedback*. The Immersive environment component is based on an Augmented Reality prototype where a series of 3D scenarios are presented with which it interacts through a gesture control device and whose objective is the practice of vowels for the support of fine-motor therapy in children from 4 to 6 years old. The Activities feedback component complements the first by determining if the child's exercise corresponds to the expected result; this is achieved with the application of the classification methods CNN, SVM, K-NN and DT, taking as a starting point the images obtained with the immersive environment. Finally, in order to measure the performance of the methods, a comparison is made between them by means of confusion matrix, classification report and cross validation as performance metrics, considering that they are techniques that have been addressed in studies such as [20] [21] [26], presenting a comparison between the methods when recognizing images in fields such as the identification of diseases in plants, numbers and elements of the daily life using data sets such as MNIST and CIFAR 10 and finally, identification of types of rice grains.

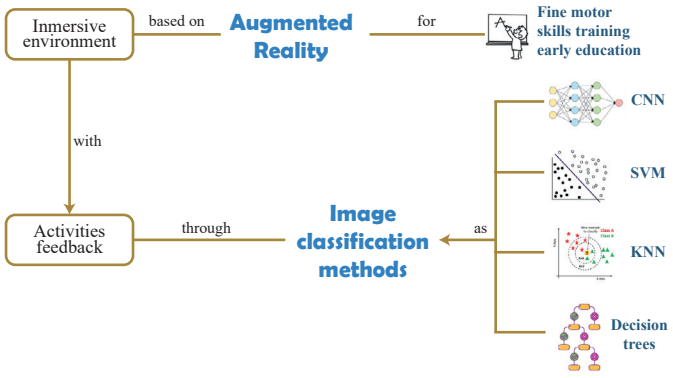


Fig. 1. General scenario. Source: Authors

A. Immersive Environment

The immersive environment used is presented in [23], where the authors developed a set of scenarios based on Unity 3D. There are five modules, one for each vowel and each one consists of scenarios

where multimedia elements, such as videos, related words and images, are presented for the learning of the vowel, and a training scenario for its practice. The interaction with the scenarios is done through the Leap Motion gesture control device which detects and tracks the position of the participant's hands and fingers in space through two monochromatic stereo cameras and three infrared LEDs [27], such as it is shown in Fig. 2. In this way, the user can play the videos, change the scenario and move through the different modules.

Fig. 2. Use of the leap motion device.Source: Leap motion [28].

The training scenario presents a blank board, indicating the vowel to be written, where the child should use their hands to write on the board, through the Leap Motion device (Fig. 3); this allows the practice of the writing of the vowels and at the same time the fine-motor skills are worked by involving the movement of the hands and the hand-eye coordination. From this platform the images of the vowels are obtained by taking a screenshot of the training board.

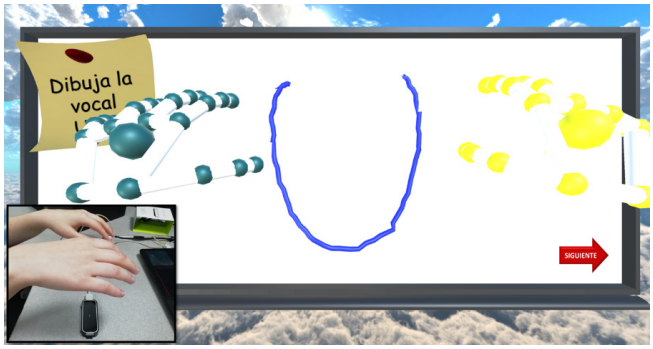


Fig. 3. Vowel writing in the immersive environment. Source: Authors.

B. Activities Feedback

Fig. 4 presents the approach to carry out the development of the module that corresponds to Activities Feedback. This component has as its starting point the training of machine learning models such as Support Vector Machine (SVM), K-Nearest Neighbor (K-NN) and Decision Tree (DT), using the Grid Search method, in order to perform hyperparameter optimization and obtain the best model of each technique according to the vowel training dataset in accordance with the nature and operation of the immersive environment presented previously. Additionally, an already defined model of convolutional neural network (CNN) will be trained, which has been taken from studies such as [15], exposing as a result a base model of a CNN, with the objective of recognizing images acquired from an immersive environment called AR-Sandbox, presenting an accuracy in prediction phase of 0.87. Once the models have been trained and selected, the evaluation and comparison of the models is carried out, using metrics such as confusion matrices, a classification report made up of precision score, f1 score and recall score and, finally, performing cross-validation.

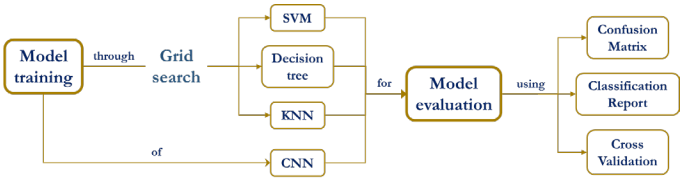


Fig. 4. General approach for Activities Feedback component. Source: Authors.

In order to perform the training of the SVM, K-NN, DT, and CNN models, a training dataset made up of 5 classes corresponding to the vowels (A, E, I, O, U) was used. The dataset is determined by a total of 1600 images, where, each class has a total of 320 images, taken from the internet or made in some basic drawing tool. Fig. 5 presents the training set used.

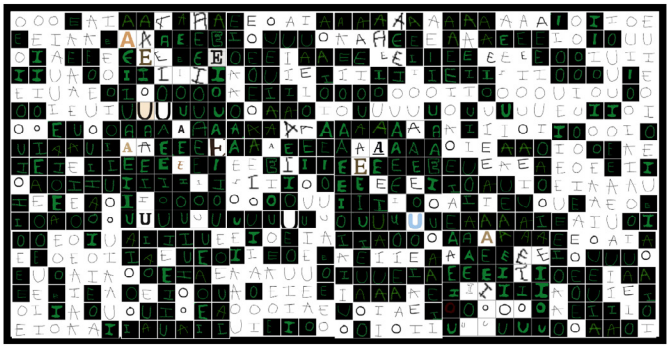


Fig. 5. Training dataset. Source: Authors

1. CNN Model

As mentioned above, the selection of the convolutional neural network model was based on a study carried out by [15], where the definition of a base model of a CNN was carried out, performing hyperparameter optimization by Random Search, defining a hyperparameters dictionary. These parameters are taken randomly in the base model, with the purpose of training a total of 50 models, to make the selection of a final model, having as decision criteria the accuracy, the average square error and the loss function, metrics presented in this study. The selected model had the purpose of recognizing and classifying geometric figures acquired directly from an immersive environment such as an AR-Sandbox. The model is presented in Fig. 6.

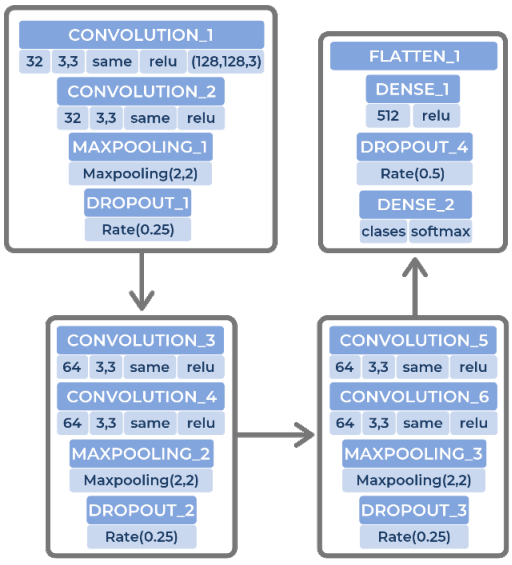


Fig. 6. CNN model. Source: Restrepo, Casas et al., 2019 [15]

When the base model is defined, we proceed to train it with the training dataset described above, obtaining an average of accuracy of 0.9758 and a loss function of 0.0874. Table I, presents the defined compilation and training hyperparameters.

TABLE I. Compilation and Training Hyperparameters

Hyperparameter	Value
Optimizer	RMSProp
Loss Function	Cross Entropy
Metrics	Accuracy
Batch_size	32
Epochs	10

Source: Authors

2. SVM Model

In order to obtain the SVM model by optimizing hyperparameters using Grid Search, a hyperparameters dictionary is defined, declaring variable parameters such as, the kernel that will use the algorithm and the penalty value for the error. Table II presents the defined hyperparameters dictionary.

TABLE II. SVM Hyperparameter Dictionary

Hyperparameter	Possible Values
Kernel	'linear', 'poly', 'rbf', 'sigmoid'
Penalty parameter	1, 10, 100, 100, 1000

Source: Authors

As a result, a total of 20 models were trained, taking the accuracy value as the decision criterion, the model with the best performance had as hyperparameters a ***'poly'*** type kernel and a ***penalty value for the error*** of 1.0, obtaining an average accuracy of 0.595. Fig. 7, presents the model scheme with its hyperparameters.

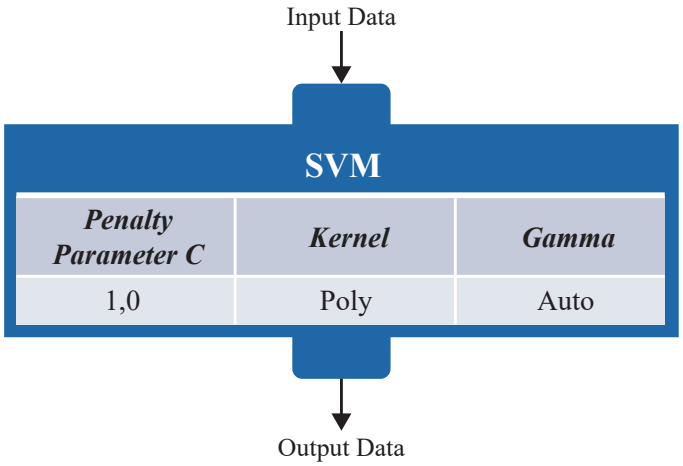


Fig. 7. SVM scheme. Source: Author

3. K-NN Model

On the other hand, in order to select the K-NN model using Grid Search, Table III presents the hyperparameters dictionary defined, declaring variable the number of neighbors and the weight function used in the prediction phase.

TABLE III. K-NN hyperparameter dictionary

Hyperparameter	Possible Values
Number of neighbors	2, 3, 5, 15, 30, 60, 100
Weights function	'uniform', 'distance'

Source: Authors.

When defining the hyperparameters dictionary, a total of 14 models were trained, calculating their average accuracy score, considering it as a decision criterion. The model with the highest precision value was obtained by using ***3 neighbors*** and a ***distance*** type weight function, obtained 0.625 as the score value. Fig. 8, presents the model scheme with its hyperparameters.

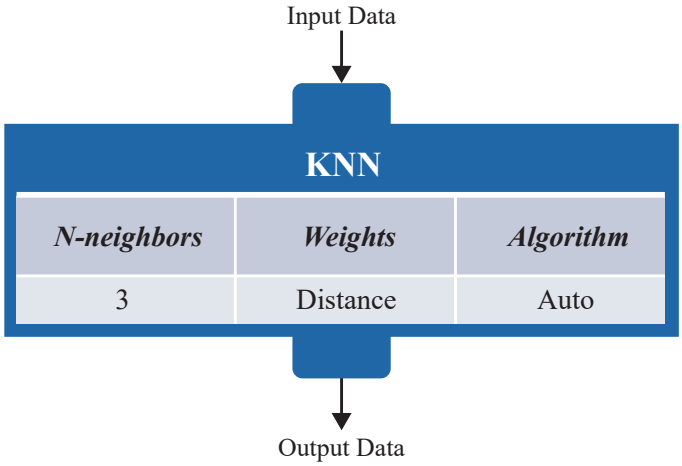


Fig. 8. K-NN scheme. Source: Authors.

4. DT model

Finishing the training of machine learning models, the decision tree model is selected using Grid Search, which is why a hyperparameters dictionary is defined, leaving the criteria to measure the quality of a division and the strategy to choose the division in each of the nodes. Table IV exposes the defined hyperparameters dictionary.

TABLE IV. Decision Tree Hyperparameter Dictionary

Hyperparameter	Possible Values
Criterion	'gini', 'entropy'
Splitter	'best', 'random'

Source: Authors.

According to the hyperparameters dictionary, only a total of 4 models were trained, where the best combination occurs when the criterion for measuring the quality of a division is made using the ***Gini*** function and the ***best*** function is used as a division strategy in each node. This model had an average accuracy value of 0.61. Fig. 9, shows the model scheme with its hyperparameters.

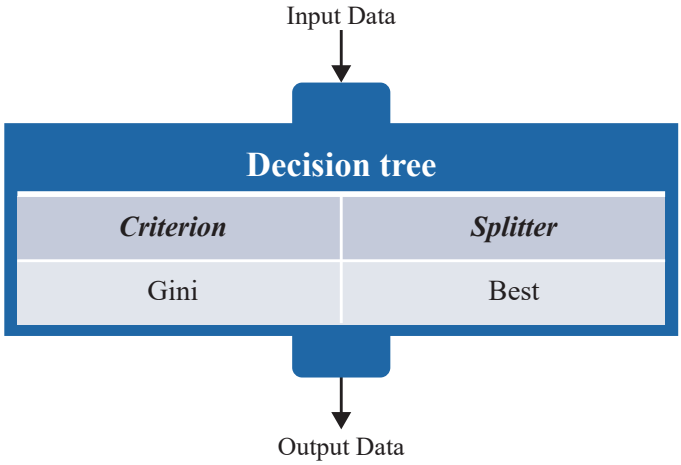


Fig. 9. Decision Tree scheme. Source: Authors.

Fig. 10. Test dataset. Source: Authors.

III. Results and Discussions

In order to perform the evaluation and comparison of the models defined above, a total of 14 children were cited, who interacted with the immersive environment and made 40 images for each class, that is, 40 images for each of the vowels, forming a set of test data of a total of 200 images. Fig. 10 exposes the test dataset.

On the other hand, the models were evaluated by means of confusion matrices, a classification report, which is composed of metrics such as f1 score, recall score and precision score and, finally, by cross-validation, the above with the purpose of obtaining approximately the accuracy value both specifically by vowel and generally from each of the models, and after that, making a comparison between each of the techniques implemented.

A. Confusion Matrix

In order to obtain the confusion matrices of each model with the test dataset, the sk-learn library was used, which was implemented in Python.

1. CNN Model

Table V, presents the number of successes in the prediction phase with the CNN model with a total of 40 samples per class, where a success rate of 82.5% was obtained.

TABLE V. Number of Hits in CNN

	A	E	I	O	U
A	36	2	0	1	1
E	3	35	1	1	0
I	2	6	30	0	2
O	4	3	0	31	2
U	0	0	3	4	33

Source: Authors.

2. SVM Model

Table VI exposes the number of hits in the prediction phase with the SVM model with a total of 40 samples per class, where a 59% hit rate was obtained.

TABLE VI. Number of Hits in SVM

	A	E	I	O	U
A	20	4	5	9	2
E	1	24	6	5	4
I	5	2	19	9	5
O	3	3	0	34	0
U	3	5	1	10	21

Source: Authors.

3. K-NN Model

Table VII shows the number of hits in the prediction phase with the K-NN model with a total of 40 samples per class, where a 52.5% hit rate was obtained

TABLE VII. Number of Hits in K-NN

	A	E	I	O	U
A	7	1	4	28	0
E	2	18	6	12	2
I	0	0	24	14	2
O	0	0	1	37	2
U	0	0	5	16	19

Source: Authors.

4. DT Model

Table VIII, presents the number of hits in the prediction phase with the DT model with a total of 40 samples for each class, where a 57.5% hit rate was obtained.

TABLE VIII. Number of Hits in DT

	A	E	I	O	U
A	21	7	7	4	1
E	8	18	4	6	4
I	6	4	21	5	4
O	3	7	0	28	2
U	1	2	6	4	27

Source: Authors.

Based on the confusion matrices presented above, it can be seen that the model that shows the best performance is CNN, since it has a total of 82.5% success rate in its predictions, exposing an increase of 23.5% compared to SVM, 30% against K-NN and 25% against DT. A specific analysis of the confusion matrices is presented in Table IX, where H is the percentage of correct answers and F the percentage of errors.

As shown in Table IX, it can be evidenced that when performing a specific analysis of each of the vowels, the CNN model significantly overcomes the rest of the models, except in the O vowel, given that the SVM gets over it by a 0.75% and K-NN by 1.5%.

B. Classification Report

With the purpose of obtaining metrics such as: precision score, recall score and f1 score, the ***sk-learn*** library was used, through its classification report module, which includes the metrics mentioned above. Table X presents the average of these metrics for each of the trained models.

TABLE IX. Confusion Matrix Analysis

Method	Vowel									
	A		E		I		O		U	
	H	F	H	F	H	F	H	F	H	F
CNN	0.9	0.1	0.875	0.125	0.75	0.25	0.775	0.225	0.825	0.175
SVM	0.5	0.5	0.6	0.4	0.475	0.525	0.85	0.15	0.525	0.475
K-NN	0.175	0.825	0.45	0.55	0.6	0.4	0.925	0.075	0.4	0.6
DT	0.525	0.475	0.45	0.55	0.525	0.475	0.7	0.3	0.675	0.325

Source: Authors.

TABLE X. Average Metrics for Classification Report

Model	Precision_score	Recall_score	F1_score
CNN	0.830	0.826	0.824
SVM	0.610	0.586	0.586
K-NN	0.700	0.516	0.510
DT	0.584	0.594	0.582

Source: Authors.

Additionally, Fig. 11. Classification report values. Source: Authors.11 graphically represents the consolidated data in Table X. As of Fig. 11. Classification report values. Source: Authors.11, it can be seen that the CNN model presents a higher level of performance compared to the rest of the techniques, in addition, this means that for the SVM, K-NN and DT models there are a greater number of false positives, therefore, the values in these metrics tend to be lower 24.13% on average compared to the CNN model.

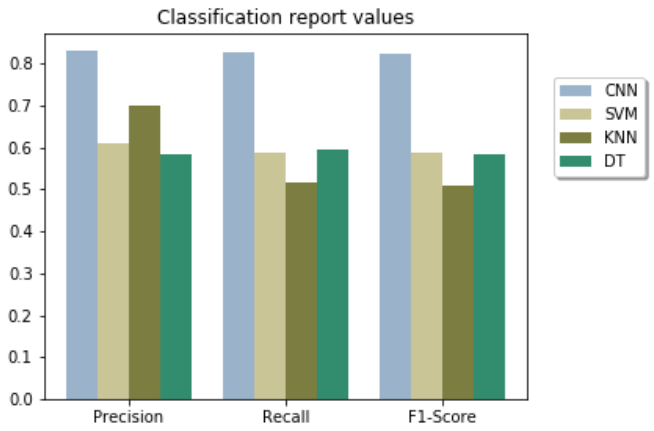


Fig. 11. Classification report values. Source: Authors.

C. Cross Validation

In order to carry out the cross-validation process, use was made of both the training dataset and the test dataset, that is 1800 images. Additionally, a total of 10 randomly divisions were established, using functions such as ***KFold*** and ***cross_val_score*** of the ***model_selection*** module of the ***sk-learn*** library. Table XI, exposes the average of the cross-validation score and its standard deviation for each of the models and for the number of divisions defined.

TABLE XI. Score and Standard Deviation in Cross Validation

Model	Average Cross Validation Score	Standard Deviation
CNN	0.8651	0.0451
SVM	0.5306	0.0485
K-NN	0.5137	0.0255
DT	0.5638	0.0457

Source: Authors

As of Fig. 12. Model comparison by cross validation. Source: Authors.12, it can be seen that when applying cross-validation and calculating its score, the CNN model shows an increase of 0.3345 against SVM, 0.3514 against K-NN and 0.3013 against DT, on average it increases by 32.9%, in addition, this model exposes a high score above 0.85. On the other hand, from Fig. 12. Model comparison by cross validation. Source: Authors.12, which presents the consolidated data in Table XI through a box diagram, it can be stated that, as a first instance, the median of the CNN model is located approximately 0.25 points above the median of the other models. In addition, it has a relatively close maximum value of 0.9 and a minimum value approximately close to 0.8, therefore its values do not vary significantly, that is, its precision value calculated when applying cross validation remains stable.

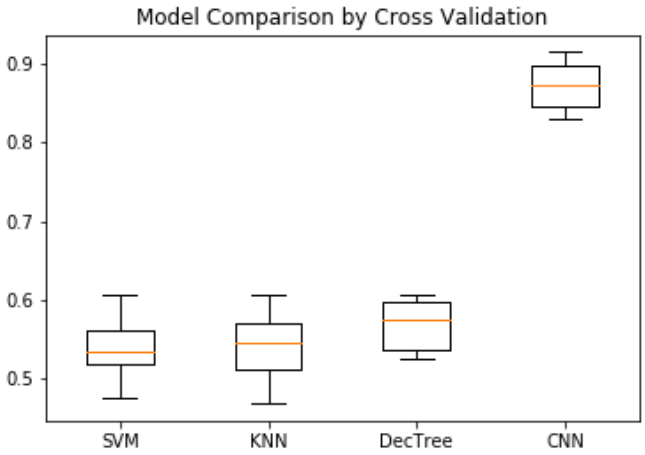


Fig. 12. Model comparison by cross validation. Source: Authors.

IV. Conclusions

Based on the results obtained in the evaluation of the confusion matrix metrics, precision score, recall score, F1 score and cross validation, it is possible to conclude that the CNN model implemented has a better performance compared to traditional methods such as K-NN, SVM and DT, even obtaining the best models of these techniques through hyperparameter optimization. Methods such as K-NN and SVM presented a greater number of successes in the prediction of the O vowel, compared to the CNN method, when analyzing the confusion matrix; nevertheless, this may imply that these methods created a predisposition to the vowel in its phase of extracting features and training.

Given the importance of the educational and motor training of children in the first school years, it is necessary to have a high level of percentage of successes in the model implemented to carry out the evaluation and feedback of the exercises performed in an immersive environment that supports the development of these skills.

Although the recognition of images associated with vowels, geometric figures, among others, has been presented in a large number of previous studies, it is important to consider that the immersive environment used involves a variety of multimedia elements, which can generate noise. Despite the above, the CNN model presented a high level of precision in its prediction stage; however, this value must be refined and increased, perhaps as the authors of [15] do, using an image cleaning technique.

The immersive environment presented in this study was taken from the work carried out by [24], where the authors have as their main objective the reinforcement of fine motor skills through the writing of vowels; however, since there is no way to evaluate the exercise, it is not guaranteed that the children have fulfilled the objective. That is why when obtaining a method with a high level of success to carry out feedback on the exercises performed in immersive environments such as the one mentioned above, the design and implementation of a monitoring and recommendation system for students is considered as future work, being applied from its early stages of education which can be decisive for its formation. Additionally, it is possible to include new scenarios by implementing different exercises by broadening the model recognition spectrum.

References

[1] Grissmer, D., Grimm, K. J., Steele, J. S., Aiyer, S. M., and Murrah, W. M. (2010). Fine Motor Skills and Early Comprehension of the World: Two New School Readiness Indicators. *Developmental Psychology,* 46(5), pp. 1008–1017. doi: 10.1037/a0020104.

[2] Case-Smith, J. (2013). *Encyclopedia of Autism Spectrum Disorders*. New York, United States: Volkmar, Fred R.

[3] Perera H, Shiratuddin MF, Wong KW and Fullarton K. (2018) EEG Signal Analysis of Writing and Typing between Adults with Dyslexia and Normal Controls. *International Journal of Interactive Multimedia and Artificial Intelligence*, 5(1), pp. 62-67. doi: 10.9781/ijimai.2018.04.005.

[4] Che Hashim, N., Abd Majid, N. A., Arshad, H. and Khalid Obeidy, W. (2018). User Satisfaction for an Augmented Reality Application to Support Productive Vocabulary Using Speech Recognition. *Advances in Multimedia,* 2018, pp. 1-10. doi: 10.1155/2018/9753979.

[5] Sveistrup, H. (2004). Motor rehabilitation using virtual reality. *Journal of NeuroEngineering and Rehabilitation,* 1(10), pp. 1-8. doi: 10.1186/1743-0003-1-10.

[6] Garcia Gaona, P. A., Martin Moncunill, D., Gordillo, K. and González Crespo, R. (2016). Navigation and Visualization of Knowledge Organization Systems using Virtual Reality Glasses. *IEEE Latin America Transactions,* 14(6), pp. 2915-2920. doi: 10.1109/TLA.2016.7555275.

[7] Garcia Gaona, P. A., Martin Moncunill, D., Gordillo, K. and Montenegro, C. (2016, june). "Navigation and Visualization of Knowledge Organization Systems using Virtual Reality Glasses: first insights". Presented in *UNESCO-UNIR ICT & Education Latam 2016,* Bogotá, Colombia.

[8] Rehman, I., Ullah, S. and Raees, M. (2019). Two Hand Gesture Based 3D Navigation in Virtual Environments. *International Journal of Interactive Multimedia and Artificial Intelligence*, 5(4), pp. 128-140. doi: 10.9781/ijimai.2018.07.001.

[9] Medina, M. A., García, C. F. J. and Olguín, M. J. A. (2018). Planning and Allocation of Digital Learning Objects with Augmented Reality to Higher Education Students According to the VARK Model. *International Journal of Interactive Multimedia and Artificial Intelligence,* 5(2), pp. 53-57. doi: 10.9781/ijimai.2018.02.005

[10] Pasaréti, O, Hajdú, H., Matuszka, T., Jámbori, A., Molnár, I. and Turcsányi-Szabó, M. (2012). Augmented Reality in education. pp. 1-10.

[11] Chen, C.-H., Huang, C.-Y. and Chou, Y.-Y. (2019). Effects of augmented reality-based multidimensional concept maps on students' learning achievement, motivation and acceptance. *Universal Access in the Information Society*, 18(2), pp. 257–268. doi: 10.1007/s10209-017-0595-z.

[12] Gutiérrez de Ravé, E., Jiménez-Hornero, F. J., Ariza-Villaverde, A. B. and Taguas-Ruiz, J. (2016). DiedricAR: a mobile augmented reality system designed for the ubiquitous descriptive geometry learning. *Multimedia Tools and Applications,* 75(16), pp. 9641–9663. doi: 10.1007/s11042-016-3384-4.

[13] Zhu, Y., Yang, X. and Wang, S. J. (2017). Augmented Reality Meets Tangibility: A New Approach for Early Childhood Education. *EAI Endorsed Transactions on Creative Technologies,* 4(11), pp. 1-8. doi: 10.4108/eai.5-9-2017.153059.

[14] Cascales, A., Laguna, I., Pérez-López, D., Perona, P. and Contero, M. (2013, July). "An Experience on Natural Sciences Augmented Reality Contents for Preschoolers". Presented in *International Conference on Virtual, Augmented and Mixed Reality VAMR 2013: Virtual, Augmented and Mixed Reality. Systems and Applications*, Las Vegas, NV, USA.

[15] Restrepo Rodríguez, A. O., Casas Mateus, D. E., Gaona García, P. A., Montenegro Marín, C. E. and González Crespo, R. (2018). Hyperparameter Optimization for Image Recognitionover an AR-Sandbox Based on Convolutional NeuralNetworks Applying a Previous Phase of Segmentation by Color–Space. *Symmetry,* 10(743), pp. 1-16. doi: 10.3390/sym10120743.

[16] Adankon, M. and Cheriet, M. (2009). *Encyclopedia of Biometrics*. Boston, MA: Li S.Z., Jain A.

[17] Ault, Aaron & Zhong, Xuan & Coyle, Edward. (2005). K-nearest-neighbor analysis of received signal strength distance estimation across environments.

[18] Ketkar, N. (2017). *Deep Learning with Python*. Apress, Berkeley, CA: Springer, pp. 63-78.

[19] Ganten, D. (2006). *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine*. Berlin, Heidelberg: Springer.

[20] Lin, P., Li, X. L., Chen, Y. M. and He, Y. (2018). A Deep Convolutional Neural Network Architecture for Boosting Image Discrimination Accuracy of Rice Species. *Food and Bioprocess Technology,* 11(4), pp. 765–773. doi: 10.1007/s11947-017-2050-9.

[21] Wang, Y. and Xin, M. (2019). Research on image classification model based on deep convolution neural network. *EURASIP Journal on Image and Video Processing,* 2019(40). doi: 10.1186/s13640-019-0417-8.

[22] Dyrmann, M., Karstoft, H. and SkovMidtiby, H. Plant species classification using deep convolutional neural network. *Biosystems Engineering,* (151), pp. 72-80. doi: 10.1016/j.biosystemseng.2016.08.024.

[23] Ariza Riaño, M., Restrepo Rodríguez, A. O., Gaona García, P. and Montenegro Marín, C. (2018, november). "3D Immersive Environments to Support Training Activities Through Fine Motor Therapy". Presented in *11th annual International Conference of Education, Research and Innovation*, Seville, Spain. doi: 10.21125/iceri.2018.1675.

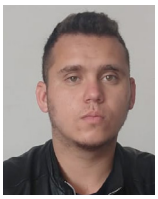
[24] Allix, K., Bissyandé, T. F., Jérome, Q., Klein, J., State, R. and Le Traon, Y. (2016). Empirical assessment of machine learning-based malware detectors for Android. *Empirical Software Engineering,* 21(1), pp. 183–211. doi: 10.1007/s10664-014-9352-6.

[25] Bisoi, R., Dash, P. K. and Das, P. P. (2018). Short-term electricity price forecasting and classification in smart grids using optimized multikernel extreme learning machine. *Neural Computing and Applications,* pp. 1-24. doi: 10.1007/s00521-018-3652-5.

[26] Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D. and Stefanovic, D. (2016). Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification. *Computational Intelligence and Neuroscience*, 2016, pp. 1-11. doi: 10.1155/2016/3289801.

[27] Tung, J. Y., Lulic, T., Gonzalez, D. A., Tran, J., Dickerson, C. R. and Roy, E. A. (2015). Evaluation of a portable markerless finger position capture device: accuracy of the Leap Motion controller in healthy adults. *Physiological measurement,* 36(5), pp. 1025-1035. doi: 10.1088/0967-3334/36/5/1025.

[28] Ultrahaptics Ltd. (2019). *Leap Motion*. In: https://www.leapmotion.com/.



Andrés Ovidio Restrepo-Rodríguez

Is an active member of Interactive Multimedia and Digital Animation research group at Engineering Faculty of Universidad Distrital Francisco José de Caldas, Bogotá, Colombia. He finished a degree on System Engineer on 2019 and currently he is doing an MSc in information science and Communication at Universidad Distrital Franscico José de Caldas. His research interests include Artificial Intelligence, Data Mining, Heterogenous Computing, Virtual Environments, Learning Analytics.

Maddyzeth Ariza-Riaño

She is a Systems Engineer and a master's student on Information Science and Communication at Engineering Faculty of Universidad Distrital Francisco José de Caldas, Bogotá, Colombia. Her research interests include Data Mining, Software Architecture and Semantic Web.

Paulo Alonso Gaona-García

Is Associate Professor and active member of GIIRA research group at Engineering Faculty of Universidad Distrital Francisco José de Caldas, Bogotá – Colombia. He obtained his Ph.D. in Information of Engineering and Knowledge at University of Alcalá in 2014. He finished a degree on System Engineer on 2003 and obtained an MSc in Information Science and Communication in 2007 at Universidad Distrital Francisco José de Caldas. His research interest includes Web science, network and communications, information security, e-learning, information visualisation and semantic Web.

Carlos Enrique Montenegro-Marin

Is a Ph.D.in systems and computer services for internet from University of Oviedo, Asturias, Spain (2012). He has a Diploma of advanced studies 2008 of the Pontifical University of Salamanca. He is MSc. Science in Information and Communication Systems from the Universidad Distrital Francisco José de Caldas. He is a System engineer. His research interests include Object-Oriented technology, Language Processors, Modeling Software with, DSL and MDA.

Íñigo Sarría

Íñigo Sarría holds a Doctorate (Cum Laude and special prize) from the International University of La Rioja, a Licentiate degree in Mathematics from the University of the Basque Country, a University Expert qualification in Analysis of the Knowledge Society from The International University of La Rioja, and a Certificate in Teaching from the Complutense University of Madrid. He is Director of the Computer Science and Technology Area of the Higher School of Engineering and Technology at the International University of La Rioja and is the Coordinator of the Master's in Mathematical Engineering and Computing. ORCID code: https://orcid.org/0000-0002-2584-9671

LA UNIVERSIDAD
EN INTERNET

Rectorado
Avenida de la Paz, 137
26006 Logroño (La Rioja)
t (+34) 941 21 02 11

www.unir.net