International Journal of Interactive Multimedia and Artificial Intelligence

March 2019, Vol. V, Number 4

ISSN: 1989-1660

I believe this artificial intelligence
is going to be our partner.
If we misuse it, it will be a risk. If we use it right,
it can be our partner.
Masayoshi Son

Special Issue on Artificial Intelligence Applications

INTERNATIONAL JOURNAL OF INTERACTIVE MULTIMEDIA AND ARTIFICIAL INTELLIGENCE
ISSN: 1989-1660 –VOL. 5, NUMBER 4

Editor's Note

The term artificial inteligence (AI) basically refers to using machines to do things that we consider to be "intelligent", that is, being able to either simulate or do things that we describe people as doing with their cognitive faculties[1]. A more complete definition presents artificial intelligence as the ability of a machine to perform cognitive functions we associate with the human minds, such as perceiving, reasoning, learning, interacting with the environment, problem solving, and even exercising creativity[2]. The concept is related to machines that process huge amounts of information, learn from the results of their actions and never rest. The concept of AI makes us think that human genius has managed to create something that seems to exceed its own capacity.

The term was introduced by Alan Turing in 1950[3], so it is actually a pretty well-known field. However, we have seen an acceleration lately in the use of the AI due to two main factors. First, computational power is rising with exponential growth, and second, the amount of available data has grown at an impressive rate in recent years. To a large extent, then, the exponential growth in data and in computational power has led to the hype of AI.

One of the remarkable aspects of AI is the degree to which it is an extension of the features that we have seen in data and analytics [4] [5] [6]. One of the enabling factors for machine learning to take hold is the large amounts of data. We have seen more and more data collected by companies and all kinds of organizations, be it transactional data, voice data, or data from the Internet of Things in the physical world. When you have all that data, you can extend the work you have done in analytics with AI techniques. Therefore, you will see methodologies about machine learning and deep learning with new neural networks that are applied to those vast amounts of data. In this sense, there are four technology systems of which machine learning is just part of and where some of the recent advancements and developments have been happening:

1. Physical AI, i.e., robotics and autonomous vehicles [7] [8].
2. Computer vision, i.e., image processing or video processing [9].
3. Natural-language processing, be it spoken language particularly, or written language. We are seeing a lot of natural-language work being done[10] [11].
4. Virtual agents or conversational interfaces; this is the ability of systems to roughly converse with you whether by voice or online through chats[12] [13].

To some extent, artificial intelligence is going through a bit of a hype cycle. There are a lot of applications, a lot of industries and activities, and a lot of value is at stake. But it can be said that today we are in a phase where we have applications which we would call narrow AI. Those are very specific tasks that machines today can do better than human beings. But there is that question about a general AI, where you have a broader spectrum of capabilities that can be managed by a machine[14] [15].

We are beginning to enter that phase. And we should not forget that the speed of development is exponential in those key technologies. It is coming much, much faster than we can imagine. As Peter Diamandis says, cumulative "intelligence" (both artificial and human) is the single greatest predictor of success for both a company or a nation[16].

Therefore, the suggestion for leaders of all types of organizations (companies, hospitals, government agencies, etc.) would be to start an analytics transformation now if have not done so already. This will require them to build capabilities, build technology and start the change in the organization, which will also be necessary to ultimately go into AI-enabled processes.

We can ask ourselves whether there is a case for a portfolio-of-initiatives approach, where one considers what can be done here and now [17].

In this sense, one possible suggestion would be to take the right use cases at the right point in time. Getting started now with the easier and simpler use cases also prepares us to take the more advanced use cases in the future. Empowering organizations to become analytics- or AI-driven is key to success in the future.

With these ideas in mind, we have prepared this special issue. It has been designed with the primary objective of demonstrating the diversity of fields where AI is used and, consequently, how it is gaining increasing importance as a tool for analysis and research. In this sense, there are works related to the following topics: the use of AI with the IoT, campaign management, topic models and fusion methods, sales forecasting, price forecasting for electricity market, NLP techniques in computational medicine, evaluation of patient triage in hospital emergency settings, algorithms for solving the assignment problem, scheduling strategy for scientific workflow, driver fatigue detection mechanisms, virtual reality and specialized training, image segmentation, web service selection, multimedia documents adaptation, 3D navigation in virtual environments, multi-criteria decision-making methods and emotional states classification.

The first paper of this special issue, "A Review of Artificial Intelligence in the Internet of Things", is written by Cristian González García, Edward Rolando Núñez-Valdez, Vicente García-Díaz, B. Cristina Pelayo G-Bustelo and Juan Manuel Cueva Lovelle, a group of researchers at the Universidad de Oviedo, in Spain. It presents some examples of the use of AI with the IoT and how this fusion can create very important and interesting applications. The paper touches on four of the most important parts of AI: Machine Learning, Computer Vision, Fuzzy Logic, and Natural Language Processing. These four subfields are very related to each other; mainly, by the algorithms that are used to process the different information or the techniques that are used to teach the information necessary to create the model that different techniques need. Furthermore, these subfields are used to create smarter machines or programs to help us in our daily life, like the Internet of Things is trying to do. For instance, Machine Learning could be used to create a better module with enough intelligence to automate and make decisions, Computer Vision to automate visual scenarios, Fuzzy Logic to give intelligence and make decisions to save money, electricity, and so on, and NLP to improve the understanding between machines and humankind.

The paper by Ramón Alberto Carrasco, María Francisca Blasco, Jesús García-Madariaga and Enrique Herrera-Viedma, "A Fuzzy Linguistic RFM Model Applied to Campaign Management", states that in the literature there are some proposals for integrated schemes for campaign management based on segmentation from the results of the RFM model. RFM is a technique used to analyze customer behavior by means of three variables: Recency, Frequency and Monetary Value. It is very much in use in the business world because of its simplicity of use, implementation and interpretability of its results. However, RFM applications to campaign management present limitations like lack of precision because the scores of these variables are expressed by an ordinal scale. In this paper, the authors propose to link customer segmentation methods with campaign activities in a more effective way by incorporating the 2–tuple model to both the RFM calculation

DOI: 10.9781/ijimai.2018.07.004

process and its subsequent exploitation by means of segmentation algorithms, specifically, k-means. This yields greater interpretability of these results and also allows computing these values without loss of information. Therefore, marketers can effectively develop more effective marketing strategies.

The paper, "Topic Models and Fusion Methods: a Union to Improve Text Clustering and Cluster Labeling", written by Mohsen Pourvali, Salvatore Orlando and Hosna Omidvarborna, focuses on modeling algorithms. Topic modeling algorithms are statistical methods that aim to discover the topics running through the text documents. Using topic models in machine learning and text mining is popular due to its applicability in inferring the latent topic structure of a corpus. This paper presents an enriching document approach, using state-of-the-art topic models and data fusion methods, to enrich documents of a collection with the aim of improving the quality of text clustering and cluster labeling. The authors propose a bi-vector space model in which every document of the corpus is represented by two vectors: one is generated based on the fusion-based topic modeling approach, and one simply is the traditional vector model. The experiments carried out in on various datasets show that using a combination of topic modeling and fusion methods to create documents' vectors can significantly improve the quality of the results in clustering the documents.

The next paper, "Sales Prediction through Neural Networks for a Small Dataset", written by Rosa María Cantón Croda, Damián Emilio Gibaja Romero, Santiago Omar and Caballero Morales, all of them researchers from the Universidad Popular Autónoma del Estado de Puebla (UPAEP), in Mexico, is also related to digital marketing. Sales forecasting allows firms to plan their production outputs, which contributes to optimizing firms' inventory management via a cost reduction. However, not all firms have the same capacity to store all the necessary information over time. Therefore, time-series with a short length are common within industries, and problems arise because small time series do not fully capture sales' behavior. This paper shows the applicability of neural networks in a case where a company reports a short time-series given the changes in its warehouse structure. Given neural networks' independence from statistical assumptions, this paper uses a multilayer-perceptron to get the sales forecasting of the company. The authors found that learning rates variations do not significantly increase the computing time, and the validation fails with an error less than five percent.

The paper "Day-ahead price forecasting for Spanish electricity market" is written by Álvaro Romero, José Ramón Dorronsoro, and Julia Díaz from the Universidad Autónoma de Madrid and the "Instituto de Ingeniería del Conocimiento", in Madrid, Spain. In recent years, electrical systems around the world and in particular the Spanish electric sector have undergone great changes with the focus of making them more liberalized and competitive markets. For this reason, in many countries like Spain, there have appeared electric markets where producers sell and electricity retailers buy the power we consume. All agents involved in this market need predictions of generation, demand and especially prices to be able to participate in them in a more efficient way and obtain a greater profit. The present work explores the development of a tool that allows us to predict the price of electricity for the next day in the most precise way possible. For such a target, this document analyzes the electric market to understand the calculation of prices and identify the agents that can make prices vary. Traditional proposals in the literature range from the use of Game Theory to the use of Machine Learning, Time Series Analysis or Simulation Models. In this project it was proposed a normalization of the target variable on an hourly and daily basis because of a strong seasonal component in order later to benchmark several models of Machine Learning: Ridge Regression, K-Nearest Neighbors, Support Vector Machines, Neural Networks and Random Forest. After observing that the best model is Random Forest, a discussion was carried out on the appropriateness of the normalization for this algorithm. From this analysis, it was obtained that the model that gives the best results without applying the normalization function is Random Forest. This is because of the loss of the close relationship between the objective variable and electricity demand, which obtains an Average Absolute Error of 3.92€ for the whole period of 2016.

The article presented by Antonio Moreno Sandoval, Julia Díaz, Leonardo Campillos Llanos and Teófilo Redondo, "Biomedical Term Extraction: NLP techniques in Computational Medicine" shows a completely different approach. Artificial Intelligence and its Natural Language Processing branch are chief contributors to recent advances in classifying documentation and extracting information from assorted fields. Medicine is one that has received a lot of attention because of the amount of information generated in public professional journals and other means of communication within the medical profession. The typical information extraction task from technical texts is performed via an automatic term recognition extractor. Automatic Term Recognition (ATR) from technical texts is applied for the identification of key concepts for information retrieval and, secondarily, for machine translation. Term recognition depends on the subject domain and the lexical patterns of a given language, in our case, Spanish, Arabic and Japanese. This article presents the methods and techniques for creating a biomedical corpus of validated terms, with several tools for optimal exploitation of the information therewith contained in said corpus. This paper also shows how these techniques and tools have been used in a prototype.

The paper "Evaluation of a Diagnostic Decision Support System for the Triage of Patients in a Hospital Emergency Department", written by J.C. Nazario Arancibia, F.J. Martín Sánchez, A.L. Del Rey Mejías, J. González del Castillo, J. Cháfer Vilaplana, M.A. García Briñón, M.M. Suárez-Cadenas, J. Mayol Martínez and G. Seara Aguilar, is another example of the important AI applications in the field of healthcare, specifically, evaluating a computer-aided diagnosis decision support system. In developed countries, there has been a significant increase in the use of hospital emergency services. Triage is the first evaluation and classification process used to prioritize patients who arrive at the emergency department (ED). One of the greatest challenges for the management of the service is to provide tools that make it possible to expedite the management of patients in the shortest time possible from the moment of their arrival, especially for those who present pathologies that are not selected as a high priority by the classification systems, and thus generate unnecessary overcrowding of the ED. Diagnostic decision support systems can be a powerful tool for guiding diagnosis, facilitating correct classification and ultimately improving patient safety. With an observational criterion and without interfering with the emergency process, the authors use in parallel the Mediktor, the brand registered by Teckel Medical (Mediktor Corp.) system, which uses a sequence of questions guided by its algorithm, to obtain the most frequent expected diagnostic possibilities. The researchers compare the results obtained by the system with the final diagnosis of the usual emergency procedure. The level of accuracy of Mediktor as a support tool for establishing the final diagnosis of patients was 76.5%, higher than that published in similar programs.

The paper "Hybrid Algorithm for Solving the Quadratic Assignment Problem," written by Mohammed Essaid Riffi and Fatima Sayoti, both from the University of Chouaib Doukkali, El Jadida (Morocco), focuses on an optimization problem that has multiple applications. The Quadratic Assignment Problem (QAP) is a combinatorial optimization problem; it belongs to the class of NP-hard problems. This approach is applied in various fields such as hospital layout, scheduling parallel production lines and analyzing chemical reactions for organic compounds. This paper describes an application of the Golden Ball algorithm mixed with Simulated Annealing (GBSA) to solve QAP. The simulated annealing search can be blocked in a local optimum because of the unacceptable movements. The strategy proposed in this paper guides the simulated annealing search to escape from the local optima and to explore the search space in an efficient way. To validate the proposed approach, numerous simulations were conducted on 64 instances of QAPLIB to compare GBSA with existing algorithms in the literature of QAP. The numerical results obtained show that the GBSA produces optimal solutions in reasonable time; it has a better computational time. This work demonstrates that the solution proposed is effective in solving the quadratic assignment problem.

The paper, "Data-Aware Scheduling Strategy for Scientific Workflow Applications in IaaS Cloud Computing", written by Sid Ahmed Makhlouf and Belabbas Yagoubi focuses on the optimization of scientific workflows. Scientific workflows benefit from the cloud computing paradigm, which offers access to virtual resources provisioned on pay-as-you-go and on-demand basis. Minimizing resources costs to meet user's budget is very important in a cloud environment. Several optimization approaches have been proposed to improve the performance and the cost of data-intensive scientific Workflow Scheduling (DiSWS) in cloud computing. However, in the literature, the majority of the DiSWS approaches focused on the use of heuristic and metaheuristic as an optimization method. Furthermore, the tasks hierarchy in data-intensive scientific workflows has not been extensively explored in the current literature. Specifically, in this paper, a data-intensive scientific workflow is represented as a hierarchy, which specifies hierarchical relations between workflow tasks, and an approach for data-intensive workflow scheduling applications is proposed. In this approach, first, the datasets and workflow tasks are modeled as a conditional probability matrix (CPM). Second, several data transformation and hierarchical clustering are applied to the CPM structure to determine the minimum number of virtual machines needed for the workflow execution. In this approach, the hierarchical clustering is done with respect to the budget imposed by the user. After data transformation and hierarchical clustering, the amount of data transmitted between clusters can be reduced, which can improve cost and makespan of the workflow by optimizing the use of virtual resources and network bandwidth. The performance and cost are analyzed using an extension of Cloudsim simulation tool and compared with existing multi-objective approaches. The results demonstrate that the approach presented in this paper reduce resources cost with respect to the user budgets.

"Driver Fatigue Detection using Mean Intensity, SVM, and SIFT" is the title of the paper written by Saima Naz, Sheikh Ziauddin and Ahmad R. Shahid from the COMSATS Institute of Information Technology, Islamabad (Pakistan). Driver fatigue is one of the major causes of accidents. This has increased the need for driver fatigue detection mechanisms in vehicles to reduce human and vehicle loss during accidents. In the proposed scheme, the authors capture videos from a camera mounted inside the vehicle. From the captured video, they localize the eyes using the Viola-Jones algorithm. Once the eyes have been localized, they are classified as open or closed by using three different techniques, namely, mean intensity, SVM, and SIFT. If the eyes are closed for a considerable amount of time, it indicates fatigue and, consequently, an alarm is generated to alert the driver. The experiments show that SIFT outperforms both mean intensity and SVM, achieving an average accuracy of 97.45% on a dataset of five videos, each of which are two minutes long.

The article "PRACTICA. A Virtual Reality Platform for Specialized Training Oriented to Improve the Productivity", written by Juan Manuel Lombardo, Miguel Ángel López, Vicente M. García, Mabel López, Francisco Fernandez Muela, Rubén Cañadas, Ismael Medina, Susana Velasco, Mónica León and Felipe Mirón, all members of the R & D Open Source Foundation (FIDESOL), analyzes the possibility of using virtual reality to increase productivity.The growth of virtual reality glasses that are coming to a market mostly oriented to the purchase of video games is opening new possibilities of virtual reality (VR) exploitation. Therefore, the PRACTICA project is defined as a new service which offers specialized training companies a system for creating courses based on a VR simulator that brings students an experience close to reality. The issue in creating these virtual courses is the need for programmers that can generate them. To overcome this obstacle, PRACTICA allows the creation of courses without the need to program source codes. In addition, elements of virtual interaction that cannot be used in a real environment because of risks for staff, have been incorporated, such as the introduction of fictitious characters or obstacles that interact with the environment. To do this, artificial intelligence techniques have also been incorporated so that environment elements can interact with the user in the stage. This feature offers the opportunity to create situations and scenarios that are even more complex and realistic. This project aims to release a service to bring virtual reality technologies and artificial intelligence to non-technological companies, so that they can generate (or acquire) their own content and give them the desired use for their purposes.

The research, "Multilevel Thresholding for Image Segmentation Using an Improved Electromagnetism Optimization Algorithm" presented by Ashraf M. Hemeida, Radwa Mansour and M. E. Hussein, focuses on image processing. Image segmentation is considered one of the most important tasks in image processing, which has several applications in different areas such as; industry agriculture, medicine, etc. In this paper, the authors develop an electromagnetic optimization (EMO) algorithm based on levy function, EMO-levy, to enhance the EMO performance for determining the optimal multi-level thresholding of image segmentation. In general, EMO simulates the mechanism of attraction and repulsion between charges to develop the individuals of a population. EMO takes random samples from search space within the histogram of image, where, each sample represents each particle in EMO. The quality of each particle is assessed based on Otsu's or Kapur objective function value. The solutions are updated using EMO operators until determine the optimal objective functions. Finally, this approach produces segmented images with optimal values for the threshold and a few number of iterations. The proposed technique is validated using different standard test images. Experimental results prove the effectiveness and superiority of the proposed algorithm for image segmentation compared with well-known optimization methods.

The work "QoS based Web Service Selection and Multi-Criteria Decision Making Methods" is presented by Pallavi Bagga, Aarchit Joshi and Rahul Hans, a group of researchers at Lovely Professional University and DAV University, both in Punjab (India). With the continuing proliferation of web services that offer similar efficacies, around the globe, it has become a challenge for a user to select the best web service. In literature, this challenge is exhibited as a 0-1 knapsack problem of multiple dimensions and multiple choices, known as an NP-hard problem. The Multi-Criteria Decision Making (MCDM) method is one way to solve this problem and helps users to select the best service based on his/her preferences. In this regard, this paper assists researchers in two ways: firstly, to witness the performance of different MCDM methods for a large number of alternatives and attributes, and secondly, to perceive the possible deviation in the ranking obtained from these methods. To carry out the experimental evaluation, in this paper, five different well-known MCDM methods are implemented and compared over two different scenarios of 50 as well as 100 web services, where their ranking is defined on an account of several Quality of Service (QoS) parameters. Additionally, a Spearman's Rank Correlation Coefficient has been calculated for different pairs of MCDM methods in order to provide a clear depiction of MCDM methods which show the least deviation in their ranking. The experimental results aid web service users in making an appropriate decision about the selection of a suitable service.

The paper, "Multimodal Generic Framework for Multimedia Documents Adaptation", written by Hajar Khallouki and Mohamed Bahaj, focuses on multimedia documents adaptation. Nowadays, people are increasingly capable of creating and sharing documents (which generally are multimedia oriented) via the internet. These multimedia documents can be accessed at anytime and anywhere on a wide variety of devices, such as laptops, tablets and smartphones. The heterogeneity of devices and user preferences has raised a serious issue for multimedia contents adaptation. The research focuses on multimedia documents adaptation and more specifically on interaction with users and exploration of multimodality. The authors propose a multimodal framework for adapting multimedia documents based on a distributed implementation of W3C's Multimodal Architecture and Interfaces applied to ubiquitous computing. The core of the proposed architecture is the presence of a smart interaction manager that accepts context related information from sensors in the environment as well as from other sources, including information available on the web and multimodal user inputs. The interaction manager integrates and reasons over this information to predict the user's situation and service use. A key to realizing this framework is the use of an ontology that braces up the communication and representation, and the use of the cloud to insure the service continuity on heterogeneous mobile devices. Smart city is assumed as the reference scenario.

The reaserch, "Two Hand Gesture Based 3D Navigation in Virtual Environments", carried out by I. Rehman, S. Ullah and M. Raees focuses on natural interaction. This issue is gaining popularity due to its simple, attractive, and realistic nature, which realizes direct Human Computer Interaction (HCI). In this paper, the authors present a novel two hand gesture based interaction technique for 3 dimensional (3D) navigation in Virtual Environments (VEs). The system uses computer vision techniques for the detection of hand gestures (colored thumbs) from real scene and performs different navigation (forward, backward, up, down, left, and right) tasks in the VE. The proposed technique also allows users to efficiently control speed during navigation. The proposed technique is implemented via a VE for experimental purposes. Forty participants performed the experimental study. Experiments revealed that the proposed technique is feasible, easy to learn and use, having less cognitive load on users. Finally gesture recognition engines were used to assess the accuracy and performance of the proposed gestures.

The last paper of this special issue, "Are instructed emotional states suitable for classification? Demonstration of how they can significantly influence the classification results in an automated recognition system", was written by M. Magdin and F. Prikler, from Constantine the Philosopher University in Nitra, Faculty of Natural Sciences, Department of Computer Science, Nitra, Slovakia. At the present time, various freely available or commercial solutions are used to classify a subject's emotional state. Classification of the emotional state helps us to understand how the subject feels and what they are experiencing in a particular situation. Classification of the emotional state can thus be used in various areas of our life such as neuromarketing, the automotive industry (determining how emotions affect driving), and implementing such a system into the learning process. The learning process, which is the (mutual) interaction between the teacher and the learner, is an interesting area in which individual emotional states can be explored. Several research studies were carried out in this pedagogical-pyschological area. These studies in some cases demonstrated the important impact of the emotional state on the results of the students. However, for comparison and unambiguous classification of the emotional state, most of these studies used the instructed (even constructed) stereotypical facial expressions of the most well-known test databases (Jaffe is a typical example). Such facial expressions are highly standardized, and the software can recognize them with a fairly high degree of accuracy, but this does not necessarily point to the actual success rate of the subject's emotional classification in such a test because the similarity to real emotional expression remains unknown. Therefore, the authors examined facial expressions in real situations and subsequently compared these facial expressions with the instructed expressions of the same emotions (the Jaffe database).

Francisco Mochón

References

[1] Peter Breuer and Michael Chui. (2018). Artificial intelligence in business: Separating the real from the hype. McKinsey. Podcast. November 2017.

[2] Michael Chui, Vishnu Kamalnath and Brian McCarthy. (2018). An executive's guide to AI. McKinsey Quarterly. February 2018.

[3] A. M. Turing. (1950). Computing Machinery and Intelligence. Mind 49, pp. 433-460.

[4] Antonio Moreno, Teófilo Redondo. Text Analytics: the convergence of Big Data andArtificial Intelligence, International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 3, Nº6, pp. 25-32.

[5] Hsinchun Chen, Roger H. L. Chiang and Veda C. Storey. (2012). BusinessIntelligence and Annalytics: from big data to big impact. MIS quarterly, 2012.

[6] Vishal Gupta and Gurpreet S. Lehal. (2009). A Survey of Text MiningTechniques and Applications. Journal of Emerging technologies in web intelligence, Vol. 1, Nº1 August 2009.

[7] Nick Statt. (2017). The next big leap in AI could come from warehouse robots - The Verge. 1 Jun. 2017.

[8] Alex Owen-Hill. (2017). What's the Difference Between Robotics and Artificial Intelligence? Robotiq. July 19, 2017.

[9] Raj Talluri. (2017). Conventional computer vision coupled with deep learning makes AI better. Network World, Nov. 29, 2017.

[10] Marek Bardoński. (2017). Natural Language Processing in Artificial Intelligence is almost human-level accurate. Worse yet, it gets smart!. Towards Data Science. Nov. 2, 2017.

[11] Egor Dezhic. (2017). Artificial Intelligence in Natural Language Processing. Becoming Human. Artificial Intelligence. June 11, 2017.

[12] Gil Press. (2017). AI By The Numbers: 33 Facts And Forecasts About Chatbots And Voice Assistants. Forbes. May 15, 2017.

[13] Michael McTear, Zoraida Callejas, David Griol. (2016). Conversational Interfaces: Devices, Wearables, Virtual Agents, and Robots. Springer. May 20, 2016.

[14] Cade Met Building A.I. That Can Build A.I. - The New York Times. Nov. 2, 2017.

[15] Gaurav Batra, Andrea Queirolo, Nick Santhanam. (2018).Artificial intelligence: The time to act is now. McKinsey Article. January 2018.

[16] Peter H. Diamandis. (2016). Exponential Growth Will Transform Humanity in the Next 30 Years. Singularity Hub. Dec. 21, 2016.

[17] Gautam Narula. (2018). Everyday Examples of Artificial Intelligence and Machine Learning. TechEmergence.Environments, 24(3), 2016, 590-605. doi:10.1080/10494820.2014.908927

TABLE OF CONTENTS

OPEN ACCESS JOURNAL

ISSN: 1989-1660

A Review of Artificial Intelligence in the Internet of Things

Cristian González García*, Edward Rolando Núñez-Valdez, Vicente García-Díaz, B. Cristina Pelayo G-Bustelo, Juan Manuel Cueva Lovelle

MDE Research Group, Department of Computer Science, University of Oviedo, Oviedo (Spain)

Received 10 October 2017 | Accepted 3 February 2018 | Published 24 March 2018

Abstract

Humankind has the ability of learning new things automatically due to the capacities with which we were born. We simply need to have experiences, read, study… live. For these processes, we are capable of acquiring new abilities or modifying those we already have. Another ability we possess is the faculty of thinking, imagine, create our own ideas, and dream. Nevertheless, what occurs when we extrapolate this to machines? Machines can learn. We can teach them. In the last years, considerable advances have been done and we have seen cars that can recognise pedestrians or other cars, systems that distinguish animals, and even, how some artificial intelligences have been able to dream, paint, and compose music by themselves. Despite this, the doubt is the following: Can machines think? Or, in other words, could a machine which is talking to a person and is situated in another room make them believe they are talking with another human? This is a doubt that has been present since Alan Mathison Turing contemplated it and it has not been resolved yet. In this article, we will show the beginnings of what is known as Artificial Intelligence and some branches of it such as Machine Learning, Computer Vision, Fuzzy Logic, and Natural Language Processing. We will talk about each of them, their concepts, how they work, and the related work on the Internet of Things fields.

Keywords

Artificial Intelligence, Machine Learning, Learning Systems, Computer Vision, Fuzzy Logic, Natural Language Processing.

DOI: 10.9781/ijimai.2018.03.004

I. Introduction

THIS was the question that was asked by Alan Mathison Turing and to which he responded with the proposal of The Imitation Game [1]. The Imitation Game is played by three people, 'A' who is a man, 'B' who is a woman, and 'C' who is the interrogator and can have any of the genders.

The interrogator is in a room other than 'A' and 'B'. During the game, 'C' can ask them any kind of question to determine who is the man and who is the woman, for example, about the length of the hair. 'C' knows the participants under the labels 'X' and 'Y'. When the game ends, 'C' must say 'X' is 'A' and 'Y' is 'B' or vice versa. In this game, the tone of voice does not help 'C', since 'A' and 'B' must write the answers, preferably using a machine, which are sent by a teletype or through another intermediary to 'C'.

This problem can be thought with a variant that is based on replacing the man 'A' with a machine. In this case, 'C' could ask 'A' and 'B' to write poetry, but the machine could refuse. Given this, 'C' could ask a mathematical operation and see that the machine solved it or ask them if they play chess. However, to avoid undue comparisons in the game, the interrogator is prevented from asking for demonstrations, since in some cases the machine is better and in others, the human being is better, as well as each having different skills. For these reasons, this game is about the machine trying to imitate the behaviour of the human being, assuming it can give answers just like a human being would do it naturally. From this theory, the well-known Turing Test emerged.

* Corresponding author.

E-mail address: gonzalezcristian@uniovi.es

However, the imitation from a machine is very difficult to be given, as Turing says from different points of view, namely: the theological, the reluctant people to know the problem (Head-in-the-sand), the mathematician, the one of the conscience, the one of the different disabilities, from the point of view of the memories of Ada Lovelace [2], from the continuity of the nervous system, from the argument of the familiarity of the behaviour and from the extrasensory perception.

For the reasons expressed from the different points of view, Alan Turing focused on what Ada Lovelace commented when she says that a man can 'inject' an idea into a machine. Thus, based on this, he hypothesised that he could inject knowledge if there was a space large enough for it, or at least, so that the machine could play The Imitation Game and leaving as a problem how to program the machines to overcome the game, because according to Turing, they already had the hardware capacity.

To carry out this idea, he proposed to make a child's brain instead of trying to make an adult brain, and thus educate it to obtain the brain of an adult. In this way, he divides the problem into two parts: the 'child' program and the education process. He explains that we should not expect the first 'child' machine to come out on the first attempt and that we should teach it to see how its learning evolves. Thus, after several attempts, they would be getting better machines (or even worse), something that Turing compared with the process of evolution and that several researchers would later develop under the name of genetic algorithms [3].

Based on this, Turing talked about different experiments he carried out and the need of machines to simulate a human being, as is the case of the incorporation of random elements to imitate the possibility of failure that human beings have. To conclude, he commented on the thinking of many people that one could start making machines compete with humans in abstract activities, such as playing chess. However, from his point of view, it was better to mount the machines on the better 'organs' that money could buy and teach them to understand and speak English.

II. What is Artificial Intelligence?

Computers can only process zeros and ones. However, years ago, **Artificial Intelligence (AI)** was born to offer the possibility of creating programs that would allow computers to learn. This was the purpose of the first computer scientists, such as Alan Mathison Turing, John von Neumann and Norbert Wiener, among others, who tried to equip computers with programs that contained **intelligence**, the ability to self-copy, learn and control their environment, being, the principle of this, trying to model the human brain, imitate human learning and simulate biological evolution [4].

Research in **Artificial Intelligence** began after World War II, possibly being Alan Mathison Turing the first to investigate in this field in 1947 [5], and publishing an article about whether machines could think in 1950 [1]. This article is where the famous **The Imitation Game** comes from, from which the **Turing Test** emerged. Both serve to know if a machine is intelligent enough to confuse it with a human being. In that article, Turing also explained how an **AI** system should be created, starting with the creation of a child machine that would be learning to be as an adult. Subsequently, in the decade of the 50s, many more researchers joined this field [5].

Later, John McCarthy, Marvin L. Minsky, Nathaniel Rochester and Claude E. Shannon coined the term **Artificial Intelligence** in 1955 in the Dartmouth summer research project [6]-[8]. In this project a stay of two months was proposed to investigate the principles about learning or another type of intelligence and see if the computers could simulate it, or, rather, with the name they coined, to investigate in **Artificial Intelligence**.

But what exactly is **Artificial Intelligence**? According to John McCarthy [5], **Artificial Intelligence** is the science and engineering that tries to make machines intelligent, trying to get them to understand human language and to reach problems and goals as well as a human being. However, there are some authors who affirm that the objective of the **AI** is to put the human mind inside a computer, although they may speak metaphorically. One of the ways to know if a machine is intelligent is the **Turing Test**.

To answer the previous question, one must also understand what is meant by **intelligence**, for which a reference can be made to [5]. **Intelligence** is the computational part of the ability to achieve objectives and has various types and degrees as occurs in humans, many animals and some machines. However, it is not necessary to consider intelligence to answer 'yes' or 'no' to a question, because, although this is sometimes the goal, machines can do much more, such as, for example, solving problems.

Some of the fields where **Artificial Intelligence** can be applied are, for example, video games, data mining, discourse recognition, natural language understanding, Computer Vision, the creation of expert systems and data classification [5]. Thus, **Artificial Intelligence** can be divided into several branches among which stand out Automatic Learning, Computer Vision, Fuzzy Logic, Natural Language Processing, Heuristics and Intelligent Agents.

III. The Six Rules of Artificial Intelligence

As for robotics, there are the three rules, created by Isaac Asimov in the first literature about robots, Satya Nadella, CEO of Microsoft, offered in an interview to the online magazine Slate[1] an outline of six rules that should be observed by the designers of Artificial Intelligence [9].

1. **AI must be designed to assist humanity**. It needs to respect human autonomy, using collaborative robots to perform dangerous works, such as mining, thus safeguarding human workers.
2. **AI must be transparent, being aware of how the technology works and its rules**. Or as Satya Nadella said: 'Technology will know things about humans, but humans should know about machines'. This would allow humans to understand how technology sees and analyses the world because ethics and design must go hand in hand.
3. **AI must maximise effectiveness without destroying the dignity of the people**. It must preserve cultural commitments, strengthening diversity. A broader, deeper and more diverse commitment to the population is needed because technology should not dictate the values or the virtues of the future.
4. **AI must be designed for intelligent privacy**. Sophisticated methods of protection are needed to ensure personal information to gain trust.
5. **AI must have algorithmic responsibility so that humans can undo the unintended damage**. This is why it is necessary to design AI technology for what was expected and what was not expected.
6. **AI must avoid bias by ensuring adequate and representative research so that an erroneous heuristic cannot be used to discriminate**.

However, Satya Nadella also clarified that humans 'should' prioritise and cultivate a series of characteristics to be able to coexist with AIs, such as:

1. **Empathy** to create relationships with others and perceive their feelings and thoughts because this is something very difficult to replicate in the machines.
2. **Education** that will be necessary to believe and manage innovations that are not understood today. For this, it will be necessary to increase the investment in education that allows developing the knowledge and the necessary skills to implement the new technologies and solve problems that will need much time.
3. **Creativity** because this is the most coveted human skill, that machines will continue to enrich and increase.
4. **Judgment** and **responsibility** to be willing to accept a diagnosis or legal decision made by a machine, but from which we expect that still a human being remains the ultimate responsible for making the final decision.

IV. Machine Learning

Machine Learning, is one of the branches of Artificial Intelligence, and, maybe, one of the most important subfields of AI [10]. It is based on the development of learning techniques to allow machines to learn [11], based on the analysis of data and applying some algorithms for this processing and others for decision making from the data already analysed.

One of the goals of **machine learning** is to facilitate the process of analyzing data intelligently, examples of which are spam filters

1 http://www.slate.com/

[11], **Optical Character Recognition** (OCR) techniques [12], quantification of data relationships [12], recommendation systems [13], [14], computer-aided medical diagnostics [12], search engines [15], **Computer Vision** [16], [17] or **Data Mining** [18]–[20].

A. Taxonomy of Techniques According to the Type of Learning

Machine Learning algorithms can be classified based on the type of **feedback** received. In this subsection, we present a total of six categories: supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning, transduction learning and multitasking learning.

The first is **supervised learning**. These algorithms infer a function by means of labelled training data that they receive as input [12], [21], [22]. An example of them is the algorithms used to classify images by using Computer Vision techniques. These algorithms must be fed with images labelled, for example, as correct or incorrect. In this way, the algorithm is fed so that it learns to classify the images according to the previously labelled images supplied. Therefore, it can be said that the algorithm learns for itself [23]. Some of the algorithms found in this group are some of those belonging to **Artificial Neural Networks**, **Bayesian Networks**, **Support Vectors Machines** and **Decision Trees**.

Another taxonomy is **unsupervised learning** [22], [23]. The difference with the previous one is that the input data is not labelled, and it is the system itself that must infer algorithms or patterns to recognise and label the same type of data. In this group, we find some types of **Artificial Neural Networks**, the **Association Rules** and some **Grouping Algorithms**.

In contrast, in **semi-supervised learning**, a combination of the two previous algorithms is performed [24]. In this way, the system must consider both the tagged elements and the untagged elements.

Another type of learning is **reinforcement learning** [25]. This is based on the algorithm learning based on the world that surrounds it. In this case, the algorithm learns from the answers given by the outside world based on its actions, criticizing them if they were good or bad. Therefore, this is an algorithm based on the classic trial-error learning.

An algorithm similar to the supervised learning is **transduction learning**, presented in [26]. The difference is that, in this case, the function is not created explicitly, but rather the idea is to predict the categories of future examples based on the input examples. In this way, this learning tries to create groups by labelling the elements of the received inputs. An algorithm found in this category is the **Transductive Support Vector Machine** (TSVM) [27], [28].

The last taxonomy is **multitasking learning**. These algorithms use previously learned knowledge in similar problems already experienced to improve the learning of the problem to be solved [29].

B. Taxonomy of Algorithms Used in Machine Learning

There are different types of algorithms to perform the learning that can be grouped according to the type of output of the same. In this subsection, we will talk about **decision trees, association rules, genetic algorithms, Artificial Neural Networks, Deep Learning, Support Vector Machines, clustering algorithms and Bayesian networks.**

The first is the use of **decision trees** [19], [20], of which information on their use in Artificial Intelligence can be found since 1986 in [30]. **Decision trees** are a set of nodes arranged like a binary tree. In the intermediate nodes, there are the necessary conditions to reach the leaves, where these conditions or predicates may imply one or more characteristics of the element. Meanwhile, at the end of each branch, what is known as a leaf, there is the output that the object must return after analyzing the different possibilities obtained in the input and traversing the tree, that is, the leaves contain the decision to be made. Based on this, these algorithms map possible entries and the possibilities of such entries that the object can contain. The operation of the decision trees to classify an element consists of beginning to analyse that element by the root of the tree and, based on whether it fulfils or not the predicate of the node in which it is found, move that element to the left or right child of the current node, repeating the process until it reaches a leaf that returns the result of the tree.

Association rules contain those algorithms that try to discover associations between variables of the same group [19]. In this way, they allow you to discover facts that occur in a certain set of data. A possible example of these is that, if someone publishes a tweet with a photo and the hashtag 'selfie', then the photo was made with his mobile. A real case is the one studied by [31], where they discovered that 90% of the purchases that contained bread and butter also had milk.

Another type of algorithms are **genetic algorithms**. These are nothing more than a heuristic search process that attempts to simulate natural selection following a process of evolution of individuals through random actions. Thus, these algorithms contain methods such as mutation, crossing and selection to find the best solution to a given problem. Some examples of the use of these algorithms are in the field of automated designs of equipment and machinery, in game theory, in the learning of Fuzzy Logic rules, natural language processing and prediction, among others. The creation of genetic algorithms is due to John Holland and his students at the University of Michigan, who developed them between the 60s and 70s and presented them theoretically in [3]. In this way, his goal was not to search for algorithms to solve a specific problem, but to import into computers the phenomenon that occurs in nature by which it adapts to what happens around it [4].

Another case is the **Artificial Neural Networks** (ANN) [22], [32]. ANNs allow automatic learning by simulating the neuron system of the nervous system of animals. Within this system, the neurons collaborate with each other to produce the output, considering that each connection between neurons has a weight, which is adapted according to the experience collected throughout its use. ANNs consist of an input layer, one or more hidden layers and a final output layer, all composed of neurons. Therefore, neural networks can learn and are used for a variety of tasks such as translation of languages or learning patterns.

Deep Learning is a set of algorithms that are based on modelling high-level abstractions composed of multiple non-linear transformations [33], [34]. The use of Deep Learning allows learning data representations. For example, an image can be represented in many ways, but not all of them facilitate the task of face recognition, so, what it does is try to define which is the best representation method to perform this task. Some examples of its use are in **Computer Vision** and in the recognition of speech and music [34]. The algorithms used can be both **supervised learning** algorithms and **unsupervised learning** algorithms [35]. The deep learning algorithms have different internal transformations between the input of the data and the output response, each transformation consisting of different weights and thresholds that can be trained [35]. However, according to [35], there is no definition that exactly says the limits between **Deep Learning**, **Shallow Learning** and **Very Deep Learning**, but they propose that it should be considered Very Deep Learning when it exceeds 10 levels of depth.

Support Vector Machines (SVMs) [36], also used in similar investigations such as **Support Vector Networks** (SVN) [37] and **Support Vector Regression** (SVR) [38], are a set of supervised methods that are used to solve problems of classification and regression, producing non-linear limits by using linear limits and transforming the feature space version [18], [22]. The operation of this type of algorithm is based on the selection of a small number of critical limit cases, known as support vectors, to build the discriminant function to separate existing cases as much as possible [18], [20]. For example, by using SVM you can classify a set of images from a previous training with tagged images to teach the algorithm how to classify. An example of this is the optical character recognition presented in the first article that dealt with them [36].

Clustering algorithms are those that classify similar data based on certain criteria in subgroups, also called clusters [18], [20], [39], [40]. The success of **clustering** is measured based on how useful this clustering is for humans [18]. There are many different algorithms to perform this task because the notion of 'cluster' is not defined very precisely [41]. To do this, it looks for the distance or similarity of the data according to the defined function. These types of algorithms are commonly used in data mining [18], [20], [42], as well as in other cases of statistical analysis [43]. These algorithms are **unsupervised learning** algorithms.

Another type of algorithm is **Bayesian networks** [8], [44]. Bayesian networks are probabilistic models that represent random variables and their unconditional independencies in a directed acyclic graph. That is useful, for example, to represent probabilistic relationships between diseases and symptoms. In this way, we can calculate the probability that a disease is present or not in the body.

V. Computer Vision

Within Artificial Intelligence, one of the fields is **Computer Vision**. It is the field that allows computers to 'learn' to recognise an image and the characteristics of this image. The goal that Computer Vision seeks is that machines can understand the world [45]. Getting the machines to be able to recognise elements by their image and discern what they are, calculate how far they are and their place, and all this in a quick way. However, performing this task presents many challenges.

The first challenge is how to model the objects so that the computer can discern the different types of objects to differentiate what is a person or what is a car, for example. However, these can be viewed from many different positions. Here comes another problem that is to recognise or differentiate if a person is sitting, standing, behind a vehicle or crouched, or even if the car is in front or profile. In addition, there are different light stages throughout the day, not only day and night, but also cloudy or sunny. This creates another problem that is that the same object may be in the same position that the computer recognises but that it is not able to recognise it due to light changes.

Another challenge is the one that arises about how to process the image. Related tasks must be quick and efficient searches of the image should be performed, which must be processed in different ways to normalise it and to avoid some problems previously discussed.

Currently, one of the most important fields in Computer Vision is the detection of people [46]. This is due to the boom to create intelligent systems, mostly for cars, to detect pedestrians to try to avoid or at least to reduce the severity of an accident. Some even try to recognise human movement [47].

A. General Concepts

Every part of the image counts. We must work with the colour, the size of the window that we will use to analyse the image and the intensity of the image. Therefore, the same image can be differentiated in the colour of the light and the sensitivity of the camera. Any variant will make the same take a different image, although they are the same components in it. It must also be borne in mind that the material of the surface of the different elements affects the refraction of light, which also affects the images made. All this is because the cameras we use are based on the human eye because they have three types of sensors, one for each cone existing in the human eye. The cones are photoreceptors found in the retina and there is one for each wavelength: Cones-S (short), Cones-M (middle), Cones-L (long) [46]. Thus, the sensors equivalent to these cones are the red, green and blue sensors (RGB).

Sometimes we also work with other types of images such as near-infrared images (RGB-NIR), which add a non-visible channel in the shortest wavelength of the infrared spectrum. In other cases, thermal images can be used, which reproduce the correlation between temperature and infrared emission of objects. Another type of image is depth images (RGBD) that allow the devices to measure, for example, a Kinect II, the depth of the scene with respect to the sensor.

The process of **Computer Vision** consists of several steps to be performed [16], [46], [48]. In Figure 1, a possible diagram corresponding to these processes is shown. The first step to create a computer vision system is to perform the **features extraction** to build the descriptors of the objects in the images. An example of a descriptor is the colour of the pixel, which would allow defining the object as a region of connected pixels that have the same colour. However, the descriptors usually contain tens or hundreds of different characteristics to be more stable and better. This leads to performance problems and slow scanning of the images.

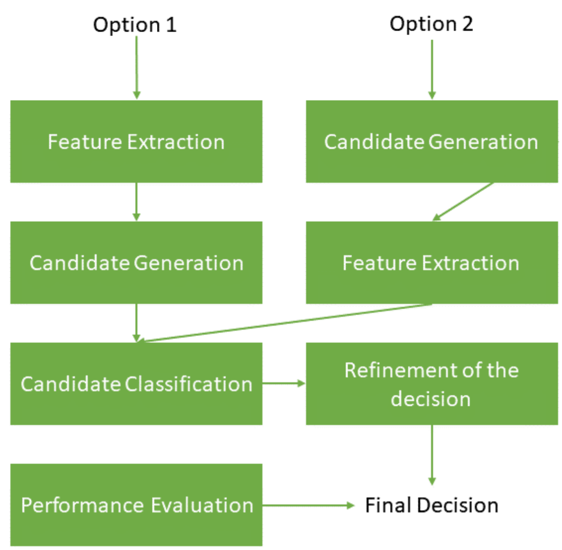


Fig. 1. Cycle diagram of creation of a Computer Vision module.

One of the problems that arise is when there are changes in the intensity of the images to process since a minimum change in it will make the colour different and difficult to be recognised. Therefore, the first step should be to process the images to eliminate the effects of intensity and colour changes and create **descriptors** using a colour invariant to the intensity and the effects of colour. This prevents an image taken with a different intensity or colour changes due to light or camera from being different for the computer vision system's **classifier**.

The second step is the **generation of candidates**. In this step, the regions of interest in the image are defined to solve the problem of the location of the object within it. To do this, the image will be analysed in search of the previously defined objects and will be boxed with a window within the step known as **window generation**. In this generation, there are multiple ways to create windows and how to apply them, such as the **sliding window** and the **sliding window pyramid** (**multiscale**). This helps to solve one of the problems that must be solved: different size that objects can have and the depth of which they can be found.

The next step is the **classification of candidates**. In this phase, we must study and verify that the descriptor is effective for the different instances of the object that we want to detect and verify that it is different from the objects that we do not want to. Thus, we must also create the border that will define what our object is and what is not the object we want. The problems encountered in this phase range from possible variances in the tonalities to the translation, since it can happen that the windows generated in the **phase of window generation** cut objects, transfer them or cut them and detect half of an object and half of a second object as a unique object. Some of the algorithms used in this phase of candidate classification are **Histogram of Oriented Gradients** (HOG) [48], **Local Binary Patterns** (LBP) [49]–[51], **HOG-LBP** [52], **Scale-Invariant Feature Transform** (SIFT) [53] and **Speeded-Up Robust Features** (SURF) [54] as descriptors, and **Logistic Regression** [55], [56], and **Support Vector Machine** (SVM) [36] as classification methods. In addition, one of the learning methods used is **supervised learning**.

In the phase of **refinement of the decision**, the best scoring images must be chosen to avoid duplicates due to the overlap of the created windows and to remain with the best identification of the desired object.

In the **performance evaluation** phase, we must check how the classifier works and assess how accurate it is and the faults it produces, that is, whether they can be ignored or not. One way to do this is by using a confusion matrix. This matrix contains the true positives, true negatives, false positives and false negatives that were identified by the Computer Vision module. This gives us the measurement of the quality of the classifier using four parameters. The first is its accuracy, which helps us to know the distance to the perfect classification. The following is its precision, which indicates the quality of the response of the classifier. The third is the sensitivity that indicates the efficiency in the classification of all the elements that are of the kind that we are looking for. Finally, the specificity indicates the efficiency of the classifier in the classification of the elements that are not of the class for which it was trained.

However, the task of obtaining a good model to recognise a given object is very difficult. It is important to use many and very good images and try to validate it with other different images assuring that the model works correctly. In addition, there is a need to create a model to solve a specific problem, because if a generic model is created, this may have a low accuracy, due to the difficulty to recognise images. To all this, it is important to take into account the different existing problems such as the quality of the images, the different ways of training it and the time it takes to process an image.

B. Work Related to Internet of Things, the Recognition of Faces and Movements

Regarding the work related to the combination of the **Internet of Things** and **Computer Vision**, hardly anything was found. However, there is a lot of related work that focuses on the recognition of people, faces and heads. This work is mainly oriented to vehicles, to prevent outrages or reduce the damage of these, and health, by creating alarms to notify quickly and effectively when older people fall at home.

One of the first works about vehicles is that of [17]. Authors are investigating automatic driving systems that can detect pedestrians or other vehicles and thus avoid collisions, either by notifying the driver or taking control of the vehicle before the accident occurs. This system is based on the recognition of objects that move and on the recognition of pedestrians. This is not only useful for **IIoT** but also in **Smart Cities**, to create a safer city.

In contrast, in [16], they present a framework to detect faces in an extremely fast way while maintaining a good hit rate. In addition, this article presents a new way of representing images that can be computed much faster, a new classifier based on AdaBoost [57] and a new method of combining classifiers using 'waterfall'.

Other authors focus more on the low level, as it is the case of the improvement of the analysis of RGB photos to help the classification of images. These improvements are useful in the classification of images in general, even though they present their idea using the recognition of people [46]. Specifically, they focus on the colours that are opposite, that is, they consider the red-green and yellow-blue tuples and the luminosity.

Other research uses Computer Vision in Robotics, to allow robots to recognise objects or people to be able to perform actions based on it, such as dodging objects, collecting the desired objects among several recognised objects or recognizing a person. For example, in [58] they conducted a survey of the use of Computer Vision in cognitive robots, which are those that are designed based on imitating human cognition, in order that robots can recognise objects, but in such a way that robots do not get saturated with the information of an image, because even if it is small it contains a lot of information. On the other hand, in the 'survey' of [59], they deal with the state of the art of Computer Vision in unmanned aerial vehicles (UAVs), because with their use, autonomy can be extended by improving control during flight and the perception of the environment that surrounds them.

Related to health we have the work of [60]. They did a study on different systems used for the detection of falls to help the elderly. That work collects different 'wearables' device systems that capture environmental data and vision-based systems. According to the authors of the survey, vision-based systems must be better than the other approaches. Likewise, authors classify the field in three types: space-time, inactivity or change of form and movement and 3D position of the head. Among these is the proposal of [61], which proposes a fall detection system based on the combination of moving images with its own vector, so that it takes spatiotemporal data of the movement and its occurrence into account. In contrast, authors of [62] used a detection method based on recognizing the background pixels and using different algorithms to distinguish a fall from other normal movements. The authors of [63] presented a system by which they recognised the person and surrounded them with an oval in order to base their system of recognition of falls based on the change of shape of the oval and comparing this with the previous ovals by means of the histograms that were created. As an example of the third type of Computer Vision applied to falls, this corresponds to the 3D position of the head, which is described in this work [64], and which presents a system that combines 3D cameras with models of the head of the person obtained at times when it is inactive and thus provide in conjunction with the different characteristics of the movements, such as the location and duration of the movement, whether it is a fall or not.

One of the works that integrates vision by computer and IoT is the one carried out in [65], where they integrated a computer vision module in an existing IoT network, called Midgar [66], [67], to allow automation for actions and objects of different intelligence [68], based on what it is recognised in a sequence of photos. For example, if it is recognised that the sequence contains a person then the system is able to send an automatic warning, thus doing the work of a guard or places that need security much easier and automatic.

VI. Fuzzy Logic

The Fuzzy Logic term was introduced in 1965 by Lotfi Asker Zadeh [69], Professor at the University of Berkeley, as a way to deal with the problems of common sense. **Fuzzy Logic** emerged to solve problems that classical logic could not solve because classical logic can only deal with binary values (zeros or ones), while some problems require dealing with more values because these problems use expressions that they are not totally true or false, as shown in Figure 2. Thus, by having more than two values, one can have more states, that decisions can be made with more information and allow them to be represented by an intermediate value that is between absolute truth and total falsehood. Normally these states are known as ***linguistic variables***, which can represent characteristics of type 'size', which in this case could have three possible values or even more, such as 'big', 'medium' or 'small'.

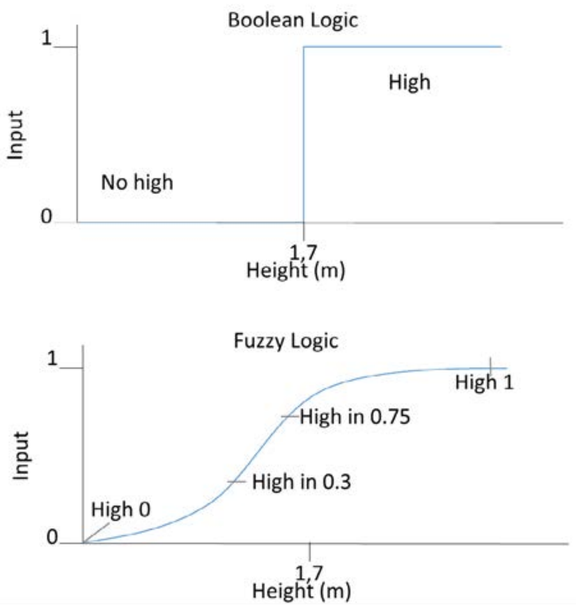


Fig. 2. Comparison between a Boolean Logic system and a Fuzzy Logic system.

Therefore, this approach to the problem is questions of the type: How big is this? The answer to this type of questions depends on individual cognition, therefore, not everyone responds to the same, because, not all of us have the same opinion and perception of reality. It is medium, or even large. An example of this perception could be the place where each person lives, because, if a person lives in a city with a lot of skyscrapers like New York, possibly these people think that a building of 8 plants is small, but in the case of that a second person lives in a town, is possible that the second person thinks that this building is very big.

A. General Concepts

The first thing that emerged in the field of **Fuzzy Logic** was the **fuzzy sets** in [70], presented by Zadeh in 1965. **Fuzzy sets** are set with 'not very well-defined' limits that have a scale between 0 and 1 and serve to measure the **degree of belief**, also known as the **degree of truth** [8], of a certain element in that set. Then, in **Fuzzy Logic**, an element can belong to a set with a certain percentage and to another set with another percentage. Therefore, these serve to represent mathematically the intrinsic imprecision of certain categories of objects and with these, to model the human representation of knowledge, and thus, improve decision systems and Artificial Intelligence. Figure 3 shows an example of the definition of two possible fuzzy sets, the left one defines the temperature and the right one defines the humidity.

Later, **Fuzzy Logic** was defined and presented in [69]. Since then, **Fuzzy Logic** has been linked in many investigations. A common problem where **Fuzzy Logic** is applied is in energy saving. However, there are a lot of examples where **Fuzzy Logic** is applied. The reason for this is that all of these need more than two values to represent the states. In the case of Grant [71], he used **Fuzzy Logic** to propose a diabetic control. In this way, he allows managing with fuzzy and ambiguous data and thus the diabetic control makes decisions as a person would do.

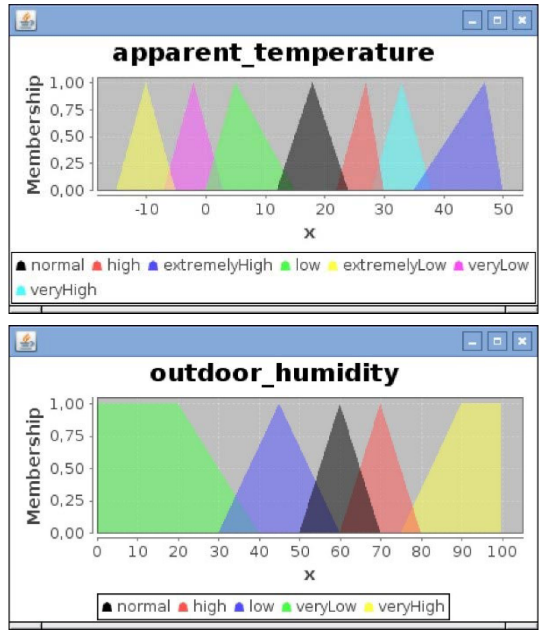


Fig. 3. Fuzzy sets defined to represent the temperature and the humidity.

To make decisions, **Fuzzy Logic** uses controllers called '**adaptative controllers**' [71] or '**expert systems**' [8]. These controllers are based on 'If X and Y then Z' rules to mimic human fuzzy thinking. These rules represent the knowledge that guides these controllers or systems to make the optimal decision. The definition of these rules is complex and requires the use of **linguistic variables** because the human representation of knowledge is fuzzy. The process of taking a **linguistic variable,** for example, in the previous case of the 'size', input, and processing it by a function to say that it is the linguistic variable with the 'small', value, is known as ***fuzzification*** [72]. Thus, a **linguistic variable** is the assignment to a numerical value a word or sentence of the natural language.

Figure 4 shows the life cycle of some data in a Fuzzy Logic system. In this figure, you can see the data entry, with two values, 18 and 85 respectively. These data are 'fuzzified' to transform them through **membership functions** [70], [73]. The difference between Boolean logic and Fuzzy Logic at this point is that, while the Boolean logic can only be 0 or 1 and belong to a set, in Fuzzy Logic this value can belong to several sets since the value can be any between 0 and 1, being a case that the value can belong in 0.3 to one set and in 0.7 to another [70]. The ***membership functions*** are executed both in the 'fuzzification' and in the 'defuzzification'. In this way, by means of **membership functions** it can be established to which or which fuzzy sets a value belongs. The **membership functions** can have different forms, such as triangular, trapezoidal, gamma, sigmoidal, Gaussian, or pseudo-exponential functions.

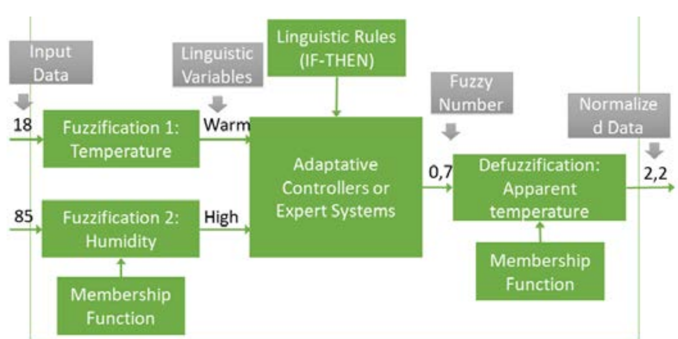


Fig. 4. Cycle diagram through a Fuzzy Logic system.

Thus, from the received values, which are 18 and 85, the corresponding **linguistic variables** are obtained, which in our example are 'small' and 'large' The next step is to apply the linguistic rules, of type IF-THEN, which are found in the controller called '**expert system**', to these variables to obtain the **fuzzy number** or **fuzzy variable**. These **linguistic rules** are responsible for obtaining the fuzzy number from the combination of linguistic variables. They are defined by hand and applying a bit of common sense [73], which implies that a test-error system may have to be made until the rules that best suit our system are found. On the contrary, if the **Fuzzy Logic** system had a training mechanism with which to learn and evolve over time, it would be said that it is an **Adaptative Fuzzy Logic System** [74].

Fuzzy numbers have a value between 0 and 1 and are the result of joining several values. The **fuzzy number** obtained is passed through the 'defuzzifier', phase in which one of the many existing methods is chosen to perform this operation in order to transform the **fuzzy number** into the normalised data that we will use to make the decision.

B. Related Work on the Internet of Things

Below are some examples of using **Fuzzy Logic**. Larios et al. in [75] locate a device avoiding many localisation errors and, therefore, achieve better accuracy. In this case, the authors also achieved savings in the energy consumption of different elements such as the Global Positioning System (GPS). On the other hand, Chamodrakas and Martakos in [76] proposed using **fuzzy set** representation methods to select a network and connect to it in an efficient way, with low power consumption, a good quality of service (QoS) and with a good performance.

In contrast, in the proposal of Bagchi [77], **Fuzzy Logic** was used to maintain streaming playback quality and achieve improvements in energy consumption. However, there are also proposals for the improvement of vehicle systems, such as the cases of [78]–[80]. In the first, the authors proposed improvements in the exchange of information between vehicles and servers in order to save energy. Meanwhile, in the second, the authors proposed a system to create applications through the use of voice. Meanwhile, others focus on energy saving in homes, as in [81], where they use **Fuzzy Logic** to make decisions based on outdoor temperature and humidity and the interior temperature of the house using two IoT networks and to know if it is necessary to turn on or to turn off the heating and/or air conditioning.

VII. Natural Language Processing

Natural Language Processing or ***NLP***, sometimes called **computational linguistics** [8], is a branch of Artificial Intelligence that tries to solve the problem of understanding natural language using machines, or in other words, allowing users to communicate with the machines in a faster and more effective way using natural language [82], [83], as if they were communicating with other humans [84]. El **NLP** requires understanding the structure of sentences, but, in addition, it also requires knowledge of the subject matter and the context in which it is found [8]. This seeks to make applications that can analyse, understand, alter and/or generate natural language according to help with tasks ranging from writing using computers to doing more human tasks through robots. For this reason, **NLP** is a field that seeks to close the existing gap in communication between humans and machines [85].

However, **NLP** is a complicated field, since it requires a linguistic corpus, a model of the domain, knowledge of the domain in which you want to work and linguistic knowledge of the language in which you are going to work, plus, of course, the software tools and necessary hardware [86]. This implies that different disciplines must work together such as computer science for methods and creation of software, linguistics to create processes and linguistic models, mathematics to identify formal models and methods, and neuroscience to explore the mechanisms of the brain and other physical activities [83].

Natural language processing has existed since the 1940s and since then it has contributed significantly in the fields of interaction with the computer, both theoretically and practically [85]. It all started in 1946 when Warren Weaver and Andrew Donal Booth discussed the feasibility of machine translation to break enemy codes in World War II. However, it was not until 1957 when it was found that solving this problem could be much more complex than it seemed [8]. Since then, this field has improved from using dictionary-based translations to what is now known as NLP. An example is NLP in medicine, where it began to be published in the year 1960 [86].

Note that NLP is very important not only in the branch of computer science but for many other areas. For example, in healthcare, it is very important because there is a lot of relevant information in the text that can be processed by a machine in order to help improve patient care and advance in medicine [86]. An example of this importance is the one that occurred in 2012, when the National Library of Medicine of the United States of America sponsored a workshop on April 24 and 25 to review the current state of the art, the challenges, the obstacles and the effective use and best practices of NLP focused on texts in English, both in language in general and in biomedical texts [86].

A. Application Areas

NLP has many application areas [82], [83], [85], [87]. Among these, some of those that can be highlighted are in the **recognition and generation of speech, natural language interfaces, speech management, understanding of history and text generation, machine translation**, and **intelligent writing assistants**.

In **speech recognition,** the goal is to convert the words spoken by a person through a microphone to represent them written on a machine. Thus, **speech understanding** systems try to improve the 'understanding' of the machine about what the user is saying while acting and/ or making decisions based on that context.

On the other hand, the **speech generation** seeks to improve the speech process of the machine by means of looking for the ideal sound representation of the written words that the machine must 'say'.

The research in **natural language interfaces** tries to close the gap between the different existing languages in order to internationalise correctly the user interfaces due to the differences between the different languages. Another example within this category is to facilitate the work with the interfaces through voice commands.

On the other hand, **NLP** is also used to index and classify text, perform summaries, index search engines, Data Mining, make question-answer dialogue systems, extract content, among other similar applications. In these cases, we are talking about **speech management**, **understanding of history**, and **text generation**.

A case similar to the previous one, where the system must have a minimum knowledge of what it reads, is that of **machine translation**, the oldest application of **NLP**. This type of application tries to translate a text from a given language into its equivalent in another required language [8].

Smart Writing Assistants is another **NLP** area. These include various types of applications such as spell-checkers, word-splitting agents, format/separation/text selection wizards, automatic thesauri, and automated document creation or maintenance or document classification environments [88], between some examples. As we can see, all these types of applications are those that try to give a smart help when are working with documents.

B. Linguistic Knowledge Models

To perform **NLP**, different **linguistic knowledge models** can be applied [83], [85], [86], [88], [89]. These models can be classified into five different types: **symbolic or knowledge-based, statisticians**, among which are **stochastic or probabilistic** and **connectionist**, **hybrid approaches** and **Artificial Intelligence.** Each approach has its advantages and disadvantages, which makes each one adapts better or worse depending on the type of **NLP** problem that you wish to solve.

Symbolic or knowledge-based models are based on the explicit representation of facts about language through schemes of representation of knowledge understood, by marking the linguistic corpus used to train it, and associated algorithms [85], [86]. [85], [86]. These types of techniques are the most studied in NLP, however, they have many deficient in comparison with stochastic and connectionist techniques. Symbolic techniques are especially used in cases where the linguistic domain is small or very well defined [85].

Stochastic or probabilistic models use several probabilistic techniques to develop approximate generalised models of the linguistic phenomenon, which are based on real examples of these phenomena. These types of techniques are very effective in those domains where symbolic techniques fail [85], [86], trying to compensate for the difficulty of creating a grammar that fits the whole language and is prepared for reasoning and uncertainty [86].

Connectionist models also use examples of linguistic phenomena to develop generalised models. However, being made from connectionist architectures, these models are less restrictive than stochastics, which is more difficult to develop than connectionists. Being based on stochastic models, connectionist techniques are used in scenarios where symbolic techniques fail [85].

Hybrid approaches use the combination of different architectures and previous models in order to find the best approach. An example of a hybrid approach is that carried out by Miikkulainen in [90], where he combines symbolic techniques with connectionist techniques to process short stories and make summaries and possible answers to questions. Hybrid techniques take the strengths of symbolic, stochastic and connectionist techniques to try to minimise the effort required to construct linguistic models and maximise their flexibility, effectiveness and robustness [85], [86]. This type of approach is the most used in recent years [86].

The methods based on **Artificial Intelligence** are very varied. These can use different types of learning as shown in [88], this work shows some examples where supervised or semi-supervised learning algorithms, based on decision trees, support vector machines, artificial neural networks, Bayesian networks or genetic algorithms were used.

C. Internal Knowledge Levels of Natural Language

NLP is a very extensive field that needs the analysis of different points of view to get to process a speech correctly. Thus, from the linguistic point of view, **NLP** has many aspects that must be studied [91].

- **Phonetics**: studies the sound of human speech from the point of view of its production and perception.
- **Phonology**: study the structure of sound and how it works.
- **Morphology**: studies the internal structure of words to delimit, define and classify words and the formation of new words.
- **Syntax**: studies the structure of the sentence by means of the combinatorial rules and principles of the words and the relationships between them.
- **Semantic**: studies the meaning and denotation of symbols, words, expressions and formal representations.
- **Linguistic variation:** study the different styles of expressing the same meaning and the dialects of a language.
- **The evolution of language:** it studies how the current language emerged and evolved.
- **Pragmatic:** studies the use and communication of language in a specific context.
- **Psycholinguistics**: it is a branch of psychology that studies the production and understanding of language by human beings to see how they acquire and use it.
- **Linguistic acquisition:** studies how language is acquired during childhood.
- **Neurolinguistics:** studies the mechanisms of the brain that facilitate knowledge, compression and language acquisition, both written and spoken.

Based on this, **NLP** is composed of different levels. These vary according to the authors and the time, since some consider four [82] [82] in 1993, another six [92] in 1995, other authors eight [83] in 1996, but which were updated to nine later [85] in 1998, but in general all agree that they can be subdivided mainly into lexical, syntactic, semantic, and pragmatic [85].

Each of these levels offers a different kind of information about the same data. Some tell us the sound of the word, others tell us its meaning, others its gender and number, another level tells us that it means depending on the context, among many other data that we can obtain depending on the way of analysing the text. Thus, architectures that add more levels, get more data and different than architectures with fewer levels. The architectures with more levels add, among other characteristics, the phonetics and knowledge of the world, maintaining the levels of the previous ones.

Next, all the levels found in the literature will be explained, showing in parenthesis what kind of classification they belong to. An example of the different levels can be seen in Figure 5, which shows an example of processing the text 'The wise dragon imitates a duck' that needs levels 4 to 6 because its lexical, syntactic and semantic composition is studied. In case you had to recognise the sentence through a microphone when it was dictated by someone, then you would need levels 1 to 3. If, on the other hand, you wanted to recognise the situation where that sentence was written in order to keep a conversation, then you would also need levels 7 to 9. If, on the other hand, you want to have a conversation with someone, you would need all the levels.

1. **Acoustic or prosody** (8, 9): study the rhythm and intonation of language and how to form their phonemes.
2. **Phonologic** (6, 8, 9): study the relationship between words and the sound that they make and how they work, as well as the formation of morphemes.
3. **Morphologic** (6, 8, 9): studies how words are constructed from their morphemes, which are the smallest unit of the language that has the lexical or grammatical meaning of its own and cannot be divided into smaller units. Thanks to this, you can know and create the gender and number of a word and thus also be able to form them. However, some authors put together the morphologic and the lexical [82].
4. **Lexical** (4, 8, 9): works with words, by means of lexemes and the meaning of words. In the event that the morphologic level does not exist, its constituents are treated at this level as morphemes and their inflectional forms. In addition, it works with the derivation of units of meaning.
5. **Syntactic** (4, 6, 8, 9): works with the combination of words that form a sentence. In this step, each word has a different category such as name, verb and adjective, which makes each word also have specific rules of combination in the given language. He also works on how to form sentences. Figure 5 shows an example of a syntactic analysis of a sentence formed by a subject (first noun phrase) and a predicate (VP) which contains the verb, and the second nominal phrase (NP). Within both nominal phrases are the

analysis of each word that is classified as a determinant (det.), adjective (adj.), noun, and preposition (prep.).

6. **Semantic** (4, 6, 8, 9): deals with the individual meaning of words, the meaning of words according to the sentence and the meaning of the whole sentence. Thus, study the meaning of the sentence independent of the context. For this, he works on how to derive meaning from sentences.
7. **Discourse** (8, 9): deals with the structural roles of the sentence or collections of sentences and how to form dialogues.
8. **Pragmatic** (4, 6, 8, 9): works with the monitoring of the context and the interpretation of sentences within this context. This affects later sentences because it influences the meaning of the interpretation of pronouns.
9. **Knowledge of the world** (6, 9): this includes information about the structure of the world of language users, thus including the beliefs and goals of the other user according to maintain, for example, a conversation.

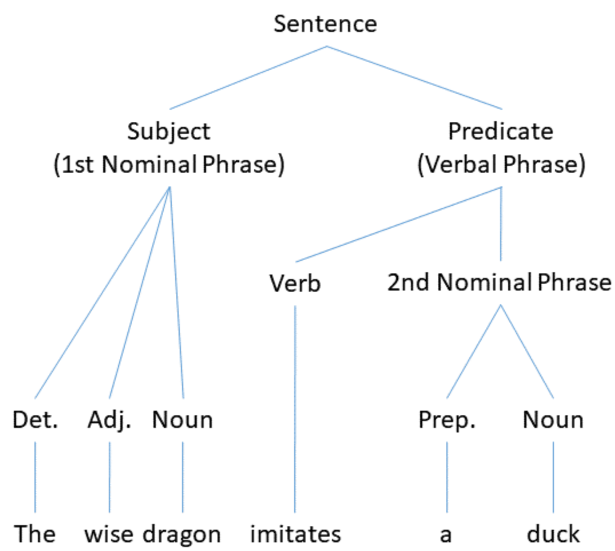


Fig. 5. Syntactic analysis of NLP.

It must have to in mind that the different levels of abstraction necessary among all these, or even a new one, will depend entirely on the domain of the application to be developed or the problem to be solved [85], since it may not be possible to they need all of them or only a few are needed. For example, a text translator will probably need only levels 3-7 or 3-8, for a speech recognition application levels 1-5 will be used, for an intelligent writing assistant levels 3-8 or 3-9 will probably be needed and a dictation system will need levels 1-5.

D. Related Work on NLP

In relation to related works in the field of **natural language processing**, there is a very large variety of these. As shows the 2012 congressional 'survey' [86], there has been much research with **NLP** in the field of medicine to improve the service provided to patients and help doctors in their task.

A first example of NLP in medicine is the work of [93], where they use **NLP** to detect problems of misuse of opioid prescriptions in order to prevent and control their measurements, since problems were detected in traditional clinics that contain documents or incomplete information or in cases in which the patient's information is obscured due to the large number of documents from the clinics. To solve this, they proposed a method that combined **NLP** and manual computer help to review the notes and detect problems with opioid misuse.

Other uses of the **NLP** are those applied to linguistics to help translate or create tools to process certain languages. In this first article [94], the authors developed a framework to promote collaboration between linguists and non-linguists in the use of **NLP** applications for the Arabic language. On the other hand, in [84] they presented a simulation of the parser for Bahasa Indonesia, which is the language of Indonesia, due to the lack of research in this area for this language.

NLP can also be used to give orders, as they do in this article [95], where it is used to handle robots. The authors tried to create a human-machine interface with natural language because they have a greater impact on users, especially when they have no training. Therefore, they presented an interface that received orders in natural language and transformed them into orders for the robot.

An example similar to the robots is the one in this article [96] where the authors used NLP to create SQL statements. With this, the authors sought to bring SQL to non-professional users, so that they could write a sentence in natural language that was transformed to the corresponding query in SQL by using an **expert system** made in Prolog that contains the experience in that domain.

Currently, there are also investigations that deal with the collection of information based on the textual requirements given by a client. In the first example, we have the lips tool, which integrates NLP within the IDE Eclipse to translate the specifications in plain text to formal models of Ecore, UML or SysML, and that can be used in the EMF to create its implementation in Java [97]. On the other hand, and quite similar, there is this second investigation [98], in which they created a prototype called Requirements Analysis and Class Diagram Extraction (RACE), which facilitates the analysis of requirements and the extraction of classes from these by means of NLP and ontologies, using the Apache OpenNLP tool, and thus help the students to see all the steps from the requirements until the creation of the classes. The third example of this type of research is [99]. In this research, the authors proposed a way to obtain the name of the classes and their details from the textual requirements delivered by a client, thus eliminating the ambiguities and possible interpretations that these documents have.

VIII. Conclusion

As we can see in this paper, the Artificial Intelligent is a very vast field. Besides, we have seen some examples of the use of AI with the IoT and how this fusion can create very important and interesting applications. However, we have only touched four of the most important parts: Machine Learning, Computer Vision, Fuzzy Logic, and Natural Language Processing.

These four subfields are very related to each other; mainly, by the algorithms that are used to process the different information or the techniques that are used to teach the necessary information to create the model that different techniques need. Furthermore, these subfields are used to create smarter machines or programs to help us in our daily life, like the Internet of Things is trying to do. For instance, Machine Learning could be used to create a better module with enough intelligence to automate and take decisions, Computer Vision to automate visual scenarios, Fuzzy Logic to give intelligence and take decisions to save money, electricity, and so on, and NLP to improve the understanding between machines and the humankind.

This is why the integration of these fields, the IoT with Artificial Intelligence, could open a new world of opportunities, as the related work has shown. Even though, they have been only a few works on this topic and it is necessary a lot of research in the intersection of these two fields.

References

[1] A. M. Turing, “Computing Machinery and Intelligence,” Mind, vol. 59, pp. 433–460, 1950.

[2] J. Fuegi and J. Francis, “Lovelace & babbage and the creation of the 1843 ‘notes,’” IEEE Ann. Hist. Comput., vol. 25, no. 4, pp. 16–26, Oct. 2003.

[3] J. H. Holland, “Outline for a Logical Theory of Adaptive Systems,” J. ACM, vol. 9, no. 3, pp. 297–314, 1962.

[4] M. Melanie, An introduction to genetic algorithms, 5th ed. MIT press, 1999.

[5] J. McCarthy, “What is Artificial Intelligence?,” 2001. [Online]. Available: http://lidecc.cs.uns.edu.ar/~grs/InteligenciaArtificial/whatisai.pdf. [Accessed: 26-Apr-2016].

[6] J. McCarthy, M. L. Minsky, N. Rochester, and C. E. Shannon, “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence,” Dartmouth, 1955.

[7] J. McCarthy, M. L. Minsky, N. Rochester, and C. E. Shannon, “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence,” AI Mag., vol. 27, no. 4, pp. 12–14, 2006.

[8] J. R. Stuart and N. Peter, Artificial Intelligence A Modern Approach, 2nd ed. New Jersey: Prentice Hall, 2003.

[9] S. Nadella, “The Partnership of the Future,” 2016.

[10] V. García-Díaz, J. Pascual-Espada, C. Pelayo G-Bustelo, and J. M. Cueva-Lovelle, “Towards a Standard-based Domain-specific Platform to Solve Machine Learning-based Problems,” Int. J. Interact. Multimed. Artif. Intell., vol. 3, no. 5, p. 6, 2015.

[11] P. Simon, Too Big to Ignore: The Business Case for Big Data, 1st ed. Wiley, 2013.

[12] M. Wernick, Y. Yang, J. Brankov, G. Yourganov, and S. Strother, “Machine Learning in Medical Imaging,” IEEE Signal Process. Mag., vol. 27, no. 4, pp. 25–38, Jul. 2010.

[13] E. R. Nuñez Valdez, J. M. Cueva Lovelle, O. Sanjuán Martínez, C. E. Montenegro Marín, G. Infante Hernández, and S. Verma, “Social Voting Techniques: A Comparison of the Methods Used for Explicit Feedback in Recommendation Systems,” Int. J. Interact. Multimed. Artif. Intell., vol. 1, no. 4, p. 61, 2011.

[14] C. González García, D. Meana-Llorián, V. García-Díaz, and E. R. Núñez-Valdez, “Social Recommender System: A Recommender System based on Tweets for points of interest,” in Proceedings of the 4th Multidisciplinary International Social Networks Conference on ZZZ - MISNC ’17, 2017, p. 7.

[15] A. McCallum, K. Nigam, J. Rennie, and K. Seymore, “A Machine Learning Approach to Building Domain-Specific Search Engines,” Int. Jt. Conf. Artif. Intell., vol. 16, pp. 662–667, 1999.

[16] P. Viola and M. J. Jones, “Robust Real-Time Face Detection,” Int. J. Comput. Vis., vol. 57, no. 2, pp. 137–154, May 2004.

[17] C. G. Keller, T. Dang, H. Fritz, A. Joos, C. Rabe, and D. M. Gavrila, “Active Pedestrian Safety by Automatic Braking and Evasive Steering,” IEEE Trans. Intell. Transp. Syst., vol. 12, no. 4, pp. 1292–1304, Dec. 2011.

[18] I. H. Witten, E. Frank, and M. A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, 3rd ed. 2011.

[19] T. Menzies and Y. Hu, “Data mining for very busy people,” Computer (Long. Beach. Calif)., vol. 36, no. 11, pp. 22–29, Nov. 2003.

[20] J. Leskovec, A. Rajaraman, and J. D. Ullman, Mining of Massive Datasets. Cambridge University Press, 2014.

[21] M. Mohri, A. Rostamizadeh, and A. Talwalkar, Foundations of machine learning. MIT press, 2012.

[22] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed., vol. 27, no. 2. New York, NY: Springer New York, 2009.

[23] N. Lange, C. M. Bishop, and B. D. Ripley, “Neural Networks for Pattern Recognition,” J. Am. Stat. Assoc., vol. 92, no. 440, p. 1642, Dec. 1997.

[24] X. Zhu, “Semi-Supervised Learning Literature Survey,” Sci. York, pp. 1–59, 2007.

[25] J. A. Hertz, A. S. Krogh, R. G. Palmer, and A. S. Weigend, Introduction to the Theory of Neural Computation, vol. I, no. June. Basic Books, 1991.

[26] A. Gammerman, V. Vovk, and V. Vapnik, “Learning by transduction,” in Fourteenth conference on Uncertainty in artificial intelligence, 1998, no. l, pp. 148–155.

[27] T. Joachims, “Transductive Inference for Text Classification Using Support Vector Machines,” in ICML ’99 Proceedings of the Sixteenth International Conference on Machine Learning, 1999, vol. 99, pp. 200–209.

[28] T. Joachims, “Transductive Support Vector Machines,” 2006, pp. 105–118.

[29] R. Caruana, “Multitask Learning,” Carnegie Mellon University, 1997.

[30] J. R. Quinlan, “Induction of Decision Trees,” Mach. Learn., vol. 1, no. 1, pp. 81–106, 1986.

[31] R. Agrawal, T. Imieliński, and A. Swami, “Mining association rules between sets of items in large databases,” ACM SIGMOD Rec., vol. 22, no. 2, pp. 207–216, Jun. 1993.

[32] S.-C. Wang, “Artificial Neural Network,” in Interdisciplinary Computing in Java Programming, Boston, MA: Springer US, 2003, pp. 81–100.

[33] Y. Bengio, A. Courville, and P. Vincent, “Representation Learning: A Review and New Perspectives,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[34] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[35] J. Schmidhuber, “Deep Learning in neural networks: An overview,” Neural Networks, vol. 61, pp. 85–117, 2015.

[36] B. E. Boser, I. M. Guyon, and V. N. Vapnik, “A training algorithm for optimal margin classifiers,” in Proceedings of the fifth annual workshop on Computational learning theory - COLT ’92, 1992, pp. 144–152.

[37] C. Cortes and V. Vapnik, “Support-Vector Networks,” Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995.

[38] K.-R. Müller, A. J. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen, and V. Vapnik, “Predicting time series with support vector machines,” in Artificial Neural Networks—ICANN’97, vol. 1327, no. X, 1997, pp. 999–1004.

[39] M. S. Aldenderfer and R. K. Blashfield, Cluster Analysis. Sage Publications, Inc., 1984.

[40] R. Duda, P. Hart, and D. Stork, Pattern Classification and Scene Analysis. John Wiley & Sons Inc, 1973.

[41] V. Estivill-Castro, “Why so many clustering algorithms,” ACM SIGKDD Explor. Newsl., vol. 4, no. 1, pp. 65–75, Jun. 2002.

[42] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “The KDD process for extracting useful knowledge from volumes of data,” Commun. ACM, vol. 39, no. 11, pp. 27–34, Nov. 1996.

[43] L. Greiner, “What is Data Analysis and Data Mining?,” 2011. [Online]. Available: http://www.dbta.com/Editorial/Trends-and-Applications/What-is-Data-Analysis-and-Data-Mining-73503.aspx. [Accessed: 26-May-2016].

[44] N. Friedman, D. Geiger, and M. Goldszmidt, “Bayesian Network Classifiers,” Mach. Learn., vol. 29, no. 2/3, pp. 131–163, 1997.

[45] C. Wang, N. Komodakis, and N. Paragios, “Markov Random Field modeling, inference & learning in computer vision & image understanding: A survey,” Comput. Vis. Image Underst., vol. 117, no. 11, pp. 1610–1627, Nov. 2013.

[46] R. Muhammad Anwer, D. Vázquez, and A. M. López, “Opponent Colors for Human Detection,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 6669 LNCS, 2011, pp. 363–370.

[47] T. B. Moeslund and E. Granum, “A Survey of Computer Vision-Based Human Motion Capture,” Comput. Vis. Image Underst., vol. 81, no. 3, pp. 231–268, Mar. 2001.

[48] N. Dalal and B. Triggs, “Histograms of Oriented Gradients for Human Detection,” in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), 2005, vol. 1, pp. 886–893.

[49] Dong-Chen He and Li Wang, “Texture Unit, Texture Spectrum And Texture Analysis,” in 12th Canadian Symposium on Remote Sensing Geoscience and Remote Sensing Symposium, 1989, vol. 5, pp. 2769–2772.

[50] Dong-chen He and Li Wang, “Texture Unit, Texture Spectrum, And Texture Analysis,” IEEE Trans. Geosci. Remote Sens., vol. 28, no. 4, pp. 509–512, Jul. 1990.

[51] L. Wang and D.-C. He, “Texture classification using texture spectrum,” Pattern Recognit., vol. 23, no. 8, pp. 905–910, Jan. 1990.

[52] X. Wang, T. X. Han, and S. Yan, “An HOG-LBP human detector with partial occlusion handling,” in 2009 IEEE 12th International Conference on Computer Vision, 2009, no. Iccv, pp. 32–39.

[53] D. G. Lowe, “Object recognition from local scale-invariant features,” in Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999, vol. 2, no. [8, pp. 1150–1157 vol.2.

[54] R. Funayama, H. Yanagihara, L. Van Gool, T. Tuytelaars, and H. Bay, "Robust interest point detector and descriptor," US 8,165,401 B2, 2006.

[55] S. H. Walker and D. B. Duncan, "Estimation of the Probability of an Event as a Function of Several Independent Variables," Biometrika, vol. 54, no. 1/2, p. 167, Jun. 1967.

[56] D. R. Cox, "The Regression Analysis of Binary Sequences," J. R. Stat. Soc., vol. 20, no. 2, pp. 215–242, 1958.

[57] Y. Freund and R. E. Schapire, "A decision theoretic generalization of on-line learning and an application to boosting," Comput. Syst. Sci., vol. 57, pp. 119–139, 1997.

[58] M. Begum and F. Karray, "Visual Attention for Robotic Cognition: A Survey," IEEE Trans. Auton. Ment. Dev., vol. 3, no. 1, pp. 92–105, Mar. 2011.

[59] Y. Liu and Q. Dai, "A survey of computer vision applied in Aerial robotic Vehicles," in 2010 International Conference on Optics, Photonics and Energy Engineering (OPEE), 2010, vol. 1, no. 201, pp. 277–280.

[60] M. Mubashir, L. Shao, and L. Seed, "A survey on fall detection: Principles and approaches," Neurocomputing, vol. 100, pp. 144–152, Jan. 2013.

[61] H. Foroughi, A. Naseri, A. Saberi, and H. Sadoghi Yazdi, "An eigenspace-based approach for human fall detection using Integrated Time Motion Image and Neural Network," in 2008 9th International Conference on Signal Processing, 2008, pp. 1499–1503.

[62] Zhengming Fu, E. Culurciello, P. Lichtsteiner, and T. Delbruck, "Fall detection using an address-event temporal contrast vision sensor," in 2008 IEEE International Symposium on Circuits and Systems, 2008, pp. 424–427.

[63] H. Foroughi, B. S. Aski, and H. Pourreza, "Intelligent video surveillance for monitoring fall detection of elderly in home environments," in 2008 11th International Conference on Computer and Information Technology, 2008, no. Iccit, pp. 219–224.

[64] B. Jansen and R. Deklerck, "Context aware inactivity recognition for visual fall detection," in 2006 Pervasive Health Conference and Workshops, 2006, pp. 1–4.

[65] C. González García, D. Meana-Llorián, B. C. Pelayo G-Bustelo, J. M. Cueva Lovelle, and N. Garcia-Fernandez, "Midgar: Detection of people through computer vision in the Internet of Things scenarios to improve the security in Smart Cities, Smart Towns, and Smart Homes," Futur. Gener. Comput. Syst., vol. 76, pp. 301–313, Nov. 2017.

[66] C. González García, C. P. García-Bustelo, J. P. Espada, and G. Cueva-Fernandez, "Midgar: Generation of heterogeneous objects interconnecting applications. A Domain Specific Language proposal for Internet of Things scenarios," Comput. Networks, vol. 64, no. C, pp. 143–158, Feb. 2014.

[67] C. González García, J. P. Espada, E. R. N. Valdez, and V. García-Díaz, "Midgar: Domain-Specific Language to Generate Smart Objects for an Internet of Things Platform," in 2014 Eighth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, 2014, pp. 352–357.

[68] C. González García, D. Meana-Llorián, B. C. P. G-Bustelo, and J. M. C. Lovelle, "A review about Smart Objects, Sensors, and Actuators," Int. J. Interact. Multimed. Artif. Intell., vol. 4, no. 3, pp. 7–10, 2017.

[69] L. A. Zadeh, "Fuzzy logic and approximate reasoning," Synthese, vol. 30, no. 3–4, pp. 407–428, 1975.

[70] L. a. Zadeh, "Fuzzy sets," Inf. Control, vol. 8, no. 3, pp. 338–353, Jun. 1965.

[71] P. Grant, "A new approach to diabetic control: Fuzzy logic and insulin pump technology," Med. Eng. Phys., vol. 29, no. 7, pp. 824–827, Sep. 2007.

[72] E. Portmann, A. Andrushevich, R. Kistler, and A. Klapproth, "Prometheus - Fuzzy information retrieval for semantic homes and environments," in 3rd International Conference on Human System Interaction, 2010, pp. 757–762.

[73] O. G. Duarte, "Sistemas de lógica difusa. Fundamentos," Rev. Ing. e Invetigación, no. 42, pp. 22–30, 1999.

[74] W. A. Kwong and K. M. Passino, "Dynamically focused fuzzy learning control," IEEE Trans. Syst. Man Cybern. Part B, vol. 26, no. 1, pp. 53–74, 1996.

[75] D. F. Larios, J. Barbancho, F. J. Molina, and C. León, "LIS: Localization based on an intelligent distributed fuzzy system applied to a WSN," Ad Hoc Networks, vol. 10, no. 3, pp. 604–622, 2012.

[76] I. Chamodrakas and D. Martakos, "A utility-based fuzzy TOPSIS method for energy efficient network selection in heterogeneous wireless networks," Appl. Soft Comput., vol. 11, no. 4, pp. 3734–3743, Jun. 2011.

[77] S. Bagchi, "A fuzzy algorithm for dynamically adaptive multimedia streaming," ACM Trans. Multimed. Comput. Commun. Appl., vol. 7, no. 2, pp. 1–26, Feb. 2011.

[78] G. Cueva-Fernandez, J. Pascual-Espada, V. García-Diaz, and M. Gonzalez-Rodriguez, "Kuruma: The Vehicle Automatic Data Capture for Urban Computing Collaborative Systems," Int. J. Interact. Multimed. Artif. Intell., vol. 2, no. 2, p. 28, 2013.

[79] G. Cueva-Fernandez, J. Pascual Espada, V. García-Díaz, and R. Gonzalez-Crespo, "Fuzzy decision method to improve the information exchange in a vehicle sensor tracking system," Appl. Soft Comput., vol. 35, pp. 1–9, Oct. 2015.

[80] G. Cueva-Fernandez, J. P. Espada, V. García-Díaz, R. G. Crespo, and N. Garcia-Fernandez, "Fuzzy system to adapt web voice interfaces dynamically in a vehicle sensor tracking application definition," Soft Comput., pp. 1–14, May 2015.

[81] D. Meana-Llorián, C. González García, B. C. Pelayo G-Bustelo, J. M. Cueva Lovelle, and N. Garcia-Fernandez, "IoFClime: The fuzzy logic and the Internet of Things to control indoor temperature regarding the outdoor ambient conditions," Futur. Gener. Comput. Syst., vol. 76, pp. 275–284, Nov. 2017.

[82] B. Z. Manaris and W. D. Dominick, "NALIGE: a user interface management system for the development of natural language interfaces," Int. J. Man. Mach. Stud., vol. 38, no. 6, pp. 891–921, Jun. 1993.

[83] B. Z. Manaris and B. M. Slator, "Interactive Natural Language Processing: Building on Success," Computer (Long. Beach. Calif)., vol. 29, no. 7, pp. 28–32, Jul. 1996.

[84] E. M. Sibarani, M. Nadial, E. Panggabean, and S. Meryana, "A Study of Parsing Process on Natural Language Processing in Bahasa Indonesia," in 2013 IEEE 16th International Conference on Computational Science and Engineering, 2013, pp. 309–316.

[85] B. Manaris, "Natural Language Processing: A Human-Computer Interaction Perspective," in Advances in Computers, vol. 47, no. C, 1998, pp. 1–66.

[86] C. Friedman, T. C. Rindflesch, and M. Corn, "Natural language processing: State of the art and prospects for significant progress, a workshop sponsored by the National Library of Medicine," J. Biomed. Inform., vol. 46, no. 5, pp. 765–773, Oct. 2013.

[87] K. W. Church and L. F. Rau, "Commercial applications of natural language processing," Commun. ACM, vol. 38, no. 11, pp. 71–79, 1995.

[88] F. Colace, M. De Santo, L. Greco, and P. Napoletano, "Text classification using a few labeled examples," Comput. Human Behav., vol. 30, pp. 689–697, 2014.

[89] P. Nesi, G. Pantaleo, and G. Sanesi, "A hadoop based platform for natural language processing of web pages and documents," J. Vis. Lang. Comput., vol. 31, pp. 130–138, Dec. 2015.

[90] R. Miikkulainen, Subsymbolic natural language processing: An integrated model of scripts, lexicon, and memory. MIT press, 1993.

[91] A. Akmajian, R. A. Demer, A. K. Farmer, and R. M. Harnish, Linguistics: An introduction to language and communication. MIT press, 2001.

[92] J. Allen, Natural language understanding, 2nd ed. Pearson, 1995.

[93] D. S. Carrell, D. Cronkite, R. E. Palmer, K. Saunders, D. E. Gross, E. T. Masters, T. R. Hylan, and M. Von Korff, "Using natural language processing to identify problem usage of prescription opioids," Int. J. Med. Inform., vol. 84, no. 12, pp. 1057–1064, Dec. 2015.

[94] S. Alansary, M. Nagi, and N. Adly, "A suite of tools for Arabic natural language processing: A UNL approach," in 2013 1st International Conference on Communications, Signal Processing, and their Applications (ICCSPA), 2013, pp. 1–6.

[95] P. Drews and P. Fromm, "A natural language processing approach for mobile service robot control," in Proceedings of the IECON'97 23rd International Conference on Industrial Electronics, Control, and Instrumentation (Cat. No.97CH36066), 1997, vol. 3, pp. 1275–1277.

[96] F. Siasar djahantighi, M. Norouzifard, S. H. Davarpanah, and M. H. Shenassa, "Using natural language processing in order to create SQL queries," in 2008 International Conference on Computer and Communication Engineering, 2008, pp. 600–604.

[97] O. Keszocze, M. Soeken, E. Kuksa, and R. Drechsler, "Lips: An IDE for model driven engineering based on natural language processing," in 2013

1st International Workshop on Natural Language Analysis in Software Engineering (NaturaLiSE), 2013, pp. 31–38.
[98] M. Ibrahim and R. Ahmad, "Class Diagram Extraction from Textual Requirements Using Natural Language Processing (NLP) Techniques," in 2010 Second International Conference on Computer Research and Development, 2010, pp. 200–204.
[99] A. Tripathy and S. K. Rath, "Application of Natural Language Processing in Object Oriented Software Development," in 2014 International Conference on Recent Trends in Information Technology, 2014, pp. 1–7.

Cristian González García

Cristian González García is an assistant professor in the Department of Computer Science of the University of Oviedo (Spain). He is a Technical Engineer in Computer Systems, M.Sc. in Web Engineering, and a Ph.D. in Computers Science graduated from School of Computer Engineering of Oviedo in 2011, 2013, and 2017 (University of Oviedo, Spain). He has been a visitor Ph.D. candidate in the University of Manchester, United Kingdom. Besides, he has been working in different national and regional projects, in projects with private companies, and in the University of Oviedo as a professor. He has published 10 journals articles, 4 conference articles, and 2 book chapters.

Edward Rolando Núñez-Valdez

Edward Rolando Núñez Valdez is an assistant professor in the Department of Computer Science of the University of Oviedo (Spain). He is Ph.D. from the University of Oviedo in Computer Engineering. Master in Software Engineering from the Pontifical University of Salamanca and B.S. in Computer Science from Autonomous University of Santo Domingo. He has participated in several research projects, He has taught computer science at various schools and universities and has worked in software development companies and IT Consulting for many years. He has published several articles in international journals and conferences. His research interests include Web Engineering, Artificial Intelligence, Recommendation Systems, Modeling Software with DSL and MDE.

Vicente García-Díaz

Vicente García Díaz is an associate professor in the Department of Computer Science of the University of Oviedo (Spain). He has a PhD from the University of Oviedo in Computer Engineering. He has been a visiting professor in the University of Manchester (UK), Distrital University Francisco José de Caldas (Colombia) and the University of Lisbon (Portugal) and taught courses on computer science in other international institutions. He has published more than 70 research papers in international conferences and journals. He has graduated in prevention of occupational Risks and is a Certified associate in project management through the Project Management Institute. His research interests include domain specific languages, data science, user of computers for learning and teaching, machine learning and extraction of knowledge from data, extracted from social networks or internet of things-based networks.

B. Cristina Pelayo G-Bustelo

B. Cristina Pelayo García-Bustelo (Santander, 1971) Computer Science Engineer from Oviedo University, Spain (2004). Ph. D. from Oviedo University, Spain (2007). From 2004 he is Associate Professor at the Languages and Computers Systems Area in Oviedo University (Spain). His research interests include the IoT (Internet of Things), Web Engineering, Modelling Software with BPM, DSL and MDA.

Juan Manuel Cueva Lovelle

Juan Manuel Cueva Lovelle (Oviedo, 1958) Mining Engineer from Oviedo Mining Engineers Technical School in 1983 (Oviedo University, Spain). Ph. D. from Madrid Polytechnic University, Spain (1990). From 1985 he is Professor at the Languages and Computers Systems Area in Oviedo University (Spain). ACM and IEEE voting member. His research interests include the IoT (Internet of Things), Language Processors, Human-Computer Interface, Web Engineering, Modelling Software with BPM, DSL and MDA.

A Fuzzy Linguistic RFM Model Applied to Campaign Management

Ramón Alberto Carrasco[1], María Francisca Blasco[1], Jesús García-Madariaga[1], Enrique Herrera-Viedma[2] *

[1] Department of Management and Marketing, Complutense University, Madrid 28223 (Spain)
[2] Department of Computer Science and Artificial Intelligence, University of Granada, Granada 18071 (Spain)

Received 12 November 2017 | Accepted 11 February 2018 | Published 23 March 2018

Abstract

In the literature there are some proposals for integrated schemes for campaign management based on segmentation from the results of the RFM model. RFM is a technique used to analyze customer behavior by means of three variables: Recency, Frequency and Monetary value. It is s very much in use in the business world due to its simplicity of use, implementation and interpretability of its results. However, RFM applications to campaign management present known limitations like the lack of precision because the scores of these variables are expressed by an ordinal scale. In this paper, we propose to link customer segmentation methods with campaign activities in a more effective way incorporating the 2–tuple model both to the RFM calculation process and to its subsequent exploitation by means of segmentation algorithms, specifically, k-means. This yields a greater interpretability of these results and also allows computing these values without loss of information. Therefore, marketers can effectively develop more effective marketing strategy.

Keywords

2-Tuple Model, Campaign Management, Relational Strategy, RFM.

DOI: 10.9781/ijimai.2018.03.003

I. Introduction

Most marketers have difficulty in identifying the right customers to engage in successful campaigns. Customer segmentation is a popular method that is used for selecting appropriate customers for a launch campaign. In order to link customer segmentation methods with campaign activities Chan [1] presents an approach that combines customer targeting and customer segmentation for campaign strategies based on RFM model.

RFM (Recency, Frequency and Monetary) is a model used to analyze customer behavior proposed by Hughes in 1994 [2]. "Recency" represents the length of a time period since the last purchase, while "Frequency" denotes the number of purchases within a specified time period and "Monetary value" means the amount of money spent in this specified time period. The RFM models were developed as a logical step in the evolution of marketing segmentation techniques. When the shotgun approaches (marketing everything to everyone) proved inefficient in terms of returns, the marketing campaigns started separating customers in segments based on socio-demographics attributes [3]. Commonly, RFM methods have been used to measure the Customer Lifetime Value (CLV), i.e., the predicted value a customer is going to generate in his entire lifetime [4]-[6].

With RFM analysis, organizations could discover these most valuable customers easily by observing their past behaviours [7]. In fact, these three variables belong to behavioral variables and can be used to make predictions based on the behavior in the transactional database [8]. Therefore, in a RFM process, the goal is to obtain the customer purchase behavior (the most loyal customers, dormant customers…) from these transactional data to proactively trigger appropriate direct marketing actions (retention, reactivation campaigns…) [9], [10].

* Corresponding author.

E-mail address: viedma@decsai.ugr.es

In recent years, more sophisticated statistical and data-mining techniques have been employed in direct marketing field: chi-squared automatic interaction detection (CHAID), logistic regression, neural network models, etc. Despite the deployment of these methods, marketers continue to employ RFM models. There are several reasons for the popularity of RFM among which the following are worth mentioning [11]: It is easy to use; it can generally be implemented very quickly; and is a method that managers and decision makers can understand.

McCarty and Hastak [11] have compared RFM, CHAID, and logistic regression as analytical methods for direct marketing segmentation, using two different datasets. It turns out that CHAID tends to be superior to RFM when the response rate to a mailing is low and the mailing would be send to a relatively small portion of the database. However, RFM is an acceptable procedure in other circumstances.

RFM approaches present known limitations like the lack of precision. Indeed, the scores of these RFM variables are expressed by an ordinal scale. The most common scale is the set {1,...,5} that refer to the customer contributions to revenue for enterprises. The 5 refers to the most customer contribution to revenue and 1 refers to the least contribution to revenue [12].

On the other hand, the fuzzy linguistic approach is a tool intended for modeling qualitative information in a problem. It is based on the concept of linguistic variable [13] and has been satisfactorily used in many problems [14]-[17]. The 2-tuple fuzzy linguistic approach is a model of information representation that carries out processes of "computing with words" without the loss of information [18] that has been widely used in many business and management applications [19]-[25].

In this paper, we propose to link customer segmentation methods with campaign activities in a more effective way incorporating the 2–tuple model both to the RFM calculation process and to its subsequent exploitation by means of segmentation algorithms, specifically, k-means. This yields a greater interpretability of these results and also allows computing these values without loss of information. Therefore, interpreting these linguistic results, decision makers can effectively identify valuable customers and consequently develop more effective marketing strategy. Additionally, we present an IBM SPSS Modeler [26] implementation of this model. This enables us to be more applicable at the practical level and not remain solely confined to the theoretical one.

The rest of the paper is organized as follows: Section II revises the preliminary concepts, i.e., the integrated scheme of customer segmentation with campaign activities using the RFM model and 2-tuple model. In Section III we propose to modify this integration scheme by incorporating the 2-tuple model in two directions: in the RFM scores computation and the subsequent segmentation algorithm. Additionally we show an implementation and use case of this new model using IBM SPSS Modeler comparing it with the previous one. Finally, we point out some concluding remarks and future work.

II. Preliminaries

In this section we present the basic elements needed to understand our new proposal: an integrated scheme of customer segmentation with campaign activities based on the RFM model and the 2-tuple fuzzy linguistic approach.

A. Integrated Scheme of Customer Segmentation with Campaign Activities

The RFM analytic approach is a common model that identifies customer purchase behavior, i.e., that differentiates important customers from large data by three variables [9]:

- **Recency** (R): The time (in units such as days, months, years…) since the most recent purchase transaction or shopping visit.
- **Frequency** (F): The total number of purchase transactions or shopping visits in the period examined.
- **Monetary** value (M): The total value of the purchases within the period examined.

In order to link customer segmentation methods with campaign activities [1] the following scheme that integrates the RFM model (Fig. 1):

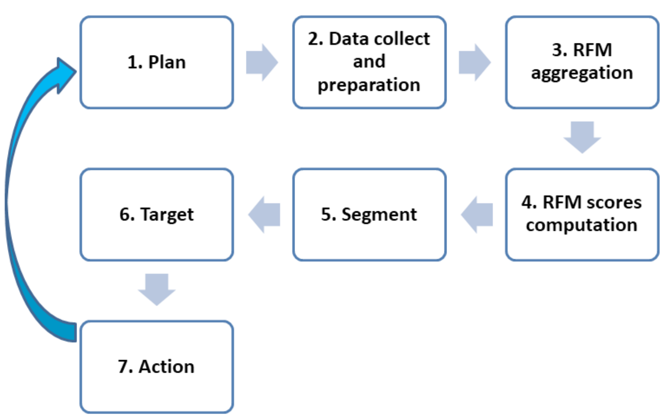


Fig. 1. Integrated scheme of customer segmentation with campaign activities (based on [1]).

1. **Plan**. Conducting a campaign requires first determining series of sequential marketing plans to establish the corresponding relational strategy (Kang 2015), for example, aimed at promoting customer loyalty or preventing customer churn. We can formalize this plan using the following table of campaigns: *Campaigns (CampaignID, RelationalStrategyDES)*, where *CampaignID* contains the campaign code and *RelationalStrategyDES* is the description of the relational strategy of that campaign.
2. **Data collect and preparation**. The second step is collecting related customer information. Once we have chosen the period to analyze, the customers are selected if they have at least purchased during this period. Transactional data on these customers must be retrieved, audited, cleaned, and prepared for subsequent operations. Let *Transactions (CustomerID, Date, Amount)* be the table where these transactional data on purchases are included. The customer code is stored in *CustomerID*. *Date* and *Amount* are the corresponding date and amount spent by the customer in the purchase.
3. **RFM aggregation**. Transactional data is aggregated at a customer level, i. e., on the *CustomerID* attribute. Thus, we obtain the table *CustomerTransactions (CustomerID, Recency, Frequency, Monetary)* with the RFM information summarized for each customer identified by *CustomerID*. *Recency* would be the days since the last purchase of such customer (using a later fixed reference date for all customer purchases). *Frequency* is the number of times the customer has purchased. *Monetary* contains the total amount of those purchases.
4. **RFM scores computation**. Customers are sorted according to the respective RFM measure and are grouped in classes of equal size, typically quintiles. Customers are sorted independently according to each of the individual RFM components and then binned into five groups of 20 per cent. This result is included into the table *CustomerRFM (CustomerID, RecencyScore, FrequencyScore, MonetaryScore, RFMScore)* with *RecencyScore, FrequencyScore, MonetaryScore* ∈ {1,…, 5}. Therefore the RFM measures are transformed into ordinal scores such that the value 1 includes the 20% of customers with the worst values and the 5 the 20% of customers with the best values in the corresponding measure. Especially for the *Recency* attribute, the scale of the derived ordinal score, *RecencyScore*, should be reversed so that larger scores represent the most recent buyers. Sometimes it can be useful to have a unique measure, *RFMScore*, which characterizes together the RFM scores. In order to provide this continuous RFM score, the R, F, and M bins are summed, with appropriate user-defined weights, i.e., w_R, w_F, w_M. The RFM score is the weighted average of its individual components and is calculated as follows:

$$RFMScore = RecencyScore \times w_R + FrequencyScore \times w_F + MonetaryScore \times w_M. \quad (1)$$

5. **Segment**. Once the results of the previous step are validated, the marketers of the enterprise apply this RFM knowledge in order to search the most suitable customer group for each plan campaign. For this, segmentation or clustering techniques are especially useful. Clustering or segmentation is the process of grouping a set of objects into groups of similar objects. In this way, clustering based on RFM scores of the table *CustomerRFM* provides more behavioral knowledge of customers' actual marketing levels than other cluster analyses [27].

K-means is one of the well-known algorithms for clustering [28], [29] of which various modifications have been proposed including fuzzy logic [30]. In this algorithm each cluster is characterized by its center point i.e. centroid. K-means is a partitioning cluster algorithm by grouping *n* vectors (customers in our case) based on attributes into k partitions, where $k < n$, according to some measure, usually Euclidean distance. The name comes from the fact that k clusters are determined and the center of a cluster is the mean of all vectors within this cluster. The algorithm starts with k initial centroids, then assigns vectors to the nearest centroid using Euclidean distance and re-computes the new centroids as means of the assigned data vectors. This process is repeated over and over again until vectors no longer changed clusters between iterations [27], [31]. Thus, using a k-means algorithm, the centroid results of this algorithm are: $v_s = (v_{s1}, v_{s2}, v_{s3})$, with $s = 1..k$, one for each cluster. These centroids are quite interpretable from the point of view of business as explained in the previous stage (5 best values and 1 the worst).

6. **Target**. It is necessary to identify the most profitable groups of customers for each campaign plan. Thus, once the segmentation is concluded and validated, marketers should determine the targeted clusters that can be associated with the subsequent campaign and then get the clients that belong to those groups which are stored in the table *CustomerTarget (CustomerID, CampaignID)*.
7. **Action**. The last step is to implement effective campaign management oriented selected target.

Much of aforementioned approach can be solved with several data science or data mining tools. In Fig. 2 we show an example using IBM SPSS Modeler [26].

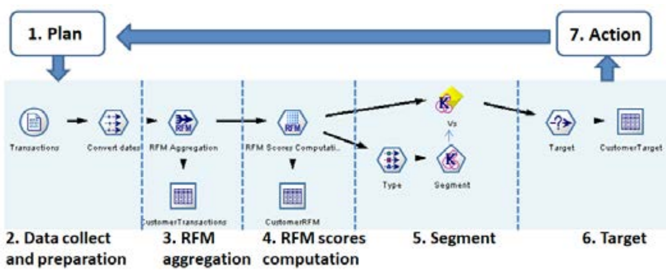


Fig. 2. Integrated scheme of customer segmentation with campaign activities with IBM SPSS Modeler.

Following, we explain each stage of this stream:

1. **Plan**. Based on Ref. [1], in Table I we present the set of campaigns to be carried out.

TABLE I. Campaign Marketing Plan Included in Campaigns Table

CampaignID	*RelationalStrategyDES*
Best	They are the most valuable clients for the company. The relational strategy will be aimed at managing the relationship in order to maintain the value of the clients. For example, we could offer them free services.
New/ Reactivates	They are new clients or ex-clients that have been reactivated in the period analyzed. Welcome gifts, bonus... could be offered associated with the next purchase.
Growing	They are a low-value client which it is necessary to carry out a growth strategy by up-selling or cross-selling.
Churn	The customers who are possible to leave or turn to other competitors. The strategy could be retention by means special discounts…
Worse	They are clients with a minimum degree of relationship with the company so they can be considered ex-clients. The relational strategy will be similar to that of leads, that is, acquisition. Free trial could be offered for example.

2. **Data collect and preparation**. Transactional data of the two last years were retrieved, audited, cleaned, and prepared (casting to date type) for subsequent stages. Inactive customers with no purchases during this period were not included into the *Transactions* table. We have based this example on a file obtained based on data referenced in [32] with 69659 purchase transactions, corresponding to $n = 23570$ distinct customers.
3. **RFM aggregation**. For this step we use the RFM *Aggregate* node (labeled as RFM *Aggregation* in Fig. 2) that simplifies the computation of this stage. We only have to designate the required transaction fields (*CustomerID*, *Date* and *Amount*) and the fixed date to compute *Recency* as the time of difference (days, hours, minutes or seconds) between *Date* and this date (2018-01-01), see Fig. 3.

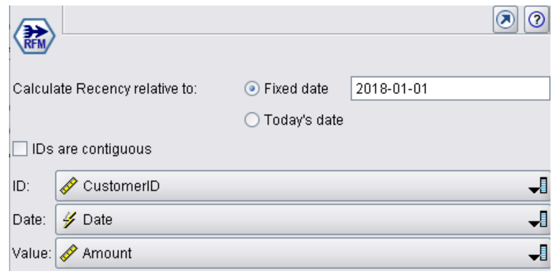


Fig. 3. Detail of the *RFM Aggregate* node settings.

4. **RFM scores computation**. IBM SPSS Modeler also offers a node named *RFM Analysis* (called *RFM scores computation* in Fig. 2) that can directly group the R, F, and M measures into the selected number of quantiles (five in our case). This node also computes the *RFMScore* using the Eq. (1) (with w_R=1/3, w_F=1/3 and w_M= 1/3), see Fig. 4.

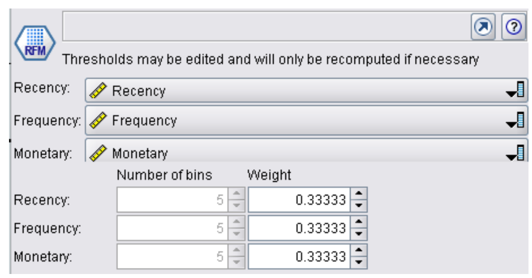


Fig. 4. Detail of the *RFM Analysis* node settings.

Before performing segmentation using scores may address certain plan campaigns, for example, in order to identify the most valuable customers (campaign identified by *Best* in Table I). This could be solved with the RFM model sorting in descending order by the field *RFMScore*. As can be seen in Table II, there are many clients with equal score because when grouping customers in quintiles the procedure results in a total of 5 × 5 × 5 = 125 distinct values as much of *RFMScore*. This lack of precision can be a problem when selecting customers for the different campaigns.

TABLE II.
Detail of the *CustomerRFM* Table Ordered by *RFM Score* Descending (the First 2640 Clients Have the Same *RFMScore* = 5)

	CustomerID	Recency	Frequency	Monetary	Recency Score	Frequency Score	Monetary Score	RFM Score
1	22524	7153	15	185.830	5	5	5	5.000
2	8957	7243	26	345.100	5	5	5	5.000
3	10399	7157	16	228.870	5	5	5	5.000
4	8969	7160	42	630.540	5	5	5	5.000
5	8970	7166	24	400.380	5	5	5	5.000
6	8975	7239	16	182.440	5	5	5	5.000
7	8976	7216	12	175.300	5	5	5	5.000
8	8986	7193	24	277.410	5	5	5	5.000
9	8989	7185	17	255.140	5	5	5	5.000
10	5195	7221	11	137.460	5	5	5	5.000
…								
2640	17142	7167	23	388.310	5	5	5	5.000
2641	12635	7380	13	179.530	4	5	5	4.667
2642	19377	7434	12	253.510	4	5	5	4.667

5. **Segment**. IBM SPSS Modeler provides the appropriate module for k-means. We specify in its settings the number of groups to obtain, this is, i.e. $k = 5$. Before using this node, you must specify with a *Type* node which variables are going to be used in the segmentation, i.e. the RFM scores. We proceed to analyze the different clusters obtained, associating them, if possible, to the campaigns of the specified plan. In Table III, we show the results of this process of clustering and the association to plan.

TABLE III. Results of the K-means Clustering

S	C_s	Recency Score v_{s1}	Frequency Score v_{s2}	Monetary Score v_{s3}	RFM Patter	Campaign ID
cluster-1	2440	3.72	1.97	2.25	R↑F↓M↓	New/ Reactivates
cluster-2	4067	2.74	2.85	3.11	R↕F↕M↕	Growing
cluster-3	7054	4.58	4.56	4.60	R↑F↑M↑	Best
cluster-4	1921	2.11	4.03	4.19	R↓F↑M↑	Churn
cluster-5	8088	1.73	1.23	1.48	R↓F↓M↓	Worse

6. **Target**. Simply, we select those clients that belong to the chosen cluster.
7. **Action**. Based on the description of the strategy that has been made in Table I, the action of each of the proposed campaigns would be implemented for each customer.

B. The 2-Tuple Model

The *2-tuple fuzzy linguistic approach* [18] is a continuous model of information representation that has been used in many business and management applications [19]-[25]. This model carries out processes of "computing with words" without the loss of information which are typical of other fuzzy linguistic approaches. Following, we explain the basic notations and operational laws to understand our proposal:

Let $S = \{s_0, \ldots, s_T\}$ be a linguistic term set with odd cardinality, where the mid-term represents a indifference value and the rest of terms are symmetric with respect to it. We assume that the semantics of labels are given by means of triangular membership functions and consider all terms distributed on a scale on which a total order is defined, i.e. $s_i \leq s_j \Leftrightarrow i < j$. In this fuzzy linguistic context, if a symbolic method aggregating linguistic information obtains a value $b \in [0,T]$, and $b \notin \{0,\ldots,T\}$, then an approximation function is used to express the result in S.

Definition 1 [18]. *Let b be the result of an aggregation of the indexes of a set of labels assessed in a linguistic term set S, i.e. the result of a symbolic aggregation operation, $b \in [0,T]$. Let $i = round(b)$ and $\alpha = b - i$ be two values, such that $i \in [0,T]$ and $\alpha \in [-0.5,0.5)$, then α is called a Symbolic Translation.*

The 2-tuple fuzzy linguistic approach [18] is developed from the concept of symbolic translation by representing the linguistic information by means of 2-tuple (s_i, α_i), $s_i \in S$ and $\alpha_i \in [-0.5, 0.5)$, where s_i represents the information linguistic label, and α_i is a numerical value expressing the value of the translation from the original result b to the closest index label, i, in the linguistic term set S. The value (s_i, α_i) also can be represented as $s_i \pm \alpha_i$ (+ or - depending on the sign of α_i).

This model defines a set of transformation functions between numeric values and 2-tuple:

Definition 2 [18]. *Let $S = \{s_1, \ldots, s_T\}$ be a linguistic term set and $b \in [0,T]$ a value representing the result of a symbolic aggregation operation, then the 2-tuple that expresses the equivalent information to b is obtained with the following function:*

$$\Delta: [0,T] \rightarrow S \times [-0.5,0.5)$$

$$\Delta(b) = (s_i, \alpha), \textit{ with } \begin{cases} s_i, & i = round(b) \\ \alpha = b - i, & \alpha \in [-0.5,0.5). \end{cases} \quad (2)$$

where round(·) is the usual round *operation, s_i has the closest index label to b and α is the value of the symbolic translation.*

For all Δ, there exists Δ^{-1}, defined as:

$$\Delta^{-1}(s_i, \alpha) = i + \alpha. \quad (3)$$

The negation operator is defined as:

$$neg((s_i, \alpha)) = \Delta(T - (\Delta^{-1}(s_i, \alpha))). \quad (4)$$

Information aggregation consists of obtaining a value that summarizes a set of values. Hence, the result of the aggregation of a set of 2-tuples must be a 2-tuple. Using the functions Δ and Δ^{-1} that transform numerical values into linguistic 2-tuples and vice versa without loss of information, any of the existing aggregation operators can be easily extended for dealing with linguistic 2-tuples. Below, we describe the aggregation operators which we use in our model:

Definition 3. *Let $A = \{(l_1, \alpha_1), \ldots, (l_n, \alpha_n)\}$ be a set of linguistic 2-tuple and $W = \{w_1, \ldots, w_n\}$ be their associated weights. The 2-tuple weighted average $\bar{A}^w$ is:*

$$\bar{A}^w[(l_1, \alpha_1), \ldots, (l_n, \alpha_n)] = \Delta\left(\frac{\sum_{i=1}^{n} \beta_i \cdot w_i}{\sum_{i=1}^{n} w_i}\right) \quad (5)$$

Definition 4. *Let $A = \{(l_1, \alpha_1), \ldots, (l_n, \alpha_n)\}$ be a set of linguistic 2-tuple. The 2-tuple average $\bar{A}$ is:*

$$\bar{A}[(l_1, \alpha_1), \ldots, (l_n, \alpha_n)] = \Delta\left(\frac{\sum_{i=1}^{n} \beta_i}{n}\right) \quad (6)$$

III. Applying the 2-Tuple Approach for Campaign Management

As explained in Section II.A, although *RFM* analysis is a very *useful tool for campaign management*, it has its limitations such as its lack of precision in the calculation of scores. This is due to the representation as an ordinal number of these RFM scores (for example, see the Table II where you cannot identify which are really the best customers). In this section, we propose to incorporate the 2-tuple model in order to improve this campaign management. This is possible because by incorporating the 2-tuple model we will get results from the RFM model with more linguistic interpretability and above all with more precision.

The campaign management scheme followed is the same as shown in Fig. 1 and explained in Section II.A but changing stages 3 and 4 as explained in the following two sub sections in which we also show the implementation in SPSS Modeler of the model:

A. RFM Scores Computation

The basic idea is to compute and store the scores included into the output table of this step (*CustomerRFM*), i.e., *RecencyScore, FrequencyScore, MonetaryScore* and *RFMScore* using the 2-tuple model [25].

First, we need to define the symmetric and uniformly distributed domain S using five linguistic labels. These labels have a semantic meaning for these four variables of the RFM model referred to the degree of agreement on the goodness of the variable:

Let $S = \{s_0, \ldots, s_T\}$, $T = 4$: s_0 = *Strongly Disagree* = *SD*, s_1 = *Disagree* = *D*, s_2 = *Neutral* = *N*, s_3 = *Agree* = *A*, and s_4 = *Strongly Agree* = *SA*, with the definition showed in Fig. 5.

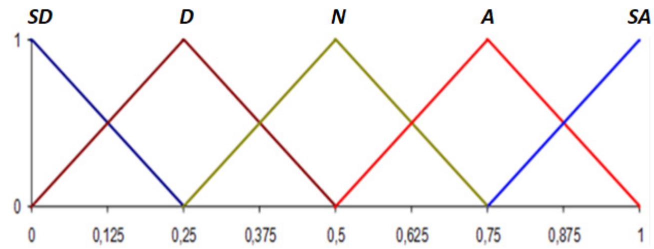


Fig. 5. Definition of the set S.

Therefore, we have the variables to calculate: *RecencyScore, FrequencyScore, MonetaryScore, RFMScore* $\in S \times [-0.5, 0.5)$.

For each customer $i = 1, \ldots, n$, we obtain $A_i = (A_{i1}, A_{i2}, A_{i3})$ with A_{i1} = *RecencyScore*$_i$, A_{i2} =*FrequencyScore*$_i$ and A_{i3} =*MonetaryScore*$_i$. Firstly, customers are sorted in ascending order according to each of the individual RFM components $B_i = (B_{i1}, B_{i2}, B_{i3})$, with B_{i1} = *Recency*$_i$, B_{i2} =*Frequency*$_i$ and B_{i3} =*Monetary*$_i$, stored in *CustomerTransactions* (obtained as explained in phase 3 of Section II.A). Now, we define $rank_{ij} \in \{1, \ldots, n\}$ as the ranking of each client respect to each of these variables:

$$percent\_rank_{ij} = (rank_{ij} - 1) / (n - 1)$$

with $percent\_rank_{ij} \in [0, 1]$, $i = 1, \ldots, n$, $j = 1, \ldots, 3$ and $n > 1$. The final 2-tuple score A_{ij} is obtained as following:

$$A_{ij} = \begin{cases} \Delta((percent\_rank_{ij}), \text{ if } j \neq 1 \\ neg(\Delta(percent\_rank_{ij})), \text{ if } j = 1 \end{cases} \quad (7)$$

where $\Delta(\cdot)$ and $neg(\cdot)$ have been defined in Section II.B (Eq. 2 and 4 respectively). We use the negation function on *Recency* because the larger scores represent the most recent buyers.

The 2-tuple *RFMScore*$_i$, which characterizes together the RFM scores, is calculated for each *i*-customer using the Eq. (5) as follows:

$$RFMScore_i = \bar{A}^w\,[A_{ij}] \quad (8)$$

with the user-defined weights $W = \{w_R, w_F, w_M\}$.

In a previous paper [24] the authors have proposed both a representation data type 2-tuple as the implementation of the functions Δ and Δ^{-1} using IBM SPSS Modeler. Using these tools, the 2-tuple approach proposed in this paper has been implemented. Thus, the stream to solve the example of the Section II.A is showed in the Fig. 6.

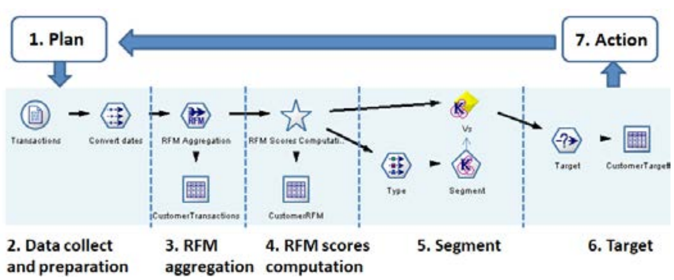


Fig. 6. Integrated scheme of customer segmentation with campaign activities with IBM SPSS Modeler using the 2-tuple model.

This stream is very similar to the presented in the Fig. 2 being its main difference the implementation of the fourth stage. For this purpose we have created a super node (symbolized by ☆) named *RFM Scores Computation* (shown in Fig. 7). A super node is similar to a procedure with inputs (labeled with *From Stream*) and/or outputs values (labeled with *To Stream*). This super node computes de 2-tuple RFM scores (using the Eq. 7 and 8) and it also computes the corresponding Δ-1 function (Eq. 3) on these values to apply conventional numerical operations necessary in the fifth stage.

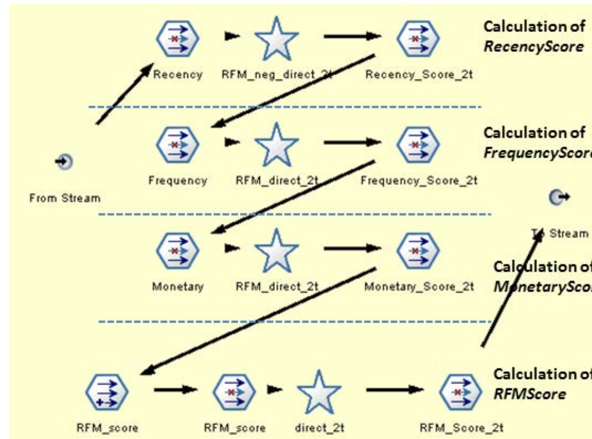


Fig. 7. IBM SPSS super node for RFM scores computation.

We execute the new version of the stream (until 4th phase in Fig. 6) on the same input data and the same user-defined weights (w_R, w_F, w_M) used on the conventional stream (Fig. 2). The selection of the best customers, i.e., the highest Δ^{-1}*(RFMScore)* is presented in the Table IV. Also in this table we show the RFM scores that were obtained according to the conventional process. In the 2-tuple implementation the interpretability of the scores is easier as they are expressed by linguistic labels instead of ordinal numbers. Also the accuracy of such scores is greater, owing the 2-tuple model, allowing a better prioritization (selection) of the best customer in order to identify the most valuable customers (campaign *Best* in Table I).

Table IV.
Results of the Conventional RFM Process Vs 2-Tuple RFM Process Ordered by Vs 2-Tuple RFM Score Descending (Top 20 Customers)

					conventional RFM process				2-tuples RFM process			
	CustomerID	Recency	Frequency	Monetary	Recency Score	Frequency Score	Monetary Score	RFM Score	Recency_Score	Frecuency_Score	Monetary_Score	RFM_Score
1	14048	185	1033	8976.330	5	5	5	5.000	SA-0.001188	SA	SA-0.000042	SA-0.000210
2	7592	186	917	13990.930	5	5	5	5.000	SA-0.003479	SA-0.000042	SA	SA-0.000526
3	3049	186	301	4262.850	5	5	5	5.000	SA-0.003479	SA-0.000339	SA-0.000339	SA-0.000810
4	2908	186	188	2914.870	5	5	5	5.000	SA-0.003479	SA-0.000721	SA-0.000849	SA-0.001230
5	8250	187	161	2582.740	5	5	5	5.000	SA-0.005622	SA-0.001188	SA-0.001018	SA-0.001726
6	3901	187	174	2411.530	5	5	5	5.000	SA-0.005622	SA-0.000976	SA-0.001188	SA-0.001832
7	7983	190	536	6973.070	5	5	5	5.000	SA-0.011880	SA-0.000085	SA-0.000085	SA-0.001854
8	1965	187	122	2010.980	5	5	5	5.000	SA-0.005622	SA-0.002143	SA-0.001739	SA-0.002362
9	16517	191	299	4185.100	5	5	5	5.000	SA-0.013407	SA-0.000382	SA-0.000424	SA-0.002367
10	2484	186	128	1784.360	5	5	5	5.000	SA-0.003479	SA-0.001994	SA-0.002291	SA-0.002440
11	21950	190	204	2982.800	5	5	5	5.000	SA-0.011880	SA-0.000636	SA-0.000806	SA-0.002450
12	6569	192	322	4968.000	5	5	5	5.000	SA-0.015189	SA-0.000297	SA-0.000212	SA-0.002487
13	12352	188	152	2102.920	5	5	5	5.000	SA-0.008040	SA-0.001315	SA-0.001612	SA-0.002547
14	20250	185	107	1592.760	5	5	5	5.000	SA-0.001188	SA-0.003034	SA-0.002970	SA-0.002709
15	10902	188	149	1966.710	5	5	5	5.000	SA-0.008040	SA-0.001379	SA-0.001824	SA-0.002712
16	710	190	130	2502.300	5	5	5	5.000	SA-0.011880	SA-0.001782	SA-0.001061	SA-0.002756
17	398	185	118	1569.330	5	5	5	5.000	SA-0.001188	SA-0.002397	SA-0.003140	SA-0.002773
18	22348	190	195	2356.890	5	5	5	5.000	SA-0.011880	SA-0.000679	SA-0.001315	SA-0.002836
19	8956	189	138	2131.020	5	5	5	5.000	SA-0.010289	SA-0.001506	SA-0.001527	SA-0.002839
20	499	194	328	4378.550	5	5	5	5.000	SA-0.019177	SA-0.000255	SA-0.000297	SA-0.003125

B. Segment

The main problem with the previous approach (4th phase in Section II.A) is the lack of precision in the representation of each individual customer (RFM scores). Consequently the results (centroids) obtained in the next stage are also imprecise. In this section, we propose applying the 2-tuple fuzzy linguistic RFM approach to customer segmentation (using k-means) to obtain more accurate results. On the other hand, it will also increase the interpretability of these centroids as we use linguistic values.

The scores obtained in the previous step (Section III.A) are 2-tuple values: $A_i = (A_{i1}, A_{i2}, A_{i3})$ with A_{i1} = *RecencyScore*$_i$, A_{i2} =*FrequencyScore*$_i$

and A_{i3} =$MonetaryScore_i$.

The objective is to *obtain the centroids* $v_s = (v_{s1}, v_{s2}, v_{s3})$ with $s = 1..k$, one for each cluster. The values of these centroids will be expressed using model 2-tuple model, thus we get a better linguistic interpretability. In order to apply the algorithm, we need to get the distance between customers and these centroids. We propose to use the Euclidean distance d_E following:

$$d_E(i,s) = \sqrt{\begin{array}{l}(\Delta^{-1}(A_{i1}) - \Delta^{-1}(v_{s1}))^2 + \\ (\Delta^{-1}(A_{i2}) - \Delta^{-1}(v_{s2}))^2 + \\ (\Delta^{-1}(A_{i3}) - \Delta^{-1}(v_{s3}))^2\end{array}}$$

for each customer $i = 1..n$ and for each cluster $s = 1..k$.

In each step of the k-mean algorithm, we recalculate the new cluster center *vs* using the Eq. 6:

$$v_s = (\bar{A}\,[A_{r1}], \bar{A}\,[A_{r2}], \bar{A}\,[A_{r3}])$$

with $r = 1..\ c_s$, that symbolizes the *r*-customer such that belongs to the *s*-cluster.

Following the example used previously, we show the result of our 2-tuple model in Table V after executing the proposed algorithm. With our model the linguistic interpretability of the clusters (centroids) is better (see the linguistic labels included in Fig. 5). But the main advantage is that these results are also more accurate as we have already commented. We can see how the distribution of groups with our model is more equitable than the conventional process, where the largest groups are those that contain the worst and best customers (see Table III).

TABLE V. Results of the 2-tuple K-means Clustering

S	C_s	Recency Score v_{s1}	Frequency Score v_{s2}	Monetary Score v_{s3}	RFM Patter	Campaign ID
cluster-1	3139	A-0.015419	N+0.018895	N+0.004251	R↑ F↕M↕	New/ Reactivates
cluster-2	5868	D+0.078401	D+0.092712	D+0.114528	R↓F↓M↓	Growing
cluster-2	5913	A+0.088279	A+0.111631	A+0.109542	R↑F↑M↑	Best
cluster-3	3276	N-0.078436	A-0.065937	A-0.060768	R↕F↑M↑	Churn
cluster-1	5374	D-0.024133	D-0.099303	D-0.115425	R↓F↓M↓	Worse

Therefore, our model could get a more appropriate and effective campaign plan.

IV. Concluding Remarks and Future Work

RFM [2] is a technique widely used a lot more now in marketing due to its simplicity of use, implementation and interpretability of its results. Even its results are better, in a practical level, than other more sophisticated techniques as CHAID and logistic regression in specific circumstances [11]. However, RFM applications to direct marketing present known limitations like the lack of precision.

In this context, we have presented an integrated relational campaign management scheme based on RFM analytic process that incorporates the 2-tuple model in order to obtain a higher precision and an easier linguistic interpretability of the RFM model results and the subsequent segmentation, in order to develop a more effective campaign plan.

Additionally, we have presented an IBM SPSS Modeler implementation of this model. In such a way, our proposal could be widely applied at a practical level on several marketing problems of this type. As an example, we have applied the implemented model on a well-known data set verifying the advantages of the new model regarding the conventional campaign management scheme.

We are currently focusing on the use of this model to several marketing problems, especially in banking industry.

Acknowledgment

This paper has been elaborated with the financing of FEDER funds in the Spanish National research project (TIN2013-40658-P), Spanish Department for Economy and Competitiveness project (TIN2016-75850-R).

References

[1] C. C. H. Chan, “Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer,” *Expert systems with applications*, vol. 34, no. 4, pp. 2754-2762, 2008.

[2] A. M. Hughes, “Strategic database marketing,” Chicago: Probus Publishing Company, 1994.

[3] J. Bernabé-Moreno, A. Tejeda-Lorente, C. Porcel, E. Herrera-Viedma, “A new model to quantify the impact of a topic in a location over time with Social Media,” *Expert Systems with Applications*, vol. 42, no. 7, pp. 3381–3395, 2015.

[4] D. R. Liu, Y. Y. Shih, “Integrating AHP and data mining for product recommendation based on customer lifetime value,” *Information & Management* vol. 42, no. 3, pp. 387-400, 2005.

[5] M. Khajvand, K. Zolfaghar, S. Ashoori, S. Alizadeh, “Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study,” *Procedia Computer Science*, vol. 3, pp. 57–63, 2011.

[6] B. Sohrabi, A. Khanlari, “Customer lifetime value (CLV) measurement based on RFM model,” *Iranian Accounting & Auditing Review*, vol. 14, no. 47, pp. 7–20, 2007.

[7] Y. M. Li, C. H. Lin, C. Y. Lai, “Identifying influential reviewers for word-of-mouth marketing,” *Electronic Commerce Research and Applications*, vol. 9, no. 4, pp. 294–304, 2010.

[8] W. Jo-Ting, L. Shih-Yen, W. Hsin-Hung, “A review of the application of RFM model,” *African Journal of Business Management*, vol. 4, no. 19, pp. 4199–4206, 2010.

[9] K. Tsiptsis, A. Chorianopoulos, “Data mining techniques in CRM: inside customer segmentation,”, John Wiley & Sons, 2011.

[10] W. Buckinx, D. Van den Poel, “Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting,” *European Journal of Operational Research*, vol. 164, no. 1, pp. 252–268, 2005.

[11] J. A. McCarty, M. Hastak, “Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression,” *Journal of business research*, vol. 60, no. 6, pp. 656–662, 2007.

[12] C. H. Cheng, Y. S. Chen, “Classifying the segmentation of customer value via RFM model and RS theory,” *Expert systems with applications*, vol. 36, no. 3, pp. 4176–4184, 2009.

[13] L. A. Zadeh, “The concept of a linguistic variable and its applications to approximate reasoning,” *Inf Sci*, vol. 8, pp. 199–249, 1975.

[14] F. J. Cabrerizo, R. Al-Hmouz, A. Morfeq, A. S. Balamash, M. A. Martínez, E. Herrera-Viedma, “Soft consensus measures in group decision making using unbalanced fuzzy linguistic information,” *Soft Computing*, vol. 21, no. 11, pp. 3037-3050, 2017.

[15] J. A. Morente-Molinera, J. Mezei, C. Carlsson, E. Herrera-Viedma, “Improving supervised learning classification methods using multi-granular linguistic modelling and fuzzy entropy,” *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 5, pp. 1078-1089, 2017.

[16] S. Alonso, I. J. Pérez, F. J. Cabrerizo, E. Herrera-Viedma, “A Linguistic Consensus Model for Web 2.0 Communities,” *Applied Soft Computing*, vol. 13, no. 1, pp. 149-157, 2013.

[17] S. Massanet, J. V. Riera, J. Torrens, E. Herrera-Viedma, “A new linguistic computational model based on discrete fuzzy numbers for computing with words,” *Information Sciences*, vol. 258, pp. 277-290, 2014.

[18] F. Herrera, L. Martínez, “A 2-tuple fuzzy linguistic representation model

for computing with words," *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 6, pp. 746–752, 2000.

[19] A. Cid-López, M. J. Hornos, R. A. Carrasco, E. Herrera-Viedma, "Applying a linguistic multi-criteria decision-making model to the analysis of ICT suppliers' offers," *Expert Systems with Applications*, vol. 57, pp. 127-138, 2016.

[20] A. Cid-López, M. J. Hornos, R. A. Carrasco, E. Herrera-Viedma, F. Chiclana, "Linguistic multi-criteria decision-making model with output variable expressive richness," *Expert Systems with Applications*, vol. 83, pp. 350-362, 2017.

[21] A. Cid-López, M. J. Hornos, R. A. Carrasco, E. Herrera-Viedma, "A hybrid model for decision-making in the Information and Communications Technology sector," *Technological and Economic Development of Economy*, vol. 21, no. 5, pp. 720–737, 2015.

[22] R. A. Carrasco, F. Muñoz-Leiva, J. Sánchez-Fernández, F. J. Liébana-Cabanillas, "A model for the integration of e-financial services questionnaires with SERVQUAL scales under fuzzy linguistic modeling," *Expert Systems with Applications*, vol. 39, no. 14, pp. 11535–11547, 2012.

[23] R. A. Carrasco, J. Sánchez-Fernández, F. Muñoz-Leiva, M. F. Blasco, E. Herrera-Viedma, "Evaluation of the hotels e-services quality under the user's experience," *Soft Computing*, vol. 21, no. 4, pp. 995-1011, 2017.

[24] R. A. Carrasco, M. F. Blasco, E. Herrera-Viedma, "An Implementation of a Linguistic Multi-Criteria Decision Making Model: An Application to Tourism," *Lecture Notes in Artificial Intelligence*, vol. 8536, pp. 232–239, 2014.

[25] R. A. Carrasco, M. F. Blasco, E. Herrera-Viedma, "A 2-tuple fuzzy linguistic RFM model and its implementation," *Procedia Computer Science*, vol. 55, pp. 1340-1347, 2015.

[26] IBM SPSS Modeler. URL (consulted 2018): http://www.ibm.com/software/analytics/spss/products/modeler/

[27] D. Birant, "Data Mining Using RFM Analysis," in *Knowledge Oriented Applications in Data Mining*, K. Funatsu and Hasegawa, K., Eds., Rijeka, Croatia: InTech , pp. 91–108, 2011.

[28] N. Settouti, M. E. A. Bechar, M. A. Chikh, "Statistical comparisons of the top 10 algorithms in data mining for classification task," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. 1, pp. 46-5, 2016.

[29] A. M. Jalil, I. Hafidi, L. Alami, E. Khouribga,"Comparative study of clustering algorithms in text mining context," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 3, no. 7, pp. 42-45, 2016.

[30] B. S. Harish, B. S. Kumar, "Anomaly based Intrusion Detection using Modified Fuzzy Clustering," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. 6, pp. 54-59, 2017.

[31] C. H. Cheng, Y. S. Chen, "Classifying the segmentation of customer value via RFM model and RS theory," *Expert Systems with Applications*, vol. 36, no. 3, pp. 4176-4184, 2009.

[32] P. S. Fader, B. G. Hardie, "Creating an RFM Summary Using Excel," http://www.brucehardie.com/notes/022/RFM_summary_in_Excel.pdf, 2008.

Ramón Alberto Carrasco

Ramón Alberto Carrasco has worked in the financial sector for over 24 years and has extensive experience acting as a leader of several departments related to Analytical Marketing and Business Intelligence. He completed his PhD in data mining at the University of Granada. His main field of research is focused on Data Science applied to Marketing (Data Mining, Multicriteria Decision-Making, Sentiment Analysis, Recommender Systems, etc) on which he has written several textbooks as well as around 60 research papers published in international journals and conferences specialized on these topics. He is currently involved in various (European, national, regional and local) Research, Development and Innovation (R&D&I) Projects. Dr Carrasco is a member of the SCI2S and SECABA Research Groups and is also serving as reviewer for several international journals and conferences. He has been an assistant professor at the University of Granada, Granada, Spain (2004-2013) and currently at the Complutense University, Madrid, Spain (from 2013).

María Francisca Blasco

Professor and Vice-Dean of the Faculty of Commerce and Tourism of Complutense University, Ph.D. in Comunication Sciences by Complutense University. She is chief editor of Journal Cuadernos de Estudios Empresariales; associate editor of International Academy of Management and Bussines Journal (IAMB); Journal and member of the scientific committee of Esic Market Journal. Her work focuses on several issues related to Data Sciences in Marketing and Tourism, Customer Behaviour, Customer Relationship Management, Category Management among others. Her research has been published in journals like Soft Computing, BRQ Business Research Quaterly, Procedia Computer Science, Tourism Review, Universia Business Review, Business Research Quaterly, etc. She works for some research competitive projects like RETO program (Horizon 2020). She is co-authored of books like Marketing del Turismo Cultural, Nuevas Tendencias en Comunicación Estratégica, Principios de Marketing (ESIC Ed.) and Diccionario de Administración de Empresas y Marketing (EcoBook).

Jesús García-Madariaga

Jesús García-Madariaga is currently Associate Professor in Management and Marketing at Complutense University and chair of the University Research team "Markco2". His present research interests encompass Tourism, Customer Relationship Management, Data Mining and Corporate Social Responsibility. His research has been published in refereed international journals, including Journal of Retailing and Consumer Services, Journal of Marketing Channels, Spanish Journal of Marketing ESIC, Journal of Database Marketing and Customer Strategy Management, Journal of International Technology an Information Management, Innovar Journal of Administrative and Social Sciences, Academy of Marketing Science News and Journal of Business Ethics among others. Jesús is member of the Doctoral Committee and Senior Lecturer in graduate courses and Master's programs at the Economics and Business School at Universidad Complutense de Madrid. He is Associate Editor at Associate Editor at International Journal of Internet Marketing and Advertising (IJIMA) and member of the Scientific Committee of some others.

Enrique Herrera-Viedma

Enrique Herrera-Viedma is Professor in Computer Science and A.I in University of Granada and currently the new Vice-President for Research and Knowlegde Transfer. His current research interests include group decision making, consensus models, linguistic preference modeling, aggregation of information, information retrieval, bibliometrics, digital libraries, web quality evaluation, recommender systems, social networks, and social media. In these topics he has published more than 200 papers in ISI journals and coordinated more than 20 research projects. Dr. Herrera-Viedma is member of the gobernment board of the IEEE SMC Society and an Associate Editor of several international journals such as the IEEE Trans. On Syst. Man, and Cyb.: Systems, Knowledge Based Systems, Soft Computing, Fuzzy Optimization and Decision Making, Applied Soft Computing, Journal of Intelligent and Fuzzy Systems, and Information Sciences.

Topic Models and Fusion Methods: a Union to Improve Text Clustering and Cluster Labeling

Mohsen Pourvali[1]*, Salvatore Orlando[1], Hosna Omidvarborna[2]

[1] Università Ca' Foscari Venezia, Venezia (Italy)
[2] Politecnico di Torino, Torino (Italy)

Received 4 June 2018 | Accepted 14 November 2018 | Published 14 December 2018

Abstract

Topic modeling algorithms are statistical methods that aim to discover the topics running through the text documents. Using topic models in machine learning and text mining is popular due to its applicability in inferring the latent topic structure of a corpus. In this paper, we represent an enriching document approach, using state-of-the-art topic models and data fusion methods, to enrich documents of a collection with the aim of improving the quality of text clustering and cluster labeling. We propose a bi-vector space model in which every document of the corpus is represented by two vectors: one is generated based on the fusion-based topic modeling approach, and one simply is the traditional vector model. Our experiments on various datasets show that using a combination of topic modeling and fusion methods to create documents' vectors can significantly improve the quality of the results in clustering the documents.

Keywords

Document Enriching, Document Clustering, Cluster Labeling, Text Mining.

DOI: 10.9781/ijimai.2018.12.007

I. Introduction

While we are overwhelming by the increasing amount of available texts, we simply do not have the human power to read and study them to provide browsing and organizing experience over such the huge amount of texts. To this end, machine learning researchers have developed probabilistic topic modeling, a suite of algorithms that aim to discover and annotate large archives of documents with thematic information. Topic modeling algorithms are statistical methods which are able to find the themes (topics) running through the text documents by analyzing their words. Using topic models in machine learning and text mining is popular due to its applicability in inferring the latent topic structure of a corpus. In document clustering, a topic model could be directly used to map the original high-dimensional representation of documents (word features) to a low dimensional representation (topic features) and then apply a standard clustering algorithm like k-means in the new feature space, or we can consider each topic as a feature of a document, thus documents with highest proportion of same topic (same feature) are located in the same cluster [1]. Specifically, in the classification problem, topic models can be interpreted as the soft (or fuzzy) classification of the collection of documents into latent classes, which means a document does not belong fully to one class but it has different degrees of membership in several classes. Besides, the results of topic models could be used to produce a hard classifier in which a document can only have one and only one category.

In this work, we present a novel approach to improve the quality of clustering using topic models [2] and fusion methods [3]. The core idea of our approach is to enrich the vectors of the documents in order to improve the quality of clustering. To this end, we apply a statistical approach to discover and annotate a corpus with thematic information represented in form of different proportions over different topics for each document. Our approach is an unsupervised method and the topics, used for enriching, are produced by the unsupervised learning method. Further, final enriched vectors, representing documents, are clustered through kmeans clustering and produced classes are hard classification classes extracted from the Latent Dirichlet Allocation (LDA) results.

We first run topic modeling several times with different parameters over the collection, we then specify a set of topics in each iteration as the special topics for each document. Finally, we combine all the special topics in each iteration to generate a single topic for every document. These generated topics are indeed the vectors which are used later in the clustering of the collection. Furthermore, we use these topics to generate labels for each cluster.

* Corresponding author.

E-mail address: pourvali.mohsen@gmail.com

II. Related Works

In this section, we briefly summarize related works on text representation models for vector-word based text clustering.

The basic text representation model, i.e., Bag of Words (BOW) model, is widely used for text clustering and classification. In this model each term is weighted by various schemes such as TF, TF-IDF [4], and its variants [5]. Using BOW representation is popular but in the short text it generates a sparse vector for the document.

To overcome data sparseness, there are several works that exploit external knowledge (e.g., Wikipedia, WordNet, etc) to extend content of the documents. Banerjee et al. [6] uses Wikipedia knowledge base to enrich document representation vector with additional features, and Hotho et al. [7] uses WordNet knowledge base to enrich the representation vectors. There are some works that use feature selection approaches to reduce the high dimensionality. Revanasiddappa et al. [8] proposed a feature selection method based on Intuitionistic Fuzzy Entropy for text categorization.

Lu et al. in [1] investigated performance of two probabilistic topic models Probabilistic Latent Semantic Analysis (PLSA) and LDA in document clustering. Authors used the topic models to generate a number1 of topics which are treated as specific features of documents. Therefore, for clustering, documents that have highest probability in a same feature (same topic) are clustered into the same cluster. In a similar way, Yau et al. [9] aims to elaborate on the ability of further other topic modeling algorithms Correlated Topic Model (CTM), Hierarchical LDA, and Hierarchical Dirichlet Process (HDP) to cluster documents. We highlight two main problems here: first, we do not know the exact number of topics running through the corpus, besides, because of frequency-based nature of topic models, we cannot claim the topic with the highest probability for a document is the main topic by which the documents must be clustered. These two problems are considered as our hypothesis in dealing with topics running through the corpus.

The supervised approaches in text classification domain [10, 11, 12] exploit topic models to enrich document representation. Vo and Ock [11] used the LDA model for topic analysis but presented new methods for enhancing features by combining external texts modeled from various types of universal datasets. In other studies [10, 12] their authors propose an approach to learn word vectors together with topics.

There are also some neural embedding methods word2vec [13] and doc2vec [14] that produce vector representations of words and documents by processing a corpus. Word2vec is a two-layer network with the main assumption that words with similar contexts have similar meaning. According to this assumption, word2vec describes semantic correlations between words in the corpus. Doc2vec (or Paragraph Vectors) is an extension of word2vec that requires labels to associate arbitrary documents with the labels. Indeed, Doc2vec learns to correlate labels and words rather than words with other words. These algorithms prefer to describe real semantic information embedded in words, sentences and documents rather than statistical relationships of the term occurrences.

In this paper, we propose a method to enrich document representation vectors to be used in partitional text clustering and cluster labeling. Our method is an unsupervised approach, needless of any external knowledge, with the aim of overcoming the two main problems about sparse vector and traditional LDA representation explained above. Since the main goal of our method is to enrich document vectors according to the statistical relationships of the term occurrences rather than real semantic information embedded in terms, we compared our results with two strong baselines in this domain. To this end, we use two unsupervised baselines: *first baseline*, i.e., BOW text representation with TF-IDF terms weighting, and *second baseline*, i.e., unsupervised usage of LDA in document representation [1, 9].

To the best of our knowledge, our work is the first to suggest a topic modeling solution to improve the quality of clustering and to perform cluster labeling based on the fusion methods.

III. Preliminary

Before we explain the main approach proposed in this paper, we briefly describe topic models and explain LDA as the topic model that we apply in our approach. We also explain two well-known data fusion methods which are used in this paper.

A. Topic Models

Topic models are based on the idea that documents are created by a mixture of topics, where a topic is a probability distribution over words. Specifically, a topic model is a statistical model by which we can create all the documents of a collection. Assume that we want to fill up every document of a corpus with the words, topic model says each document contains multiple topics and exhibits the topics in different proportion. Thus, for each document, there is a distribution over topics that according to this distribution, a topic is chosen for every word of that document, and then from that topic (i.e. distribution over vocabulary) a word is drawn [2].

B. Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is a topic model widely used in the information retrieval field. Specifically, LDA is a probabilistic model that says each document of a corpus is generated by a distribution over topics, and each topic is characterized by a distribution over words. The process of generating a document defines a joint probability distribution over both observed (i.e. words of corpus) and hidden (i.e. topics) random variables. The data analysis is performed by using that joint distribution to compute the conditional distribution of the hidden variables given the observed variables. Formally, LDA is described as follows:

$$p(\beta_{1:k}, \theta_{1:D}, z_{1:D}, w_{1:D}) = \prod_{i=1}^{k} p(\beta_i) \prod_{d=1}^{D} p(\theta_d) \left(\prod_{n=1}^{N} p(z_{d,n} | \theta_d) p(w_{d,n} | \beta_{1:K}, z_{d,n}) \right) \quad (1)$$

where $\beta_{1:k}$ are topics where each β_k is a distribution over words of the corpus (i.e. vocabulary), $\theta_{1:D}$ are topic proportions for the dth document, z_d are the topic assignments for the dth document where $z_{d,n}$ is the topic assignment for the nth word in document d, which specifies the topic that nth word in d belongs to, and w_d are the observed words for document d where $w_{d,n}$ is the nth word in document d.

C. Fusion Methods

We now introduce two baseline state-of-the-art data fusion methods, frequently used for various information retrieval tasks, namely the CombSUM and CombMNZ fusion methods [3].

Suppose there are n ranked lists which are created by n different systems over a collection of items D. Each system S_i provides a ranked list of items $L_i = < d_{i1}, d_{i2}, \dots, d_{im} >$ and a relevance score $s_i(d_{ij})$ is assigned to each of the items in the list. Data fusion techniques use some algorithms to merge these n ranked lists into one [3].

CombSUM uses the following equation:

$$g(d) = \sum_{i=1}^{n} s_i(d) \quad (2)$$

If d does not appear in any L_i, a default score (e.g., 0) is assigned to it. According to the global score $g(d)$ the items can be ranked as a new list.

Another method CombMNZ uses the equation:

$$g(d) = m \times \sum_{i=1}^{n} s_i(d) \quad (3)$$

where m is the number of lists in which item d appears.

The linear combination (i.e. general form of CombSUM) uses the equation:

$$g(d) = \sum_{i=1}^{n} w_i \times s_i(d) \quad (4)$$

where w_i is the weight assigned to system S_i.

IV. Our Method

To create an enriched vectorial representation for documents of a corpus, we propose an unsupervised technique, called Fusion- and Topic-based Enriching (FT-Enrich). Let $\mathbb{D} = \{d_1, d_2, \dots, d_n\}$ be the collection of documents that we wish to be clustered, we run LDA algorithm several times over the collection, every time with different specified number of topics. We used LDA because we want to manually specify and change the number of topics. The intuition behind using different topics in each iteration is to bring in variety of topics being discussed in documents with an ensemble approach. We start with a number of topics close to the number of clusters, for example, assuming K is the number of clusters we wish to have, the beginning number for topics is $I = K \pm \kappa$ where κ is a small integer[1]. The reason of starting with I is to emphasize the topics in an iteration which has a number of topics close to the number of clusters. Finally, for every document d_i of $\mathbb{D}$ there is a set $\mathbb{B} = \{\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_m\}$ where $\mathcal{B}_i = \{\beta_1, \beta_2, \dots, \beta_s\}$ shows s topics belonging to iteration i, and m indicates the number of iterations. At first $s = I$ which is increased by one in each iteration. Number of iterations depends on the maximum number of topics, i.e., bigger than number of clusters, involving in determining of special topics. It could be an expectation of different topics among the corpus. The clustering results in our experiments are obtained by 25 iterations. Therefore, for clustering a corpus into 4 clusters, sequence of the topics number for 25 iterations with $I = 4 - 1$ is $3,4,5, \dots, 27$.

In every iteration, for each document, we generate a set of topics, namely, special topics, which are selected from the topics within iteration i. To generate these topics, we construct a graph G_i comprising the documents of $\mathbb{D}$ and the topics generated in iteration i. Fig. 1 shows three examples of graph G in different iterations. Every circular node corresponds to a document of the collection, and the square nodes correspond to the topics generated in that iteration. The connection $\mathcal{X}_{jr}$ between a circular node d_i and a square node β_r indicates the proportion of the corresponding topic in the document. Therefore, $\mathbb{P}_i = \{\theta_{1:s}, \theta_{2:s}, \dots, \theta_{n:s}\}$ indicates topic proportions of the documents in iteration i where $\theta_j = \{\mathcal{X}_{j1}, \mathcal{X}_{j2}, \dots, \mathcal{X}_{js}\}$ shows topic proportions for document j in graph G_i where $\sum_{l=1}^{s} \mathcal{X}_{jl} = 1$. Therefore, the elements of special topics for document d_j, within iteration i, include:

- the topic with highest proportion of $\mathcal{X}_{jx}$ for document d_j,
- the topic by which document d_j finds its best couple,
- the topic by which d_j is selected as the best couple for a document.

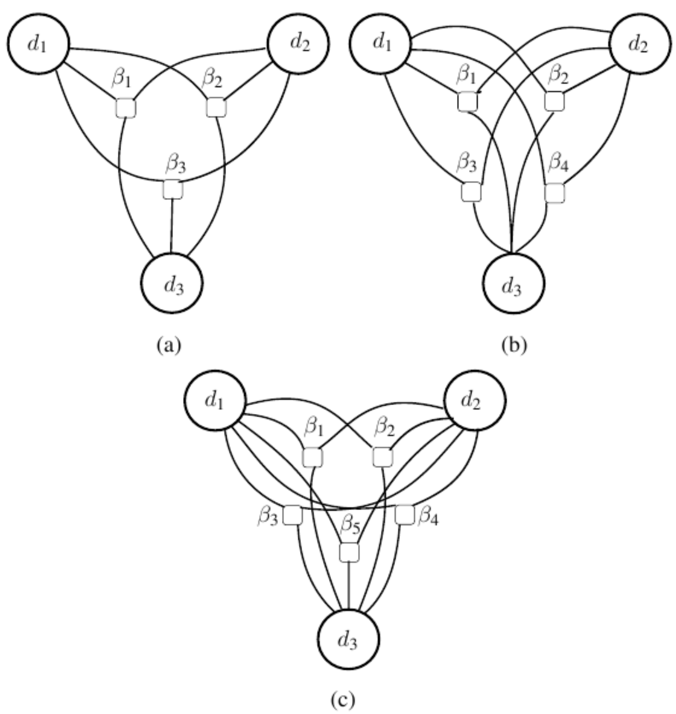


Fig. 1. Three typical graphs of G for $\mathbb{D} = \{d_1, d_2, d_3\}$ in three different iterations with: (a) three topics; (b) four topics; (c) five topics.

1 In our experiments $\kappa = 1$

Given the topics of iteration ith, the best couple for document d_j is a document d_k for which the following equation returns the highest value:

$$Couple(d_j, d_k | \mathcal{B}_i) = \arg \max_{\beta_l \in \mathcal{B}_i} \left(\frac{\mathcal{X}_{jl} \times \mathcal{X}_{kl}}{\mathcal{X}_{jl} - \mathcal{X}_{kl}} \right) \quad (5)$$

where the denominator in case of $\mathcal{X}_{jl} = \mathcal{X}_{kl}$ equals 0.1. Specifically, Equation (5) is to find documents which are similar together in a specific topic, considering their proportion in the topic. Therefore, for each document in a specific iteration, there is a special topics set $ST_i(d_j)$ where $|ST_i| \leq |\mathcal{B}_i|$. We take into account the effect of special topics for each document by combining elements of $ST_i(d_i)$. Our goal is to generate a representing vector for each document to be used in clustering where this vector is a combination of some special topics. We use the data fusion method CombSUM in two phases to generate a single topic (vector) for each document in the corpus.

In the first phase, all the topics within $ST_i(d_i)$ are combined to generate a single vector $\mathcal{V}_{ij}$ for each document d_j in iteration i. Formally, let $S_{\beta}^{norm}(b|\mathcal{B}_i)$ denotes b's **normalized** score given in distribution (topic) β, the general form of CombSUM fusion method then simply sums over the normalized b'scores given by various topics in $ST_i(d_j)$.

$$CombSum\left(b \middle| ST_i(d_j)\right) = \sum_{\beta \in ST_i} \mathcal{X}_{j\beta} \times S_{\beta}^{norm}(b|\mathcal{B}_i) \quad (6)$$

where $\mathcal{X}_{j\beta}$ is the proportion of document j in topic β.

In the second phase, all the single vectors $\mathcal{V}_{ij}$ generated in m iterations are combined to generate a **unique** vector V_j for document j. Formally, given $AV(d_j) = \{\mathcal{V}_{1j}, \mathcal{V}_{2j}, \dots, \mathcal{V}_{mj}\}$, let $S_{\mathcal{V}}^{norm}(b|\mathbb{B})$ denotes b's normalized score given in vector $\mathcal{V}$, therefore, the CombSUM fusion method sums over the normalized b's scores given by various vectors in $AV(d_j)$.

$$CombSum\left(b \middle| AV(d_j)\right) = \sum_{\mathcal{V} \in AV} S_{\mathcal{V}}^{norm}(b|\mathbb{B}) \quad (7)$$

Finally, a trade-off between V_j and traditional vector, i.e., a vector generated based on TF-IDF for document j, are used to generate the final vector. Which is the representing vector for jth document in clustering. Formally:

$$FV_j^{norm} = \alpha \times V_j^{norm} + (1 - \alpha) \times vt_j \quad (8)$$

where vt_j indicates traditional vector for jth document, and $\alpha \in [0,1]$.

V. Cluster Labelling

To label a cluster $C = \{FV_1, FV_2, \dots, FV_c\}$, we use CombMNZ data fusion method which provides good results in combining several ranked lists [3][15]. First, we create $\mathcal{L} = \{L_1, L_2, \dots, L_c\}$ where L_j is a list of terms corresponding to the vector FV_j within C, we then rank/sort the terms of L_j based on the scores/probabilities obtained for its corresponding vector FV_j in Equation (8). Therefore, $\mathcal{L}$ is updated with the new ranked lists. We then create candidate labels ${L^{[M]}}_j(C)$ which are Top-M terms within list L_j. Therefore, let $\mathcal{L}^{[M]}(C) = \bigcup_{L \in \mathcal{L}} L^{[M]}(C)$ denotes the overall candidate-labels pool which are generated based on the union of all Top-M scored labels selected from $L \in \mathcal{L}$ for cluster C. The CombMNZ is to boost label l based on the number of times that l appears in various lists. Formally:

$$CombMNZ\left(l \middle| \mathcal{L}^{[M]}(C)\right) = \#\{l \in L^{[M]}(C)\} \times \sum_{L \in \mathcal{L}} S_L^{norm}(l|C) \quad (9)$$

Finally, Top-N, i.e., $|N| < |M|$, labels of the combination result are selected as the labels of the cluster C.

VI. Experimental Setup

The principal idea of the experiments is to show the efficacy of an ensemble approach of topic modeling on clustering results through a manually predefined categorization of the corpus.

A. Datasets

We explore the utility of using representation vectors of documents generated by our method in addition to label the clusters. To this end, we used three different datasets:

Classic4: This dataset is often used as a benchmark for clustering and co-clustering[2]. It consists of 7095 documents classified into four classes denoted MED, CISI, CRAN and CACM. For our experiments, we extract randomly 500 documents from each class.

BBC NEWS: This dataset consists of 2225 documents from the BBC news website corresponding to stories in five topical areas, which are named Business, Entertainment, Politics, Sport and Tech, from 2004-2005 [18].

20NG: 20 News Group[3] (20NG) is a collection of documents manually classified into 20 different categories that each one contains about 1000 documents.

B. Preprocessing

Preprocessing is an essential step in text mining. The first classical preprocessing regards stop words removal and lower case conversion. In addition, we used L2-norm to normalize the topics/vectors generated by MALLET. The normalized vector of $v = (v_1, v_2, .., v_n)$ is a vector with the same direction but with length one. It is denoted by $\hat{v} = \frac{v}{|v|}$, where $|v| = \sqrt{v_1^2 + v_2^2 + .. + v_n^2}$.

C. Vectors Similarity Measure

For evaluating similarity of two represented vectors, we used comparative traditional measure Cosine Similarity that measures the cosine of the angle between two none zero vectors of an inner product space. Given two vectors of attributes, A and B, the cosine similarity, $\cos(\theta)$, is represented as follows:

$$Similarity = \cos(\theta) = \frac{A.B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \quad (10)$$

Cosine similarity is a judgment of orientation and not magnitude of two vectors commonly used with text data represented by word counts: its results range from -1 meaning exactly opposite, to 1 meaning exactly the same, with 0 indicating orthogonality, and in-between values indicating intermediate similarity or dissimilarity. In Information Retrieval (IR), since the term frequencies (TF-IDF weights) cannot be negative, the cosine similarity of two represented vectors of two documents will range from 0 to 1.

D. Clustering Evaluation Measures

We used two external criteria Purity and F1-measure for evaluating the clustering results.

Purity: The purity is a simple and transparent evaluation measure which is related to the entropy concept [19]. To compute the purity criterion, each cluster C is assigned to its majority class. Then we consider the percentage of correctly assigned documents, given the set of documents L_i in the majority class:

$$Precision(C, L_i) = \frac{|C \cap L_i|}{|C|} \quad (11)$$

The final purity of the overall clustering is defined as follows:

$$Purity(\mathbb{C}, \mathbb{L}) = \sum_{C_j \in \mathbb{C}} \frac{|C_j|}{N} arg \max_{L_i \in \mathbb{L}} Precision(C_j, L_i) \quad (12)$$

where N is the number of all documents, $\mathbb{C} = \{C_1, C_2, \dots, C_k\}$ is the set of clusters and $\mathbb{L} = \{L_1, L_2, \dots, L_c\}$ is the set of classes.

F1-measure: The F1-measure is defined as a harmonic mean of precision P and recall R [20]. Formally, F1-measure is defined as follows:

$$F_1 = \frac{2PR}{P+R} \quad (13)$$

where P (Precision) is defined in Equation (11), and R (Recall) is formally defined as follows:

$$Recall(C, L_i) = \frac{|C \cap L_i|}{|L_i|} \quad (14)$$

where L_i is the majority class.

E. Labelling Evaluation Measures

For evaluating the quality of cluster labeling, we use the frameworks represented in [21]. Therefore, for each given cluster, its ground truth labels where obtained by manual (human) labeling and are used for the evaluation.

We use **Match@N** (Match at top N results) and **MRR@N** (Mean Reciprocal Rank) measures proposed in [21] to evaluate the quality of the labels. They consider the categories of Open Directory Project (ODP) as the correct labels and then evaluate a ranked list of proposed labels by using the following criteria:

- **Match@N**: It is a binary indicator, and returns 1 if the top N proposed labels contain at least one correct label. Otherwise it returns zero.
- **MRR@N**: It returns the inverse of the rank of the first correct label in the top-N list. Otherwise it returns zero.

A proposed label for a given cluster is considered correct if it is identical, an inflection, or a WordNet synonym of the cluster's correct label [16].

VII. Experimental Results

A. Evaluating Results of Clustering

In our experiments, we use the software package CLUTO[4] which is used for clustering low- and high-dimensional datasets. The algorithm adopted for clustering is Partitional, and the measure of the similarity between two vectors is Cosine similarity. Every document of the corpus is represented by two vectors: one is generated based on FT-Enrich method, and one simply is the traditional vector (BOW)–classical TF-IDF weighting of terms–model.

We tested and evaluated clustering with/without applying FT-Enrich, to show the improvements in clustering purity due to a capable combination of fusion and topic modeling approaches. The obtained results of such improvement are shown in Table II and Table IV on two various datasets BBC and Classic[4]. The obtained results in Table II on BBC indicate that representing documents by only using FT-Enrich ($\alpha = 1$) considerably improve the quality of clustering compared to using traditional TF-IDF method (*first baseline*) shown in Table I. We can see in Table II the best improvement in total purity (%22) and average of F1-measures (%23) are obtained by entirely using FT-Enrich method (α=1). Furthermore, in Table I and Table II, it can be observed in cluster 4 we have about %50 improvement in purity of the cluster.

2 http://www.dataminingresearch.com/index.php/2010/09/classic3-classic4-datasets/

3 http://qwone.com/~jason/20Newsgroups/

4 http://glaros.dtc.umn.edu/gkhome/views/cluto

TABLE I. Clustering Results of Dataset BBC Using Traditional Document Representations (First Baseline) ($\alpha = 0$)

Cluster	Bus	Enter	Polit	Sport	Tech	F1	Purity
Cluster 0	58	6	254	5	11	0.676	0.760
Cluster 1	320	2	15	4	5	0.748	0.925
Cluster 2	79	24	52	7	344	0.759	0.680
Cluster 3	30	16	15	441	5	0.866	0.870
Cluster 4	23	338	81	54	36	0.736	0.635
Total Purity							**0.763**

TABLE II. Clustering Results of Dataset BBC Using Ft-Enrich Method ($\alpha = 1$)

Cluster	Bus	Enter	Polit	Sport	Tech	F1	Purity
Cluster 0	21	21	401	27	13	0.891	0.830
Cluster 1	473	5	9	1	8	0.940	0.954
Cluster 2	13	6	3	0	364	0.925	0.943
Cluster 3	1	0	1	482	4	0.961	0.980
Cluster 4	2	354	3	1	12	0.934	0.952
Total Purity							**0.932**

We investigated the variation of α by considering the amount of dispersion of documents' sizes. Our experiments show that contribution of FT-Enrich method in creating the representation vectors for corpus with low Standard Deviation (SD) with respect to its mean (ME) is major compared to the one with the high SD. Table IV shows the clustering result with $\alpha = 0.1$ on Classic4 for which $\mathrm{SD} = 143.34$ and $\mathrm{ME} = 158.47$, but on the other hand, the clustering result shown in Table II is obtained by $\alpha = 1$ for which $\mathrm{SD} = 123.64, \mathrm{ME} = 341.21$.

We also compared our method with the *second baseline*, i.e., unsupervised LDA document representation. To this end, we considered the number of topics for each dataset corpus is equal to the number of classes manually specified for the corpus. For example, for dataset BBC with 5 manually specified classes, we ran LDA topic modeling with 5 topics over the corpus. Therefore, each document of BBC news corpus is represented by 5 different representation vectors/topics. Finally, documents that have highest probability/proportion in a same topic are clustered into the same cluster. The results of the clustering are shown in Table V and Table VI. As it can be observed, clustering using LDA representation alone returns worse result on dataset Classic4 compared to the results obtained by using traditional TF-IDF method (*first baseline*) shown in Table III.

TABLE III. Clustering Results of Dataset Classic4 Using Traditional Document Representations (First Baseline) ($\alpha = 0$)

Cluster	Cacm	Cisi	Cran	Med	F1	Purity
Cluster 0	323	30	11	21	0.730	0.839
Cluster 1	55	17	479	0	0.911	0.869
Cluster 2	47	6	4	454	0.898	0.888
Cluster 3	75	447	6	25	0.849	0.808
Total Purity						**0.852**

TABLE IV. Clustering Results of Dataset Classic4 Using Ft-Enrich Method (α =0.1)

Cluster	Cacm	Cisi	Cran	Med	F1	Purity
Cluster 0	334	9	2	0	0.790	0.968
Cluster 1	71	0	485	0	0.918	0.872
Cluster 2	43	0	6	485	0.938	0.908
Cluster 3	49	491	7	15	0.925	0.874
Total Purity						**0.898**

TABLE V. Clustering Results by Grouping Documents which Have a Same Topic with Highest Probability (Second Baseline) on the BBC

Cluster	Bus	Enter	Polit	Sport	Tech	F1	Purity
Cluster 0	5	12	0	70	285	0.737	0.766
Cluster 1	0	348	6	180	5	0.753	0.646
Cluster 2	462	9	17	0	10	0.917	0.928
Cluster 3	18	12	375	11	10	0.889	0.880
Cluster 4	25	5	19	250	91	0.555	0.641
Total Purity							**0.773**

TABLE VI. Clustering Results by Grouping Documents which Have a Same Topic With Highest Probability (Second Baseline) on the Classic4

Cluster	Cacm	Cisi	Cran	Med	F1	Purity
Cluster 0	211	386	11	10	0.691	0.625
Cluster 1	258	59	0	128	0.546	0.580
Cluster 2	25	8	484	0	0.952	0.936
Cluster 3	6	47	5	362	0.787	0.862
Total Purity						**0.745**

B. Evaluating Results of Cluster Labeling

We use 20NG benchmark for our experiments in cluster labeling. Therefore, we first show the result of clustering on this dataset using representation vectors generated by our method which indeed are used in cluster labeling. We further compare our result with the clustering result obtained by using the traditional representation vectors. The results of the clustering are shown in Table VII. It shows a remarkable improvement (%68) in the total purity of clustering (TP = 0.64) which leads to achieve significant result in cluster labeling as well.

The cluster labeling method represented in this work is a direct cluster labeling method in which the candidate labels for clusters are directly extracted from content of the clusters without using external sources (e.g. Wikipedia). One of the baseline direct approaches that several clustering systems apply for cluster labeling [17] is to select the top-n terms with maximal weights from the cluster centroid as the candidate labels. In our experiments we use this approach as a baseline for comparison. Specifically, we explore the effectiveness of using candidate labels generated by our approach in addition to the highest weighted terms extracted from cluster centroid provided by: TF-IDF and FT-Enrich method.

As an example of cluster labeling, Table VIII shows top-15 labels produced by the three above explained labeling methods over first cluster of 20News dataset which is labeled "Atheism" by experts. It can be observed in Table VIII that the labels produced by CombMNZ (topic-based) method are more describing a cluster of documents with subject Atheism than other methods. Specifically, first correct proposed label atheist (i.e. inflection for Atheism) is observed with $\mathrm{N} = 7$ (Match@7=1, MRR@7=0.143) for CombMNZ (topic-based), whereas for Centroid (topic-based) with $\mathrm{N} = 14$ (Match@14=1, MRR@14=0.071), and for Centroid (TF-IDF) with $\mathrm{N} = 15$ (Match@15=1, MRR@15=0.067).

Fig. 2 reports on the Match@N and MRR@N scores of each method for increasing values of N. As it can be observed, using the highest weighted terms extracted from clusters' centroids provided by FT-Enrich method is more effective than the ones provided by TF-IDF. It further shows that using fusion method (CombMNZ(FT-Enrich)) on the representation vectors generated by FT-Enrich method provides the best performance for both label quality measures. We can further observe that, for the Match@N measure, baseline method with FT-Enrich based cluster centroid requires at list 18 terms to cover %80 of the clusters with a correct label, while the same effectiveness is

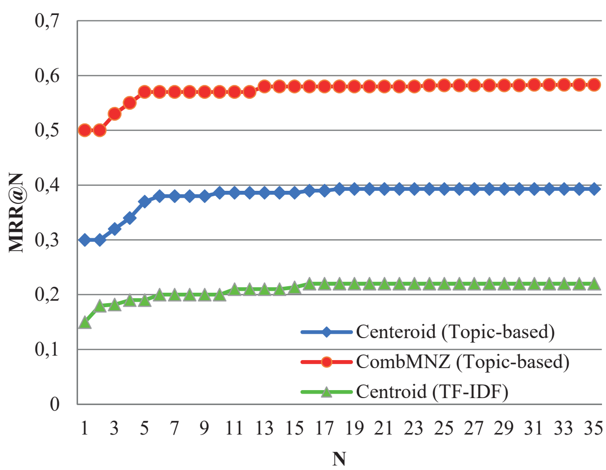
TABLE VII. Clustering Results by Grouping Documents which Have a Same Topic With Highest Probability (First Baseline) on the BBC, Using (**A**) Tf-Idf, and (**B**) Ft-Enrich Methods, Total Purity Indicated with **TP**

	Purity of Cluster																				
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	TP
A	0.25	0.39	0.16	0.5	0.21	0.3	0.46	0.48	0.25	0.3	0.16	0.29	0.71	0.24	0.84	0.35	0.41	0.4	0.71	0.45	**0.38**
B	1.0	0.96	0.89	0.99	0.96	0.64	0.63	0.43	0.23	0.42	0.28	0.47	0.95	0.59	0.88	0.95	0.94	0.94	0.35	0.81	**0.64**
	F1-measure																				
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
A	0.19	0.30	0.17	0.54	0.23	0.32	0.51	0.54	0.28	0.33	0.18	0.33	0.56	0.19	0.69	0.28	0.34	0.34	0.76	0.48	
B	0.54	0.69	0.46	0.87	0.91	0.46	0.71	0.41	0.33	0.28	0.36	0.43	0.55	0.67	0.89	0.68	0.88	0.88	0.49	0.79	

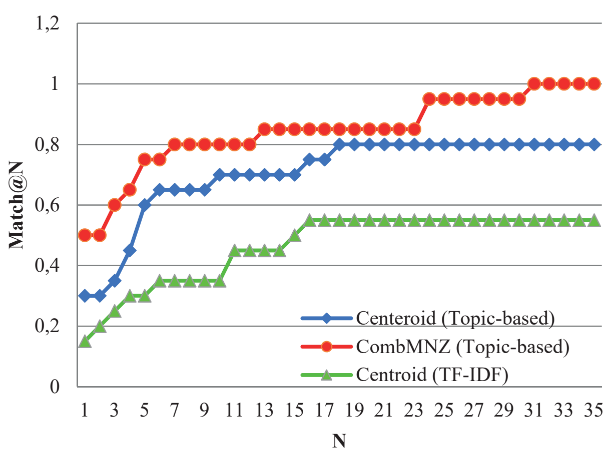
TABLE VIII. An example of Top-15 proposed labels, using three different methods; (**A**) Centroid (tf-idf), (**B**) Centroid (topic-based), and (**C**) CombMNZ (topic-based) over first cluster of 20News dataset "Atheism"

	Labels														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	caltech	keith	solntze	livesey	Livesey	sgi	wpd	Schneider	Keith	nuclear	Allan	jon	mathew	Political	Atheist
B	system	moral	person	wrong	morality	objective	murder	keith	life	jon	society	innocent	god	atheist	human
C	god	moral	person	life	wrong	morality	**atheist**	objective	murder	human	evidence	society	keith	truth	jon

achieved by a list of 7 terms only using FT-Enrich method. It is also interesting that with $N > 31$ CombMNZ (FT-Enrich) method covers %100 of the clusters with a correct label.



(A) MRR@N



(B) Match@N

Fig 2. Average (A) MRR@N and (B) Match@N values obtained for clusters of 20NG using fusion method over representation vectors generated by FT-Enrich, using top-N terms of cluster centroid weighted by FT-Enrich method, and using top-N terms of cluster centroid weighted by TF-IDF.

VIII. Conclusion

In this paper, we presented a fusion- and topic-based enriching approach in order to improve the quality of clustering. We applied a statistical approach, namely topic model, to enrich the representation vectors of the documents. To this end, an ensemble topic modeling with using different parameters for each model are represented, and then, using a fusion approach, all the generated results are combined to provide a single vectorial representation for each document. Our experiments on the different datasets show significant improvement in clustering results. We further show that putting such representation vectors in a fusion method provides interesting results in cluster labeling as well.

As a future work, we plane to exploit external sources (e.g. WordNet) in both the clustering and cluster labeling to explore the effectiveness of using topic models as well as the resources in corresponding domains.

References

[1] Yue Lu, Qiaozhu Mei and ChengXiang Zhai, Investigating task performance of probabilistic topic models: an empirical study of PLSA and LDA, Information Retrieval 14, pp. 178–203, April 2011.

[2] David M Blei, Probabilistic topic models, Communications of the ACM 55, pp. 77–84, April 2012.

[3] Shengli Wu, Data fusion in information retrieval, 13, Springer Science & Business Media, 2012.

[4] Youngjoong Ko, A study of term weighting schemes using class information for text classification, in: Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, ACM, pp. 1029–1030, August 2012.

[5] Gerard Salton and Christopher Buckley, Term-weighting approaches in automatic text retrieval, Information processing & management 24, pp. 513–523, January 1988.

[6] Somnath Banerjee, Krishnan Ramanathan and Ajay Gupta, Clustering short texts using wikipedia, in: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, ACM, pp. 787–788, July 2007.

[7] Andreas Hotho, Steffen Staab and Gerd Stumme, Ontologies improve text document clustering, in: Data Mining, 2003. ICDM 2003. Third IEEE International Conference on, IEEE, pp. 541–544, November 2003.

[8] Revanasiddappa M. B. and Harish B. S., A new feature selection method based on intuitionistic fuzzy entropy to categorize text documents, International Journal of Interactive Multimedia and Artificial Intelligence 5(3), pp. 106-117, 2018.

[9] Chyi-Kwei Yau, Alan Porter, Nils Newman and Arho Suominen,

Clustering scientific documents with topic modeling, Scientometrics 100, pp. 767–786, September 2014.

[10] Yang Liu, Zhiyuan Liu, Tat-Seng Chua and Maosong Sun, Topical word embeddings, in: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, AAAI Press, pp. 2418–2424, January 2015.

[11] Duc-Thuan Vo and Cheol-Young Ock, Learning to classify short text from scientific documents using topic models with various types of knowledge, Expert Systems with Applications 42, pp. 1684–1698, February 2015.

[12] Heng Zhang and Guoqiang Zhong, Improving short text classification by learning vector representations of both words and hidden topics, Knowledge-Based Systems 102, pp. 76–86, June 2016.

[13] Tomas Mikolov, Kai Chen, Greg Corrado and Jeffrey Dean, Efficient estimation of word representations in vector space, arXiv preprint arXiv: pp. 1301.3781, January 2013.

[14] Quoc Le and Tomas Mikolov, Distributed representations of sentences and documents, in: International Conference on Machine Learning, pp. 1188–1196, January 2014.

[15] Haggai Roitman, Shay Hummel and Michal Shmueli-Scheuer, A fusion approach to cluster labeling, in: Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval, ACM, pp. 883–886, July 2014.

[16] David Carmel, Haggai Roitman and Naama Zwerdling, Enhancing cluster labeling using wikipedia, in: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, ACM, pp. 139–146, July 2009.

[17] Douglass R Cutting, David R Karger, Jan O Pedersen and John W Tukey, Scatter/gather: A cluster-based approach to browsing large document collections, in: Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval, ACM, pp. 318–329, August 1992.

[18] Derek Greene and Pádraig Cunningham, Practical solutions to the problem of diagonal dominance in kernel document clustering, in: Proceedings of the 23rd international conference on Machine learning, ACM, pp. 377–384, June 2006.

[19] Jonathan A Silva, Elaine R Faria, Rodrigo C Barros, Eduardo R Hruschka, André CPLF de Carvalho and João Gama, Data stream clustering: A survey, ACM Computing Surveys (CSUR) 46, October 2013.

[20] Yutaka Sasaki et al., The truth of the F-measure, Teach Tutor mater 1, pp. 1–5, October 2007.

[21] Pucktada Treeratpituk and Jamie Callan, Automatically labeling hierarchical clusters, in: Proceedings of the 2006 international conference on Digital government research, Digital Government Society of North America, pp. 167–176, May 2006.

M. Pourvali

Experienced Lecturer with a demonstrated history of working in the research industry. Skilled in Word Sense Disambiguation, Text Clustering, Text Summarization, Document Enrichment, and generally in Natural Language Processing. Strong education professional with a Doctor of Philosophy (PhD) focused on Computer Science from Ca' Foscari University of Venice in Italy.

S. Orlando

MSc (1985) and PhD (1991), University of Pisa - is a full professor at Ca' Foscari University of Venice. His research interests include data and web mining, information retrieval, parallel/distributed systems. He published over 150 papers in peer reviewed international journals and conferences. He co-chaired conferences, tracks, and workshops, and served in the PC of many premier conferences.

H. Omidvarborna

She received her Bachelor Degree in Electrical Engineering from Razi University, and continued her studies in Computer and Communication Networks at Polytechnic University of Turin. Her research interests include Information retrieval and analysis and visualization of massive data.

Sales Prediction through Neural Networks for a Small Dataset

Rosa María Cantón Croda, Damián Emilio Gibaja Romero, Santiago Omar Caballero Morales *

Deanship of Graduate Studies in Engineering and Businesses, Universidad Popular Autónoma del Estado de Puebla, Puebla (Mexico)

Received 5 November 2017 | Accepted 17 February 2018 | Published 13 April 2018

Abstract

Sales forecasting allows firms to plan their production outputs, which contributes to optimizing firms' inventory management via a cost reduction. However, not all firms have the same capacity to store all the necessary information through time. So, time-series with a short length are common within industries, and problems arise due to small time series does not fully capture sales' behavior. In this paper, we show the applicability of neural networks in a case where a company reports a short time-series given the changes in its warehouse structure. Given the neural networks independence form statistical assumptions, we use a multilayer-perceptron to get the sales forecasting of this enterprise. We find that learning rates variations do not significantly increase the computing time, and the validation fails with an error minor to five percent.

Keywords

Artificial Intelligence, Artificial Neural Networks, Forecasting.

DOI: 10.9781/ijimai.2018.04.003

I. Introduction

SALES forecasting has positive implications within enterprises. It can improve the planning of production processes as well as inventory management practices [1]. Specifically, the fast growth of enterprises and their incursion into new markets make inventory optimization necessary. Even more, because sales have a close relationship with the clients' demands which depend on factors that enterprises cannot control, knowing the customers' behavior may reduce the enterprises' operational costs [2].

It is important to recall that predicting future demand is a complex problem since it relies on determining a concrete definition of what demand is [3]. Note that order times, preferences and previous sales affect the current demand that a company must satisfy or supply. Thus, forecasting accuracy depends on the information gathered by the enterprise [4]. In this sense, the literature recognizes that previous sales are information that companies register to be used as starting points. Even more, sales do not only summarize the sold quantity, but they also include information that reflects changes in consumer's behavior concerning the selling prices. Hence, sales represent a good proxy to understand the demand trends that a company faces [5].

Regarding the quantity of data required to estimate demand trends, there is an active discussion in the literature focused on determining the optimal size of a dataset for demand forecasting [6]. In general, research studies on this topic agree that dataset size depends on the prediction model [7]. For example, ARIMA MODELS need $p+q+P+Q+d+mD$ observations, where p is the number of time lags, q represents the model's order, m is the total number of periods at each season, and d is the differencing degree. Also, this model pretends to capture seasonal effects, a reason why it includes terms related to autoregressive P, moving Q and differencing averages D. On the other hand, literature has suggested the use of artificial neural networks (ANNs) as robust techniques to analyze small datasets due to their performance being independent of the total number of input data [8]. Recently, it is common to find applications of ANNs to solve forecasting problems in medicine, where small datasets prevail since patients and diseases' evolution change regularly [9]. This last factor has a direct impact on patients' medical visits, which tend to no show at them because they "feel better". A prediction of such behavior is useful to reduce costs and improve medical services efficiency [10].

Given the favorable applications of ANNs to analyze non-linear behavior in datasets [1, 9] our work was focused on the application of this technique to predict the sales of a chemical company located in Mexico. For this case, the dataset consisted of 12 monthly sales data as the company reports its sales on a monthly basis. 70.0% of this data was considered for training and 30.0% for initial validation of the ANN which was designed for this work.

This ANN consisted of a multi-layer perceptron with a sigmoid activation function. Implementation of the ANN was performed using programming with the software R v.3.2.1. We find that a small dataset does not significantly increase the number of iteration, even when we consider the maximum possible learning rate. However, as it is expected by the classical time series literature, the prediction presents a higher error, which we demonstrate in the validation phase.

The paper is organized as follows. Section II presents a brief literature review of the sales forecasting literature, and we discuss the viability and applications of ANNs. Then, Section III describes the designing of the ANN with a general overview of how these techniques work. Section IV presents the experiments associated

* Corresponding author.

E-mail addresses: rosamaria.canton@upaep.mx (R. M. Cantón Croda), damianemilio.gibaja@upaep.mx (D. E. Gibaja Romero), santiagoomar.caballero@upaep.mx (S. O. Caballero Morales).

with the ANN, highlighting interesting insights about the training, validation and prediction processes. To motivate the use of ANN, even in small databases, we compare the ANN performance with the one of using the average moving technique. Finally, Section V presents our conclusions.

II. Literature Review

Given that time series are central in determining predictions, there is a considerable literature that analyzes their characteristics and the methodologies to find the desired results [11]. This paper does not pretend to be an exhaustive exposition of the time-series forecasting literature. However, we present some of the most significant contributions in the application of time-series for demand forecasting.

A. Sales Forecasting and Industry

Industries, in general, concern about demand forecasting since their benefits depend on consumer's behavior. Hence, forecasting studies contribute to diminishing uncertainty about future sales, which implies a better production planning and inventory cost reduction [12].

Although future sales have a high degree of uncertainty, there exist different approaches to approximate the consumer's behavior in future periods. The literature on this problem considers different approaches and methodologies, which rely on the analysis of time series that summarizes how the demand evolves through time together with its related variables [3].

It is important to recall that demand and sales have a close relationship since the last summarizes the consumers' behavior. Thus, sales work as a good indicator of consumers' preferences even though they are subjective [13]. For example, demand forecasting uses variables such as sales, expenses, and benefits [14]. However, there exist other studies where demand is predicted or forecasted through different variables that depend on the market where companies and products stand. For example, Rodriguez et al. [18] built an index for water demand that summarized water consumption, land use, average incomes, and population. In the case of electricity demand, it is common to consider income, prices of related fuels, community, and technology [15]. All the previous applications used time-series to determine future events.

B. Classical Time-Series Forecasting Literature

Typically, Time-Series Forecasting relies on the application of two methods: Holt-Winters and Box-Jenkins [16]. In the case of time-series with linear behavior, the previous techniques generate a near-exact approximation of future demands. Thus, we can find applications of these methods in forecasting exchange rate [17], demand flowmeters [14], aircraft failure rates [18], general prices [19], and transport demand [20] among others.

It is important to mention that not all time-series present a linear behavior. In a non-linear situation, the Holt-Winters and Box Jenkins methods fail to provide a good forecasting of future events. Specifically, linear models cannot capture and analyze amplitude dependence, volatility clustering, and asymmetry, which prevail in finance time-series [21]. In these cases, it is appealing to assume different states [22], or considering situations with the minimum number of statistical assumptions [23].

The analysis of non-linear time-series requires the construction of models suitable for each phenomenon, which pretend to capture stochastic processes. Hence, Campbell, Lo and McKinley [24] showed that it is possible to classify most of the non-linear time series according to their deviations from their mean, the non-linear moving average model [25], or their variance (for example, the ARCH model and their variations like GARCH and EGARCH [26].

C. Neural Networks and Sales Forecasting

Classical non-linear time series models have a significant dependence on statistical assumptions, which diminish their capacity to predict a behavior that was not previously registered. To deal with such problem, Artificial Neural Networks (ANNs) appear as a suitable alternative. These techniques try to mimic neural cells behavior, which makes them non-linear models by nature [27]. Because ANNs are data-driven, they have a self-adaptive feature. This adaptation capacity makes them suitable to analyze problems where data is not complete, or when its documentation was not appropriately gathered (i.e., data with noise [28]).

Given that ANNs allow the approximation of almost all non-linear continuous function, they are suitable for sales forecasting [29]. However, in the classical literature, there is not a unique approach to sales forecasting because ANNs may have different structures according to the problem. For example, electricity demand forecasting considers a hybrid structure since data may present seasonal effects. An et al. [15] propose a two-step architecture, where seasonal effects are reduced in the first step through an empirical mode decomposition; and later, a multi-output feed-forward ANN is used to get the demand forecast for the following H periods given the demand of the previous T periods. Unlike electricity demand, water consumption follows an ANN with stochastic connections since it noisy data prevail in such situations [14]. In both cases, ANNs' architecture requires a long time-series.

Graphically, an ANN is a graph with neurons (nodes) and their synapses (edges). These models may learn and identify consumers' behavior patterns according to the input data and the architecture of the ANN. The structure of an ANN summarizes the interconnection between neurons, which are organized in layers [30]. The most uncomplicated architecture considers the input layer, the output layer, and the hidden layer, whose objective is the identification and communication of pattern to get the demand forecasting. Fig. 1 shows an example of an ANN with three layers (input, hidden and output layer).

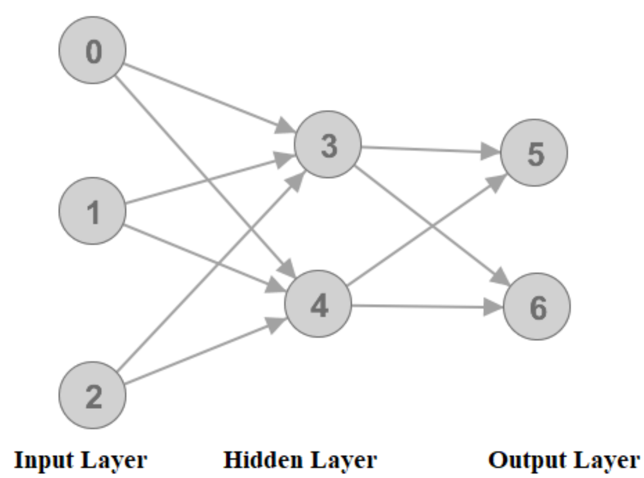


Fig. 1. ANN with only one hidden layer.

The hidden layer serves to adjust the importance of the input data in the determi nation of future behavior. Therefore, it is possible to add as many hidden layers as required [27]. Although no theoretical tool indicates the optimal number of hidden layers, experiments demonstrate that more than one hidden layer does not significantly improve the prediction's quality. Also, there is a positive relationship between the number of hidden layers and the computing time. Thus, one hidden layer is enough to get a proper prediction, mainly when the database is small [30].

III. Designing a Neural Network for Forecasting

We analyze the case of a chemical company located in Mexico. This firm owns nine central warehouses across the Mexican territory which are used to fulfill the requisitions of its clients. Because these distribution centers may generate inventory costs to the detriment of company's benefits, the company is interested if sales justify the construction of additional warehouses. This motivates the prediction of future sales through ANNs.

A. Data Selection

The chemical company builds central warehouses at different locations, which imply the absence of a long time-series that includes all central warehouses sales. Thus, we considered a time series from 2015 to 2016 that consists of the sales of the current distribution centers.

B. Data Collection

The chemical enterprise provided a database with the sales that each of its central warehouses must comply. This dataset included information related to a product, client and transportation costs. However, due to confidentiality restrictions, we considered only sales information of twelve months from 2015 to 2016 for the time series that shows the company's sales for one year.

C. Choosing the Neural Network

Forecasting literature that applies ANNs agrees in the use of the multi-layer perceptron (MLP). This ANN is characterized by a feedforward architecture where unidirectional weighted edges connect neurons. Since we have a total of 12 data, we considered an MLP with three neurons in its hidden layer. Thus, the MLP mathematical expression is

$$y_t = \alpha_0 + \sum_{j=1}^{3} \alpha_j g\left(\beta_{0j} + \sum_{i=1}^{12} \beta_{ij}\, y_{t-i}\right) + \epsilon_t, \forall\, t,$$

where α_j and β_{ij} are the connection weights, for all $j = 0, 1, 2, 3$ and $i = 0, 1, 2, \ldots, 12$. Also, g represents the activation function that defines the output of a neuron. Given the size of the dataset, we considered a hidden layer with three neurons.

There are different activation functions, and their use depends on the problem's features. Such functions may produce an activation within a discrete range, like a set {0,1} that represents "yes" or "no," or a continuous range like $\mathbb{R}$. In the case of forecasting, Thomonopoulus [2] suggested that continuous ranges are more desirable because they do not restrict neurons behavior to a finite number of results. Also, the sigmoid function was suggested as beneficial for forecasting because it allows any real number as an input, and does not present convergence problems since it is bounded between -1.0 and 1.0. Thus, in this paper, we consider the sigmoid function as the activation function

$$sig(t) = \frac{1}{1+e^{-t}}.$$

For our sales prediction problem, we compute y_t which is the company's sales during period t. To generate the prediction, the perceptron that we consider uses the sales from the three previous periods. In other words, we "learn" from months one, two and three to "predict" the fourth month. Then, the ANN computes the weights that minimize the error, and such values are saved. In the next iteration, the ANN "learns" from month two, three and four to "predict" the fifth-month sale. So, we proceed in the following way

ANN_0: Input (1,2,3) → output (4)

ANN_1: Input (2,3,4) → output (5)

ANN_2: Input (3,4,5) → output (6)

ANN_3: Input (4,5,6) → output (7)

ANN_4: Input (5,6,7) → output (8)

ANN_5: Input (6,7,8) → output (9)

We recall that the previous architecture produces a single output, and weights are actualized every time a fourth value is predicted using the three previous weights. That is to say, we choose a recurrent architecture for the neural network [30].

IV. Neural Network Implementation

There is much mathematical software that is capable of programming and implementing ANNs. We choose the software R to implement the ANN given the possibility to change the ANN parameters.

Now, the construction of an ANN requires the weights' computation, and their validation as well. To do this, we divide the database into two sets: one for training (or learning) and one for validation. By following the guidelines of [31, 32], we determined the size of both sets: the training set was considered with nine data, and the validation set was considered with three data. This corresponds to the 75.0% and 25.0% of the entire database respectively.

A. Training

The self-adaptation property allows ANNs to adjust future behavior given previous behavior. Thus, in the training phase, the ANN learned the time series behavior through computing how high the connections between neurons must be. Below, we present the code in R language.

```
## DEFINING THE NEURAL NETWORK AS A FUNCTION

neuron = function (normaldemands, learn, lA) {
## VARIABLES INITIALIZATION
weights = c (runif (length (normaldemands), min = 0, max
= 1))
weightsO = weights; weightsE = 0; sigmoid = 0;
error1 = 2; itera = 1; ErrorI = 0
while (error1> = 0.001) {
     x = 0; weightsE = weights
     for (i in 1: length (normaldemands)) {
       x = x + c (weights [i] * normaldemands [i])
     } ## end for

     sigmoid = (1 / (1 + exp (-x))) ## Sigmoid function
     error1 = (learn - sigmoid) * sigmoid * (1-sigmoid)

      ## NEW WEIGHTS COMPUTATION
     for (i in 1: length (normaldemands)) {
       weights [i] = weights [i] + lA * error1*
       normaldemands[i]
     } ## end for
     ErrorI [itera] = error1;
     iterate = iterate + 1;
} ## end while

output  =  list  (weightsO  =  weightsO,  weightsE  =
weightsE,     sigmoid = sigmoid, error = error1, ErrorI =
ErrorI, learn = learn)
return (output)
} ## end function
```

About the previous code, it is important to remark the following:

- First, we know that the sigmoid function range is the interval [0,1]. Thus, this function outputs values between zero and one. Since the time series registers the total sales, we need to normalize the values of our database. So, let x_t be the sales of period t, and y_t the normalized sales for period t. We use the following transformation

$$x_t = \frac{1}{z_{max}} z_t,$$

where $z_{max} = \max\{z_1, z_2, \ldots, z_{12}\}$ and the values of the time series are: 2975760.227, 3451339.887, 3567417.723, 3560103.948, 3933955.267, 3573328.611, 4032969.915, 3788327.879, 4350361.992, 4597373.737, 3567670.195, 4047731.206 and $z_{max} = 4597373.737$.

- Second, the training part is a supervised learning process. This means that the ANN adjusts the weights between the neurons to learn the behavior or pattern of the time series. Then, the ANN computes the weights that minimize the following error

$$error = (y_t - x_t)^2$$

In this exercise, we tolerate an error less than 0.001.

- Third, programming the ANN in R allows determining the learning rate. This indicates how fast the ANN adapts to new behavior presented in the time series. Given that we have a small dataset, we compared different learning rates to assess the ANN's performance.

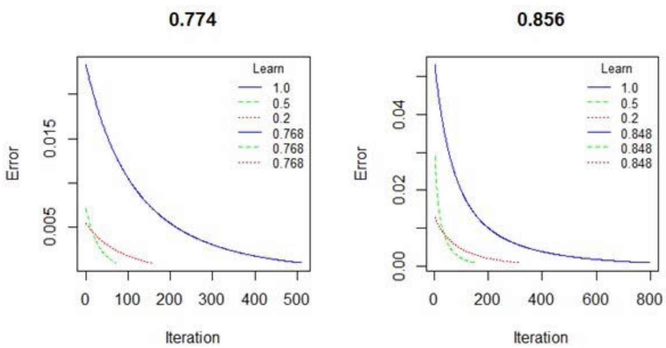


Fig. 2. Comparison between learning rates and iterations.

Fig. 2 shows the relationship between the error and the number of iterations when we consider the following learning rates: 0.2, 0.5 and 1.0. In this situation, the ANN tries to learn the real values 0.774 and 0.856. Note that the highest learning rate implies a complete understanding of the time series behavior, but it requires the most significant number of iterations. However, despite the database's size, the number of iterations is not unusual in the literature [27].

In general, the previous behavior is replicated by the ANN when we change the real values. For example, Fig. 3 shows the relationship between the three learning rates when the ANN learns the largest possible sale, which is equal to 1.0. Although the number of iteration increases while the ANN tries to reduce the error, this number does not excessively increase.

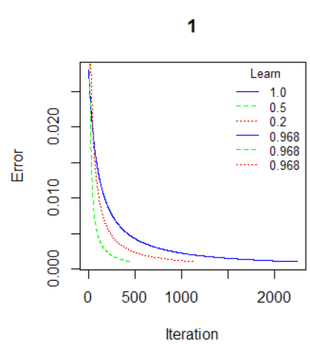


Fig. 3. Learning the largest possible sale.

Fig. 2 and 3 illustrate how fast the ANN reduces the approximation error when learning rate is switched. In all cases, the small dataset does not induce a time convergence problem, the number of iterations is relatively small. Also, as the literature suggests, when we impose a learning rate equal to 1.0, the number of iterations increases. It is interesting to note that a learning rate equal to 0.5 is faster than the learning rate equal to 0.2. However, the computation time is not an additional restriction for our work, R spends 27.82 seconds to compute all weights.

Given the limited information, we consider appropriate to choose a learning rate equal to 0.2. This assumption comes from the literature because an ANN based on small dataset requires slackness in their learning process [8].

B. Validation

For the validation process, we use the nine weights produced by the algorithm during the previous step. Remember that the fifth iteration summarizes the knowledge required by the ANN to forecast. Then, we have that

Input (7,8,9) → ANN_5 → Output (10)

Input (8,9,10) → ANN_5 → Output (11)

Input (9,10,11) → ANN_5 → Output (12)

Table I presents the weight for each value in the training set and the number of iterations that R took to compute them.

TABLE I. Weights from the Supervised Phase

MONTH	SALES (z_t)	WEIGHT	ITERATIONS
1	2975760.227	0.9544753	250
2	3451339.887	0.9833160	294
3	3567417.723	0.3501016	230
4	3560103.948	0.4076781	390
5	3933955.267	0.4222303	280
6	3573328.611	1.2014761	627
7	4032969.915	1.0023636	1090
8	3788327.879	0.3538968	172
9	4350361.992	0.4986251	274

To proceed with the validation phase, we use the sales for months 10, 11 and 12. By the ANN architecture, we follow an iterative process where the weights of the three previous periods were used. In this case, the process was unsupervised, i.e., we let the ANN adjusts the weights in a free manner to get the last three known values. Table II illustrates the weights that we use in the validation process.

TABLE II. Validation Process

MONTH	WEIGHTS		
10	1.6621	0.9737	1.2102
11	0.1779	0.7778	0.3281
12	0.4470	1.0622	0.5489

During the validation process, we find that the ANN does not predict the time series with a one hundred percent of accuracy. However, the error is not higher than the 10%, as Table III shows with the comparison between the real values and the ones produced by the ANN.

TABLE III. Validation Process

REAL SALES	TEST SALES	ERROR
1.00	.967	0.043
.776	.770	0.01
.880	.871	0.0103

C. Predictions

Although the ANN does not entirely replicate the time series behavior, this is due to the database size. Consequently, we use the weights in Table II to get the predictions for the next three months. Then, the ANN architecture assumes that the first projection, sales of month 13, is real to compute the prediction of month 14. We continue iteratively to get the prediction for month 15. Then, the algorithm does the following computations

Input (10,11,12) → ANN_5 → Output (13)

Input (11,12, Output (13)) → ANN_5 → Output (14)

Input (12, Output (13), Output (14)) → ANN_5 → Output (15)

Following the previous process, Table IV shows the weights that the ANN requires to compute the sales prediction for the next three months.

TABLE IV. Prediction Weights

MONTH	WEIGHTS		
13	0.4470030	1.0622665	0.5489295
14	1.0914885	0.5261332	0.2219956
15	0.5159739	0.9045625	0.5206516

Using the information in Table IV, we proceed to compute the sales that the chemical company may face in the first three months of the next year. In Table V we present these results in the normalized form and their corresponding transformation.

TABLE V. Main results

MONTH	NORMALIZED SALES (x_t)	SALES (z_t)
13	0.8525044	3919281.34
14	0.8448968	3884306.36
15	0.8375715	3850629.22

Finally, Fig. 4 shows the relation between the three phases: training, validation, and prediction. In this graph, we compare the prediction values with the real values, and we can see the trend for future sales. As we mentioned before, during the supervised phase, the ANN fully captures the previous behavior of the time series, while in the validation phase it presents problems to predict with a 100% of accuracy the other values. However, the distance between both lines is not significant (less than 5%).

Also, it is interesting to note that the following next three months present a decreasing behavior maybe which does not match with the behavior in the corresponding months in the database. This last point may be related to the fact that we consider the 13th month as real information to get the months 14 and 15.

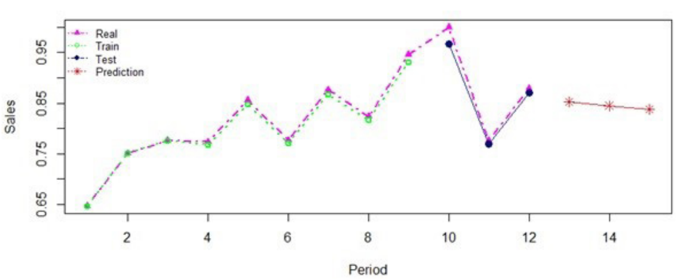


Fig. 4. Comparison between real and predicted time series with normalized data.

D. Comparison with Moving Average

As we mentioned before, moving average models (ARMA, ARIMA, …) are commonly used to analyze non-linear series [10]. In this section, we compare the sales forecasting that results from the application of a simple moving average, with the ANN's results. Since we have a small database, we forecast a month by using the following simple moving average (SMA) model

$$y_t = \frac{x_{t-1} + x_{t-2} + x_{t-3}}{3} \; for\ all\ t \geq 4.$$

In other words, this model approximates the sales of month t using the mean of the three previous month sales.

Table VI shows the sales forecasting when we use the SMA technique and the approximation error for each month. Fig. 5 illustrates the original time series with the moving average results.

TABLE VI. Moving Average Results

MONTH	SALES (z_t)	MOVING AVERAGE FORECAST (y_t)	ERROR
1	2975760.227		
2	3451339.887		
3	3567417.723		
4	3560103.948	3331505.95	0.064211
5	3933955.267	3526287.19	0.103628
6	3573328.611	3687158.98	-0.031855
7	4032969.915	3689129.28	0.085257
8	3788327.879	3846751.26	-0.015421
9	4350361.992	3798208.8	0.126921
10		4057219.93	
11		4069344.94	
12		4350361.99	

Considering the sales of the validation set (months 7, 8 and 9), we note that the SMA prediction reports an error greater than 5% percent for months 7 and 9 (see Table VI). In comparison, for the same set of months, the ANN provides predictions with an error less than 5%, see Table III. Since the sales of the validation are known, we can conclude that the ANN's predictions are more accurate than the ones from the SMA.

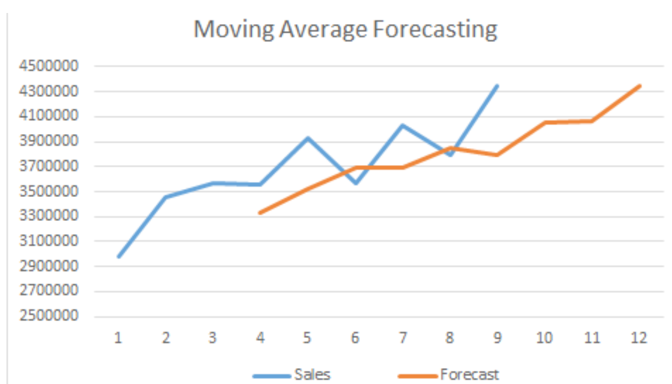


Fig. 5. Original sales (blue) versus average moving predictions (orange).

V. Conclusion

The present paper shows the flexibility of neural networks to deal with sales forecasting in the presence of a small database. As the literature suggests, the ANN does not fully capture the time series behavior as we show during the validation phase. However, it is worth noting that the prediction error is not greater than 5%, which is accurate in sales forecasting instances [31].

Even more, in our case of study, the small dataset does not increase the time convergence even when we impose a high learning rate. Also, the ANN has a better performance than the SMA concerning error.

Specifically, for months 7 and 9, whose sales are known, the SMA reports an error greater than 5% while the ANN reports an error less than 5%. The previous facts support the use of this technique in a situation when time series are small, or the database does not present all the possible information.

Therefore, ANN may be useful in a situation where time series present a non-linear behavior and do not show an explicit statistic behavior, as it is the situation in time series with a short length. However, in the present experiment, we observe the significant dependence that new predictions have when the first prediction is considered real information. In future works, we address this problem from theoretical and empirical perspectives.

References

[1] R. Carbonneau, K. Laframboise, and R. Vahidov, "Application of machine learning techniques for supply chain demand forecasting." European Journal of Operational Research, vol. 184, n° 3, pp. 1140-1154, 2008.

[2] N. Thomopoulos, "Demand Forecasting for Inventory Control. In: Demand Forecasting for Inventory Control." Springer, Cham, 2015.

[3] J. R. M Hosking, "Demand Forecasting Problems in Production Planning." Planning Production and Inventories in the Extended Enterprise. International Series in Operations Research & Management Science, vol. 151, 2011.

[4] P. Yenradeea, A. Pinnoi, and A. Charoenthavornying, "Demand forecasting and production planning for highly seasonal demand situations: a case study of a pressure container factory," Science Asia, vol. 27, pp. 271-278, 2001.

[5] R. Skyrius, G. Kazakevičienė, and V. Bujauskas, "From management information systems to business intelligence: The development of management information needs," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 2, n°3, 2013.

[6] R. J. Hyndman, and A. V. Kostenko, "Minimum sample size requirements for seasonal forecasting models," Foresight, vol. 6, pp. 12-15, 2007.

[7] P. B. Prakash, V. Ramya, and M. Yugandhar, "A Statistical GA Based Demand Forecasting Model for Automotive Batteries Manufacturing Company," i-Manager's Journal on Mechanical Engineering, vol.6, n° 3, pp-18, 2016.

[8] S. Ingrassia, and I. Morlini, "Neural network modeling for small datasets," Technometrics, vol.47, n° 3, pp. 297-311, 2005.

[9] A. Pasini, "Artificial neural networks for small dataset analysis." Journal of thoracic disease, vol. 7, n° 5, pp- 953, 2015.

[10] C. Elvira, A. Ochoa, J. C. González and F. Mochón, "Machine-Learning-Basec no Show Prediction in Outpatient Visits," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 4, n°7, 29-34, 2017.

[11] C. Chatfield, "Time-series forecasting." CRC Press, 2000.

[12] L. Šečkute and A. Pabedinskaite, "Application of forecasting methods in business," Journal of Business Economics and Management, vol. 4, n° 2, pp. 144-157, 2003.

[13] L. Kurzak, "Importance of Forecasting in Enterprise Management." Advanced Logistic systems, vol. 6, n° 1, pp.173-182, 2012.

[14] H. Rodriguez, V. Puig, J. J. Flores, and R. Lopez, "Combined holt-winters and GA trained ANN approach for sensor validation and reconstruction: Application to water demand flowmeters." 2016 3rd Conference on Control and Fault-Tolerant Systems (SysTol), on, pp. 202-207, 2016.

[15] N. An, W. Zhao, J. Wang, D. Shang, and E. Zhao, "Using multi-output feedforward neural network with empirical mode decomposition based signal filtering for electricity demand forecasting." Energy, vol. 49, pp. 279-288, 2013.

[16] L. Escobar, J. Valdes, and S. Zapata, "Redes Neuronales Artificiales en predicción de Series de Tiempo. Una aplicación a la Industria," Working Paper Series Universidad de Palermo, 2009.

[17] E. Valakevicius, and M. Brazenas, "Application of the Seasonal Holt-Winters Model to Study Exchange Rate Volatility." Engineering Economics, vol. 26, n° 4, 384-390, 2015.

[18] Y. M. Yang, H. Yu, and Z. Sun," Aircraft failure rate forecasting method based on Holt-Winters seasonal model," 2017 IEEE 2nd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), pp. 520-524, 2017.

[19] A. Singh, and G. C. Mishra, "Application of Box-Jenkins method and Artificial Neural Network procedure for time series forecasting of prices," Statistics in Transition new series, vol. 1, n° 16, pp. 83-96, 2015.

[20] S. Anvari, S. Tuna, M. Canci, and M. Turkay, "Automated Box–Jenkins forecasting tool with an application for passenger demand in urban rail systems," Journal of Advanced Transportation, vol.50, n°1, pp. 25-49, 2016.

[21] C. M. Kuan, "Lecture on the Markov switching model. Institute of Economics Academia Sinica," pp. 1-30, 2002.

[22] N. Davies, and J. Petruccelli, "Detecting Non-Linearity in Time Series," Journal of the Royal Statistical Society. Series D (The Statistician), vol. 35, n° 2, pp. 271-280, 1986.

[23] M. Troffaes, J. Gledhill, D. Škulj, and S. Blake, "Using imprecise continuous-time Markov chains for assessing the reliability of power networks with common cause failure and non-immediate repair," SIPTA working papers, 2015.

[24] J. Y. Campbell, A. W. C. Lo, and A. C. MacKinlay, "The econometrics of financial markets," Princeton University Press, 1997.

[25] P. M. Robinson, "The estimation of a nonlinear moving average model," Stochastic processes and their applications, vol. 5, n°1, pp. 81-90, 1977.

[26] D. B. Nelson, "ARCH models as diffusion approximations," Journal of econometrics, vol. 45, n° 1-2, pp. 7-38, 1990.

[27] R. Adhikari, and R. K. Agrawal, "An introductory study on time series modeling and forecasting," arXiv preprint arXiv:1302.6613, 2013.

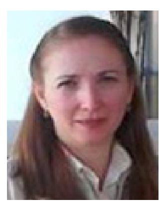
[28] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feed-forward networks are universal approximators", Neural Networks, vol. 2, pp. 359–366, 1989.

[29] M. Garetti, and M. Taisch, "Neural networks in production planning and control," Production Planning & Control, vol. 10, n° 4, pp. 324-339, 1999.

[30] I. A. Basheer, and M. Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application," Journal of microbiological methods, vol. 43, n° 1, pp. 3-31, 2000.

[31] A. S. Ahmad, M. Y. Hassan, M. P. Abdullah, H. A. Rahman, F. Hussin, H. Abdullah, and R. Saidur, "A review on applications of ANN and SVM for building electrical energy consumption forecasting," Renewable and Sustainable Energy Reviews, vol. 33, pp. 102-109, 2014.

[32] K. Haddouch, K. Elmoutaoukil, and M. Ettaouil, "Solving the Weighted Constraint Satisfaction Problems Via the Neural Network Approach." International Journal of Interactive Multimedia and Artificial Intelligence, 4(1), 56-60, 2016.



Rosa María Cantón Croda

Ph.D. in Computer Science from the Tecnológico de Monterrey, Ciudad de México, Master in Information Technologies from the Tecnológico de Monterrey, Veracruz, was graduated in Computer Science Administrator from the Tecnológico de Monterrey, Monterrey. She was Manager of Systems in transport and construction companies. More than twenty years of experience in the academic administration of Tecnológico de Monterrey and currently Dean of postgraduate studies in Engineering and Business at UPAEP. She has written some articles for national and international congresses. She was certified in PMBook, Positive Psychology by Tecmilenio and in Project-Based Learning by Aalborg University in Denmark. Currently, her research interests are Big Data and Business Intelligence.

Damián Emilio Gibaja Romero

Ph.D. in Economics from El Colegio de México, Ciudad de México, Master in Economics from El Colegio de México, México. He has written some articles for national and international congresses. He did research visits at the Paris School of Economics and the University of Glasgow. Currently, hir research interests are Game Theory and Mathematical Economics.

Santiago Omar Caballero Morales

Santiago Omar Caballero Morales is a professor-researcher in the Department of Logistics and Supply Chain Management at Universidad Popular Autónoma del Estado de Puebla (UPAEP) in Mexico. In 2009, he received a Ph.D. in Computer Science from the University of East Anglia in the United Kingdom. Since 2011 he has been a member of the National System of Researchers (SNI) in Mexico. His research interests are quality control, operations research, combinatorial optimization, pattern recognition, analysis and simulation of manufacturing processes, and human-robot interaction.

Day-Ahead Price Forecasting for the Spanish Electricity Market

Álvaro Romero, José Ramón Dorronsoro, Julia Díaz *

Universidad Autónoma de Madrid and Instituto de Ingeniería del Conocimiento, Madrid (Spain)

Received 11 November 2017 | Accepted 20 February 2018 | Published 27 April 2018

Abstract

During the last years, electrical systems around the world and in particular the Spanish electric sector have undergone great changes with the focus of turning them into more liberalized and competitive markets. For this reason, in many countries like Spain have appeared electric markets where producers sell and electricity retailers buy the power we consume. All agents involved in this market need predictions of generation, demand and especially prices to be able to participate in them in a more efficient way, obtaining a greater profit. The present work is focused on the context of development of a tool that allows to predict the price of electricity for the next day in the most precise way possible. For such target, this document analyzes the electric market to understand how prices are calculated and who are the agents that can make prices vary. Traditional proposals in the literature range from the use of Game Theory to the use of Machine Learning, Time Series Analysis or Simulation Models. In this work we analyze a normalization of the target variable due to a strong seasonal component in an hourly and daily way to later benchmark several models of Machine Learning: Ridge Regression, K-Nearest Neighbors, Support Vector Machines, Neural Networks and Random Forest. After observing that the best model is Random Forest, a discussion has been carried out on the appropriateness of the normalization for this algorithm. From this analysis it is obtained that the model that gives the best results has been Random Forest without applying the normalization function. This is due to the loss of the close relationship between the objective variable and the electric demand, obtaining an Average Absolute Error of 3.92€ for the whole period of 2016.

Keywords

Machine Learning, Big Data, Electric Market, Predictive Analysis, Prices, Random Forest.

DOI: 10.9781/ijimai.2018.04.008

I. Introduction

Electric sector and in particular the Spanish electric market is highly complex but at the same time fundamental to be able to maintain the contemporary way of life. The market pool is where the energy that reaches our homes and industries is purchased and where the electricity produced in our power plants is sold.

The Spanish electricity system has undergone a process of transformation through a liberalization of it that began in 1997. All the tasks related to the supply of electricity such as generation, transport, distribution, retail and economic and technical management of the system, have been separated.

Specifically, the spot market for electricity, managed by the OMIE [20], provides participating agents with the possibility of contracting electricity in seven sessions: the first and main, the Daily Market, and six subsequent sessions, belonging to the so-called intraday market, distributed throughout the day. It is in the first of the sessions, the Daily Market, in which this paper will focus. In this market, a price per hour is established at which each MWh of energy will be sold and purchased.

The way to establish the price follows the algorithm Euphemia that emerged in the initiative "Price Coupling of Regions" (PCR) by seven European electricity markets, among which is the Spanish one.

* Corresponding author.

E-mail addresses: alvaro.romero@iic.uam.es (Á. Romero), jose.r.dorronsoro@iic.uam.es (J. R. Dorronsoro), julia.diaz@iic.uam.es (J. Díaz).

This algorithm calculates the prices of electric energy efficiently, pursuing the maximization of welfare, which is defined as the surplus or profit, both of buyers and sellers, while optimizing the use of available capacity in interconnections.

For this welfare maximization, for both the daily and intraday markets, the Euphemia algorithm considers aggregate step curves.

In summary, the companies in charge of the generation make their offers (quantity of energy and price) and the companies in charge of retail, direct consumers, etc. demand the necessary energy at a certain price.

Once the bids are made, they are ordered according to price, in increasing order in the case of sales and decreasing order in the case of the purchase. The intersection of the supply and demand curves is called the matching point. In principle, this is the point that optimizes welfare and, therefore, establishes the price of energy for that particular hour.

All the energy offered and demanded at a price less than the matching point will be exchanged at that price, while the one with a higher price will not. This process is repeated for each of the 24 hours of a day.

II. Descriptive Analysis

The price of energy can be affected by many factors that are very complex and in some cases over which there are no data or reflect complex business strategies of companies that are not revealed to the general public.

A. Time Series Analysis

The analysis of the time series of prices indicates a strong seasonality in the data due mainly to the effect of the demand. If we analyze the price in the days of the week we can see clearly in the histogram of Fig. 1 that the days with a different price distribution are the weekend days, in which the prices are usually lower. Every day they have most of the prices around €45 but in the case of working days there are a good number of hours with prices higher than those €45, while in the case of Saturdays and Sundays there is a number much more reduced of those hours. One can also see a greater weight of hours with prices below €30 on Saturdays and Sundays, although a little higher in the latter.

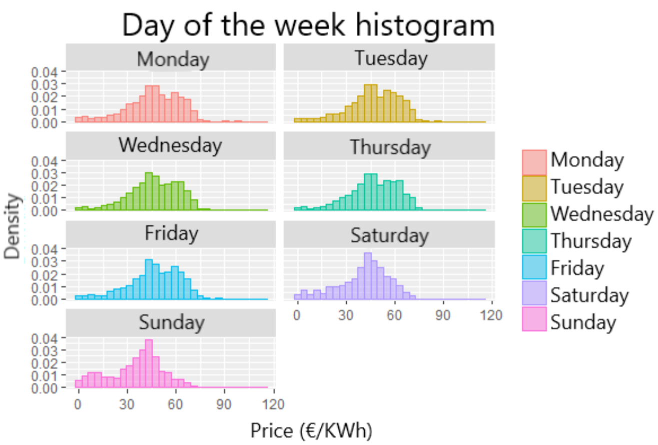


Fig. 1. Histogram by day of the week.

If we analyze the prices on an hourly basis, we can confirm the previous analysis that indicated that there were cheaper hours on weekends than on weekdays. As it happened in the daily case, in the hourly analysis the prices follow the trend of the electricity demand. So, during the night, there is less demand and, therefore, prices are lower. In Fig. 2, which has been made using the hourly average of the prices between 2014 and 2016, it can be seen that the drop in prices goes from 21:00 until approximately 5:00 where prices start to rise until 9:00, very pronounced on weekdays and until 10:00 and, not so pronounced, on weekends. The prices are approximately constant until the lunch hours where they begin to decrease. This valley comprises approximately from 13:00 to 17:00. Finally, prices reach their highest level around 20:00 – 21:00 on weekdays and 21:00-22:00 on weekends.

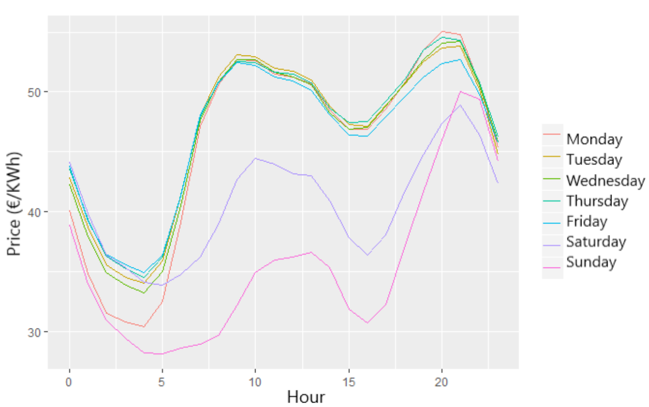


Fig. 2. Hourly mean by day of the week.

If we analyze the prices on a monthly basis as in Fig. 3, we see that in the summer months the prices reduce their hourly volatility and that the differences between weekends and midweek are maintained throughout the different months. The effect of the reduction of hourly volatility in the summer months is probably caused by the effect of refrigeration systems, the effect of tourism and the reduction of production in the industrial sector. Finally, in this price study it is interesting to observe what happens in the holidays since the patterns of electricity consumption in these days in general are clearly different from the working days.

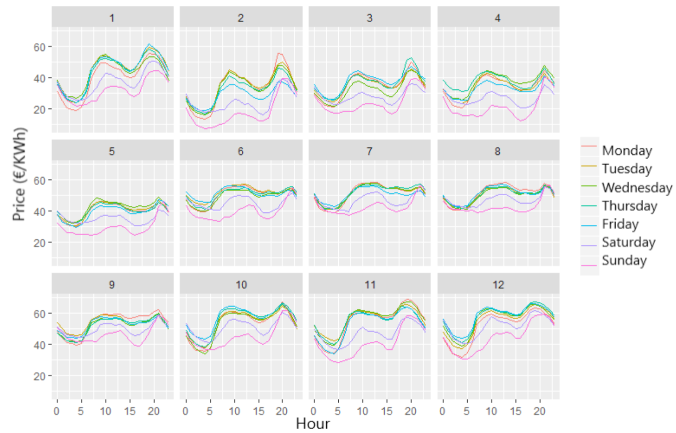


Fig. 3. Hourly Mean by week day and month.

To complete this analysis, only the national and Autonomous Community holidays, which are published in BOE (the Spanish Official Gazette), have been taken into account, but not those related to the individual cities because the effect of a holiday in small regions disappears when it is aggregated with the rest of the Spanish population. As one of the products of this analysis, the holiday coefficient for day i has been defined as the ratio of population in Spain on holidays on day i. In other words,

$$CF_i = \frac{\sum_{C \in \Omega_{fest_i}} P_C}{\sum_{C \in \Omega} P_C} \tag{1}$$

where Ω is the set of Spanish Autonomous Communities, Ω_{fest_i} is the Communities which are on holiday on day i and P_C the population of the Autonomous Community C. This coefficient allows us to see, in Fig. 4, the differences between holidays ($CF \neq 0$), working days ($CF = 0$ and day of the week other than Sunday) and Sundays (in black). It is observed that holidays have lower prices than working days and that in general the higher the holiday coefficient (therefore, the greater percentage of the Spanish population is on vacation) the lower the price.

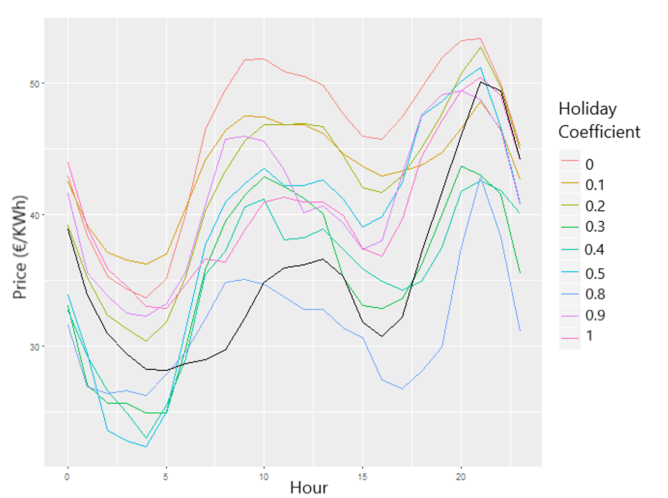


Fig. 4. Comparison of the price per hour among different holiday coefficients.

B. Variable Analysis

The variables for the price prediction problem most used in the literature [1] can be divided into the following categories: generation of renewables, electric demand, exports/imports, other countries' price, weather variables, economic variables, day type and hour.

Renewables generation. Renewables in Spain participate in the market at price 0 hence, they always lower the prices. Therefore, it is important to take them into account when predicting the price. Now, we will not have the real data for the day for which we have to give the prediction. Therefore, it is necessary to make predictions of production that will be those that enter as variables to the model. The two most important renewable generation technologies in Spain and which, therefore, should be taken into account as input variables for the model are:

- Wind energy.
- Solar energy (photovoltaic and thermosolar)

Demand. As well as the production of renewables, demand is a variable that directly affects energy prices since the more demand there is, the higher purchase offers and, therefore, the price rises.

Exports/Imports. Exports and imports are the interchange of electric energy between two countries, in the Spanish case, mainly with France. This interchange affects the price because there is more or less energy in the market and they occur due to the price of electricity in other countries.

Other Countries' Price. Because the Euphemia algorithm takes into account the price of other countries this variable is very important. Specifically, for the Spanish case the one that affects most the Spanish price is the French one, which is the main exporter of electricity to Spain. For this reason, as we will see later, there is a great correlation between the French and Spanish prices.

Weather variables. The meteorological variables affect both the production of renewables and the demand and, therefore, it will be necessary for the price predictions the usage of meteorological variables.

Economic variables. There are economic variables that affect the price of energy because they are related to demand and the production of non-renewable energy such as GDP, the price of gas and oil, etc. The difficulty of finding these variables for the necessary period of study and with sufficient granularity have meant that they are left for future work. In addition, these variables usually explain longer-term components, since they affect periods of months or even years and in this work we are more focused on the short term.

Type of day and time. The price depends on the type of day we face. This can be intuited easily, because the habits of electricity consumption are not the same in winter than in summer, or in different hours, etc. That is why all this must be taken into account when creating the model. In this study the following variables will be used:

- Hour.
- Weekday.
- Day of the month.
- Month.

In addition to the primary variables discussed above, other variables that may be of great interest have been generated in this work. In the first group of these variables are the past values of the price of electricity in Spain and France. These variables represent what happened at the same time one, two and three days before. These values are very important because of the hourly seasonal component that the price has and that was detailed before. There is a second group in which temperature appears. The temperature was available in the capitals of each province of Spain. To have a single value for the whole country, a weighted average based on population has been created so that the temperature of bigger provinces such as Madrid or Barcelona will have more relevance in the variable than temperatures in smaller ones such as Teruel or Soria.

The temperature has an effect on consumption that is not linear. Therefore, in this work the following temperature definitions have been studied:

$$\boldsymbol{T_2} = |T - \bar{T}| \tag{2}$$

$$\boldsymbol{T_3} = \begin{cases} 0 & T < 10 \\ T - 10 & T \geq 10 \end{cases} \tag{3}$$

$$\boldsymbol{T_4} = \begin{cases} -T + 15 & T < 15 \\ 0 & 15 \leq T < 20 \\ T - 15 & T \geq 20 \end{cases} \tag{4}$$

$$\boldsymbol{T_5} = \begin{cases} -T + 10 & T < 10 \\ 0 & 10 \leq T < 15 \\ T - 15 & T \geq 15 \end{cases} \tag{5}$$

$$\boldsymbol{T_6} = \begin{cases} -T + 5 & T < 5 \\ 0 & 5 \leq T < 20 \\ T - 20 & si\, T \geq 20 \end{cases} \tag{6}$$

$$\boldsymbol{T_7} = \begin{cases} -T + 5 & T < 5 \\ 0 & 5 \leq T < 25 \\ T - 25 & T \geq 25 \end{cases} \tag{7}$$

Here T is the single value of temperature explained above and $\bar{T}$ the average of this temperature during the whole period.

With the analysis of correlations, we can highlight the greater importance of some variables that we have commented previously. In descending order, the variables with the highest linear correlation are: the price in Spain 24, 48 and 72 hours before, wind generation, the price in France and electricity demand.

III. Predictive Analysis

This paper intends to apply some of the most used techniques in regression problems, more specifically, those that have been recently used in the field of predicting the price of energy with certain changes in the treatment of data. This publication serves as a modern benchmark to be measured against, since electricity markets have changed enormously in recent times.

As can be seen in [1] and in [2] the approaches taken by different authors in the past for the problem of predicting the price of energy have been diverse. These methods range from Game Theory to Computational Intelligence, through Simulation Models and Time Series.

In our study we will focus mainly on techniques related to Time Series and Computational Intelligence because these techniques are those that have been supported by researches such as [3] where it is indicated that the Economic and Game Theory methods are a good approximation but certainly insufficient in case of wanting to make precise short-term predictions.

A. Time Series

Time Series are successions of values spaced in constant periods of time; that is, the phenomenon is observed in moments taken regularly.

The analysis. of time series aims to model the underlying temporal structure in the observations taken in a certain period of time. Once the modeling is done, these algorithms serve to understand the time series but also to predict their behavior in the future. In the bibliography there are many references in which these models are used for price prediction such as [4] which uses Leipzig Power Exchange data for their experiments and ARMA models with some modifications. The work in [5], one of the first predictions of the price of electricity in Spain after the regulatory change, makes use of a Seasonal ARIMA model with different parameters for the Spanish and Californian market. In [6] and in [7] an improvement is proposed using a wavelet transformation prior to the use of the ARIMA model to reduce the volatility of the time series that is applied to electricity price data; In general, this transformation offers better results than a traditional ARIMA. There are also numerous studies that combine (S) ARIMA models with other prediction models, that use exogenous variables [8] or that make a model per day of the week or even per hour.

B. Machine Learning Models

Machine Learning is a field of Artificial Intelligence whose definition is complex and in which the different authors do not agree but we could define it as the subject that studies the techniques and algorithms that allow machines to adapt to dynamic situations and, therefore, somehow learn to predict the future, from discovering underlying patterns in the data.

The focus of this work is on supervised algorithms (data are fully labelled); more specifically, the case of price prediction belongs to a type of supervised models called regression models because the label to predict is a real value that goes, in this case, from 0 up to €180,3 (which are the minimum value and the maximum value for the prices in the Spanish market).

Some of the most used techniques for regression problems are Multiple Linear Regression, Decision Trees, K-Nearest Neighbours, Support Vector Machines, Neural Networks and ensembles that use some of the above models together to get better predictions.

1) Ridge Regression

Traditional Linear Regression has problems when there is not independence among the variables. Specifically, when there is collinearity Linear Regression does not work correctly. To avoid this, or somehow eliminate these collinearities with there are several techniques. For example, the method of Analysis of Principal Components, dimensionality reduction technique that generates orthogonal variables [9]. Another widely used resource is the use of Ridge Regression, proposed by Hoerl and Kennard [10] which introduces a regularization term in order to avoid overfitting and underfitting.

2) Decision Trees

Decision trees are a nonparametric supervised method that can be used for both classification and regression. It is a method widely used and described in depth in different references as, for example, [11].

The objective of this algorithm is to create a model that predicts the value of the objective variable by learning basic rules inferred from the variables of the data and that define regions whose edges are always parallel to the axes. Within each region a simple function is assigned, sometimes a constant.

3) K-Nearest Neighbours

The method of the K-Nearest Neighbours (K-NN) is based on inferring the variable to predict using the K cases in the training set that are more similar to the new data. The number K of neighbours to use in training can be defined by the user and, therefore, must be hyperparametrized since changing the number of neighbours to use can improve or worsen the results of the algorithm.

4) Neural Networks

Artificial Neural Networks are a type of Machine Learning algorithms that are inspired by the neuronal functioning of living beings. A Neural Network contains several processing units that connect to each other forming different architectures. Each unit or artificial neuron simulates the functioning of a neuron: it is activated if the total amount of signal it receives exceeds its activation threshold. In this case, the node is activated and emits a signal to the rest of the adjacent neurons. Therefore, each unit becomes a transmitter of the signal that can increase or decrease said signal.

Neural Networks are widely used for the problem of price prediction. The most relevant papers for this study are:

- [12], which is one of the first studies on the subject and in which an architecture is used with 15 input parameters, 15 hidden units and 1 output to predict the price for the Victorian Power System.
- [13] where data from the Spanish market are used to make a comparison between several models; in particular, one of the proposed is a multilayer perceptron with an architecture of one hidden layer and making use of wind and demand as predictor variables.
- [14] is also relevant where they use a combination of networks to predict the maximum, minimum and average value that is finally provided to 5 main neural networks to predict the price.

5) Support Vector Machines

The SVM is a model that started being used for classification and that generates a hyperplane that separates the two classes in an optimal way. In the case of regression, it is usually called SVR and its basic idea is to map the training data to a high dimension feature space through a non-linear mapping where we can perform a linear regression using a special loss function called ε-intensive loss.

SVMs are used for the prediction of prices in a large variety of publications, as in [15] which makes a comparison between Neural Networks and SVM or as in [16] that uses them to make accurate predictions and also to provide a confidence interval. There are several proposals for hybrid models using SVMs as in [17] that makes a hybrid model with SVM to capture non-linear patterns and ARIMA; this hybrid model is called in this publication SVRARIMA.

6) Random Forests

The Random Forest algorithm is a very effective ensemble of trees and is widely used today. Each tree is generated by random sampling with replacement of the original train set. The algorithm for Random Forest can be written as follows [18]:

1. Subsamples of the original data are chosen by bootstrapping.
2. For each one of the sub-samples, a regression tree is built without pruning, but in each node, instead of choosing the best possible cut among all the attributes, a subset of them is chosen randomly.
3. New data are predicted by the aggregation of all individual predictions using a mean in the case of regression.

For this method there are not many references for price prediction and most are very modern; in particular, [19] can be highlighted and it uses Random Forest for the prediction of the price in the Electricity Market of New York and uses as predictors the temporary series of prices itself lagged 3, 24, 168 and 720 hours, the demand, the temperature and a day of the week indicator.

IV. Training and Prediction Process

In order to obtain the most accurate model possible for predicting the price of the Spanish market, the following procedure was applied in this work: a normalization of the target variable and a normalization of the rest of the variables for those models that need it, a hyperparameterization of the selected models, a study of the errors and a deeper study of the best model to diminish its errors.

A. Data Normalization

Because with the study of the time series we discovered a great difference between working days and non-working days, a normalization of the price is proposed to achieve a reduction in variance. For this we must define two types of standardization: normalization of non-working days and hourly normalization. On the one hand, the normalization of non-working days conceptually aims to eliminate the effect of a decrease in the price of non-working days caused by the decrease in work activity. Mathematically normalization of non-working days can be defined as:

$$p'_{lh} = p_{lh} - \overline{p_{lh}} \quad (8)$$

$$p'_{sh} = p_{sh} - \overline{p_{sh}} \quad (9)$$

$$p'_{dh} = p_{dh} - \overline{p_{dh}} \quad (10)$$

$$p'_{fh} = p_{fh} - \overline{p_{fh}} CF_h \quad (11)$$

where p_{lh} is the hourly price of the working days, p_{sh} is the hourly price of the Saturdays, p_{dh} is the hourly price of the Sundays, p_{fh} the hourly price of the holidays and CF is the holiday coefficient. And Hourly normalization is then defined as:

$$p'_h = p_h - \overline{p_h} \quad (12)$$

where p_h is the price at hour h and $\overline{p_h}$ is the mean of every price at hour h.

For some algorithms is also important to consider the normalization of the rest of the variables. This normalization has been carried out by tuning the normalization function.

Therefore, for all the models in which it is necessary to normalize the attributes in some way, the normalization function has been hyperparametrized, always doing a grid search on the following four types of normalization:

- **MaxAbs scaler**. Scale each attribute by the maximum value that attribute can take.
- **Robust scaler**. This method uses robust statistics. Therefore, it subtracts the median and use the interquartile range to scale the data.
- **Standard scaler**. It substracts the mean and scales to obtain unit variance.
- **MinMax scaler**. It is usually used as an alternative to the previous one and is mathematically defined as follows:

$$z_i = \frac{x_i - min(x)}{max(x) - min(x)} \quad (13)$$

B. Parameter tuning

In the process of choosing the best model there are two jobs; on the one hand, the selection of the prediction algorithm and, on the other hand, we need to find the best values for the several hyperparameters that the different algorithms might receive.

To carry out the selection of the best parameters, the hyperparametrization process is applied, which basically performs trainings with a given set of train data and predictions for the validation set. In this case, because we are working with temporal data, we have opted for the use of a temporary validation following the scheme of Fig. 5.

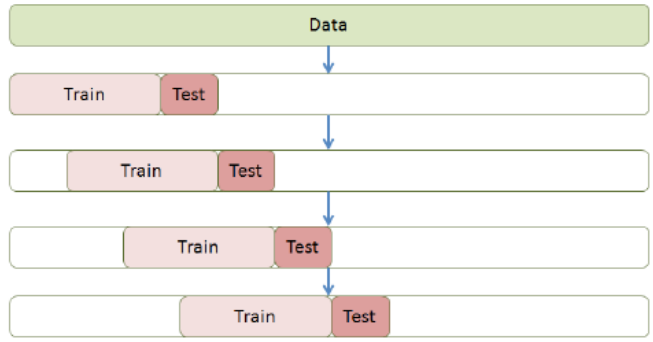


Fig. 5. Validation Workflow of this work.

In this scheme a cross validation is carried out in a continuous way so that it is impossible to take values from the future to predict past values as it could happen in a usual cross-validation as we see in Fig. 6.

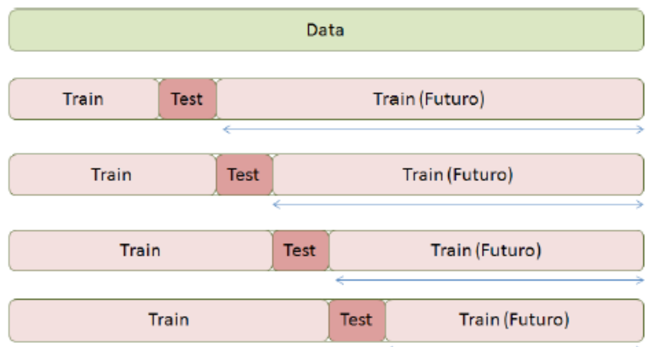


Fig. 6. Traditional cross validation.

Also, it should be noted that the test period is always a full day and the number of days taken for each train period has been hyperparametrized, going from 10 days at least up to 1 year, to then choose the most appropriate one. In general, the results show that it usually affects the chosen train period, being optimal for 7-9 months approximately. This is due to the fact that if one does not consider enough cases to train the model, it does not have enough information to generalize but one cannot take a very long period back because it would find cases that may have been produced by past macroeconomic facts that we are not taking into account in this work.

V. Results and Discussion

In this section we will discuss in detail the results of all the tests created for the prediction of the price in Spain. The period of validation is 2016 and the process for every model has been the same, training and predicting every day. For the benchmark, more than 20,000 models have been created with the different parameters discussed above.

A. Benchmark

In particular, the best results of each of the models are presented next: In the case of the Ridge Regression, the parameters that have given the best result have been α = 1.01, with 90 days for training and normalizing the data with maxAbs normalization. With this configuration we see that the average absolute error in validation is around €5.73 (see Table I).

TABLE I.
MAE by Month an Annual for the Best Ridge Regression Model

Jan	Feb	Mar	Apr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dec	Total
5.79	7.07	5.40	5.87	6.67	5.49	5.92	5.54	4.65	5.90	4.98	5.56	5.73

With the K-NN we find the best conjuration makes use, in the same way as Ridge Regression, of 270 days for the training and the robust function for the normalization of the data and 20 neighbours. With these parameters a reduction of the MAE with respect to the previous value is achieved, obtaining €5.30 in validation (see Table II).

TABLE II. MAE by Month an Annual for the Best K-NN Model

Jan	Feb	Mar	Apr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dec	Total
5.70	7.71	5.35	5.41	5.21	6.19	4.85	4.41	3.49	5.34	5.53	4.57	5.30

In the case of the Multilayer Perceptron, using the regularization parameter $\alpha = 1$, the robust normalization and, as has been said before, two hidden layers with 50 neurons, an error of €5.29 in validation is obtained (see Table III).

TABLE III. MAE by Month an Annual for the Best MLP Model

Jan	Feb	Mar	Apr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dec	Total
4.73	6.01	5.19	5.19	5.58	4.83	5.23	5.94	4.35	5.40	5.89	5.20	5.29

In the case of Support Vector Machines for Regression, an error minimization of €4,92 in validation is reached (see Table IV), when C takes the value of 100 $\varepsilon = 1$, $\gamma = 1$, with normalization minMax and with 210 days for training, where C is the penalty parameter of the error term, γ is the kernel coefficient and ε specifies the epsilon-tube within which no penalty is associated with points predicted within a distance ε from the actual value.

TABLE IV. MAE by Month an Annual for the Best SVR Model

Jan	Feb	Mar	Apr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dec	Total
5.63	6.60	4.06	4.75	4.77	4.60	4.44	4.00	5.22	4.62	5.69	4.63	4.92

Finally, the model that has given the best result has been the Random Forest with n_estimators = 710, min_samp_leaf = 2, max_feat = 0.4 and 270 training days, where n_estimators. Is the number of trees, min_samp_leaf is the minimum number of samples to be in a leaf node and max_feat is the number of features that considers in each split. With all this, a MAE of €4,44 in validation has been achieved.

B. Best Model Analysis

One of the most interesting properties of the best model, Random Forest, is that it allows us to calculate the relevance of the variables. This relevance is taken as the number of times that attribute appears in each of the created trees. Therefore, in our case, as we do a daily training, in each one we can calculate the variable frequency of appereance in each tree. Throughout the validation year, we have 365 frequencies for each variable, so if we take the average of all that series we have the average frequency of each variable. With these means we have a good measure of how important each variable is throughout the entire year. In Fig. 7, Fig. 9 and Fig. 12 the variables are numbered as follows:

0. Demand
1. Solar production.
2. Wind Production.
3. Price in Spain 24 hours before.
4. Price in France 24 hours before.
5. Price in Spain 48 hours before.
6. Price in France 48 hours before.
7. Price in Spain 72 hours before.
8. Price in France 72 hours before.
9. Holiday coefficient.
10. Month.
11. Day of the month.
12. Day of the week.
13. Time.
14. Week of the year.
15. Type of day. That takes the values of 1 if it is Saturday, 2 if it is Sunday, 3 if it is a holiday and 0 in other days.
16. Working. That takes the values of 0 if it is Saturday, Sunday or holiday and 1 the other cases.
17. T
18. T2
19. T3
20. T4
21. T5
22. T6
23. T7

1) RF with Normalized Price

Using the technique described above to calculate the relevance of the variables, we can see that the most important ones are wind energy, the price in Spain 24 hours before and the temperature in the T5 version. The low relevance that the demand obtains is surprising because the intersection point between demand and generation will be higher or lower depending on the total amount of energy demanded. This is clearly due to the normalization explained before made to the target variable; in this normalization the effect of the hours and the days of the week is eliminated and if it is not applied in the same way to the demand it causes the demand to stop being related to the prices.

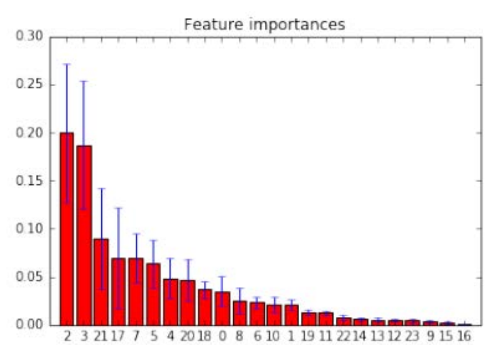


Fig. 7. Feature importance with price normalized.

If we observe the comparison between the hourly chart of demand and the price when we normalize in the manner described before, we see in Fig. 8 that any relationship between both variables has been lost. Therefore, it is necessary to consider what happens when we normalize both variables following the same method, analysis that is carried out in the following subsection.

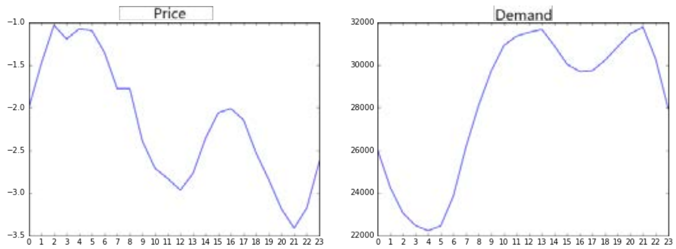


Fig. 8. Comparison between the hourly averaged Demand and the hourly averaged Normalized Price.

2) RF with Demand and Price Both Normalized

If we re-execute the hyperparametrization by normalizing the demand, we see that the results improve somewhat with respect to the previous case, not only in total, but also month by month, as we can see in Table V.

TABLE V. MAE by Month for RFR with Demand and Price Normalized

Jan	Feb	Mar	Apr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dec	Total
5.03	6.19	4.00	4.66	5.29	4.29	3.90	3.45	2.99	3.90	4.22	3.55	4.28

In addition, if we look at Fig. 9, where the most important variables of the best model are calculated, we observe that the demand is now in sixth position and the wind generation and the price 24 hours before are still in the lead.

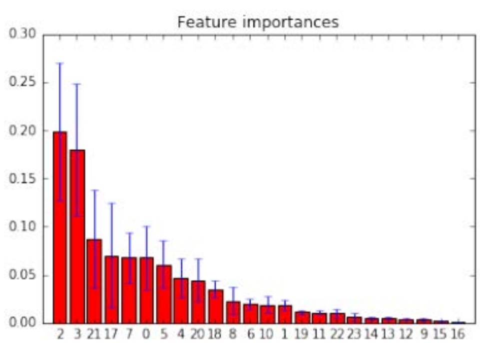


Fig. 9. Feature importance with normalization on both demand and price variables.

Somehow, by normalizing both variables in the same way, we are better preserving the common structure that they have. In addition to checking the importance, we again review the comparison of the two new demand and price variables; in Fig. 10, we observe that although the time relation between both is better preserved, this one is improvable if we compare it with the existing relation between the two variables without normalizing as we see in Fig. 11.

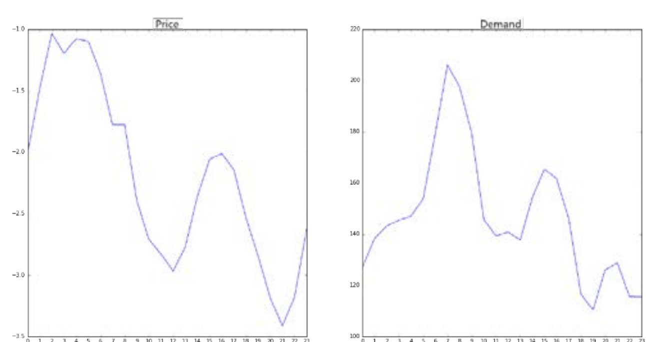


Fig. 10. Comparison between the hourly averaged demand and the hourly averaged price, both normalized.

3) RF without Normalization

When not normalizing, in the comparison between the hourly evolution of both variables of Fig. 11, a strong relation between both is observed.

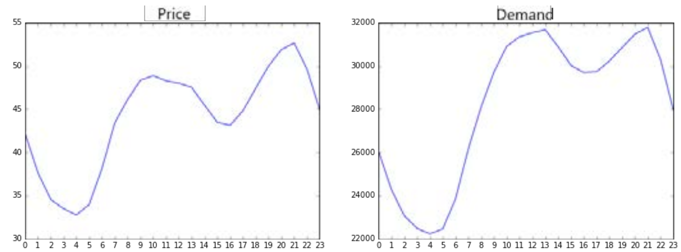


Fig. 11. Comparison between the hourly averaged demand and price.

With the hyperparametrization of Random Forest using the data without normalizing one can observe a remarkable improvement. Therefore, we can conclude that the best option is not to normalize when using Random Forest not to lose any underlying relationship between the different attributes and the price even though we have a greater variance in the target variable. Regarding the importance of the variables, they have changed radically. The demand is now in first place followed by the price 24 hours before and the wind generation, and the others have much less relevance, as we can see in Fig. 12.

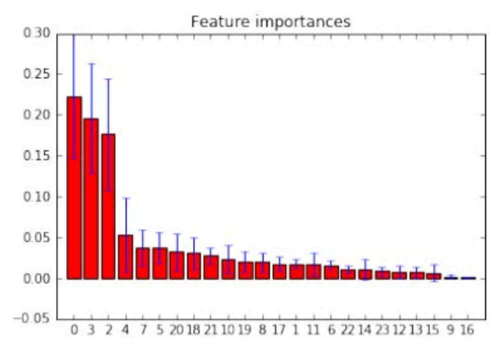


Fig. 12. Feature importance with no normalization.

By adding this improvement, we can observe a greater precision in each of the hours of the months, appreciable in the summary of the whole year that we have in Fig. 13, where softer colors can be observed in the case where the demand is normalized and even softer in the heat map corresponding to the data without normalizing. In addition, with these figures we can understand that in general there are some hours and months for which it is more difficult to predict. For example, in the summer months, especially in August and September, errors are much lower for all hours of the day than the rest of the year.

The errors of the best resulting model, Random Forest without normalization, are found summarized in Table VI.

TABLE VI. MAE by Month for RFR without Normalization

Jan	Feb	Mar	Apr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dec	Total
4.67	6.67	3.80	3.75	4.32	4.53	3.00	2.34	2.83	4.01	4.24	2.93	3.92

They are difficult to compare with those obtained by other authors in part because a large number of publications are old and the market has profoundly changed as it happens with [17], [6] or [5]. In fact, the appearance of participating agents has increased in great quantity in the last two or three years.

On the other hand, in many publications measures of certain months, weeks or even days are taken. For example, [6] and [5] measure the error using weeks of different periods. For its part, [19] measures the error on certain days of June. Even other publications such as [15], [17] make predictions for markets other than Spanish that are difficult to compare.

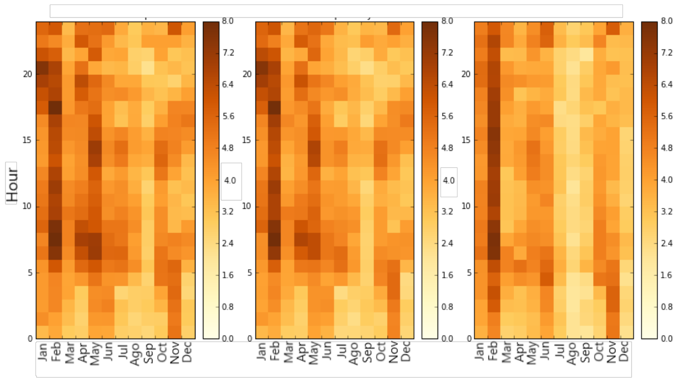


Fig. 13. Summary of errors by month and hour.

C. Results with Test Data of RF without Normalization

As the test data, the available period of 2017 has been reserved, that is, from January to July, both included. In it, the use of the best model, Random Forest has been tested without normalizing, so that it is trained every day and predicts for the next day. Following this methodology, the result is somewhat worse than for the evaluation period as is usual in this type of problem. This is due to the fact that in some way when performing the hyperparametrization the model is being adjusted to learn the validation set.

More in detail, the test error is €4.50 and broken down by months, we can see it in Table VII, where we observe that the error is degraded, especially since April. Despite this, in the actual operation of a product like this, there are two maintenance elements of the model that would improve their predictions. In the first place, there would not be a single model, but there would be several that are in follow-up and that periodically could change because there is a model that is offering better results than the one that is in production. Secondly, that starting in April, the model gets worse. It may be due to a need for re-hyperparametrization.

TABLE VII.
Test MAE by Month for RFR without Normalization in 2017

Jan	Feb	Mar	Apr	May	Jun	Jul	Total
3.30	2.92	4.60	6.28	4.60	8.01	3.25	4.50

VI. Conclusion and Future Work

In this work, different models have been implemented with the final objective of solving the problem of price prediction in the electricity market. For this, several stages have been followed.

In the first place it has been necessary to review exhaustively the different models and variables present in the bibliography that have been used for price prediction. It should be noted the difficulty in the comparison with the publications cited in this paper because of the temporal and geographic difference of the publications regarding this work.

For the choice of the model, a benchmark has been carried out between different types of models and parameters among which are Ridge Regression, K-Nearest Neighbours, Multilayer Perceptron, Support Vector Machines and Random Forests. Of all the combinations of models and parameters, the most precise has been Random Forest and for it a more detailed study has been carried out, including an analysis of the most relevant variables, a comparison of different types of normalization and an exploration of errors by months and hours.

In this work we have also seen that if we normalize the time series of prices to reduce their variance, it can cause a loss of information about the underlying patterns in the data that are very useful for prediction. In particular, it has been observed how normalization blurred the strong relationship that the price has with the demand, which is partially solved by applying the same standardization treatment to both variables. Despite this improvement, the results provide an unfavourable outcome to normalization since without it, the best results are obtained.

The best predictive model achieved, has managed to obtain a MAE in validation of 3.92 and of 4.50 in test, a result that is already useful for all the agents that participate in the market.

On the other hand, a possible future line of research is the influence of economic factors, among which one can include the price of the most used raw materials for obtaining energy (coal, oil and gas), the Gross Domestic Product, etc.

Another very important aspect to consider is the interconnections with the bordering countries. In the case of Spain, the most important connection is the French one and, therefore, the price in France was included but in addition to this price, the technical capacity of exports and imports as well as the quantities imported/exported are very important.

Labour is a key aspect in the prediction of prices due to the effect of holidays, Saturdays and Sundays on the demand for electricity. Therefore, this is a key aspect that should be further investigated. In this work we have tried to solve it through a normalization that has been proven as not very useful. Therefore, other lines of research in this regard are, on the one hand, to perform a post-process of key days (holidays, Christmas and summer time, bank holidays, etc.) and, on the other, to make separate models for those days, with the problem of the scarcity of data.

Lastly, the most important and at the same time the most difficult aspect to include in the model is the strategy of the agents in the market. Despite its difficulty, there are data that are published by the OMIE that could be used to solve this problem.

Acknowledgments

With partial support from Spain's grants TIN2016-76406-P and S2013/ICE-2845 CASI-CAM-CM. Work supported also by project FACIL--Ayudas Fundación BBVA a Equipos de Investigación Científica 2016 and the UAM--ADIC Chair for Data Science and Machine Learning.

References

[1] Sanjeev Kumar Aggarwal, Lalit Mohan Saini, and Ashwani Kumar. Electricity price forecasting in deregulated markets: A review and evaluation. International Journal of Electrical Power & Energy Systems, 31(1):13-22, 2009.

[2] Rafal Weron. Electricity price forecasting: A review of the state-of-the-art with a look into the future. International journal of forecasting, 30(4):1030-1081, 2014.

[3] Derek W. Bunn. Forecasting loads and prices in competitive power markets. Proceedings of the IEEE, 88(2):163-169, 2000.

[4] Jesús Crespo Cuaresma, Jaroslava Hlouskova, Stephan Kossmeier, and Michael Obersteiner. Forecasting electricity spot-prices using linear univariate time-series models. Applied Energy, 77(1):87-106, 2004.

[5] Javier Contreras, Rosario Espinola, Francisco J. Nogales, and Antonio J. Conejo. Arima models to predict next-day electricity prices. IEEE transactions on power systems, 18(3):1014-1020, 2003.

[6] Antonio J. Conejo, Miguel A. Plazas, Rosa Espinola, and Ana B. Molina. Day-ahead electricity price forecasting using the wavelet transform and arima models. IEEE transactions on power systems, 20(2):1035-1042, 2005.

[7] Chang-il Kim, In-Keun Yu, and YH Song. Prediction of system marginal price of electricity using wavelet transform analysis. Energy Conversion and Management, 43(14):1839-1851, 2002.

[8] Rafal Weron, Adam Misiorek, et al. Forecasting spot electricity prices with time series models. In Proceedings of the European Electricity Market EEM-05 Conference, pages 133-141, 2005.

[9] Manuel Gurrea. Análisis de componentes principales. Proyecto e-Math Financiado por la Secretaría de Estado de Educación y Universidades (MECD), 2000.

[10] Arthur E. Hoerl and Robert W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1):55-67, 1970.

[11] Christopher M. Bishop. Pattern recognition and machine learning. Springer, 2006.

[12] BR Szkuta, L. Augusto Sanabria, and Tharam S. Dillon. Electricity price short-term forecasting using artificial neural networks. IEEE transactions on power systems, 14(3):851-857, 1999.

[13] Alberto Cruz, Antonio Muñoz, Juan Luis Zamora, and Rosa Espínola. The effect of wind generation and weekday on Spanish electricity spot price forecasting. Electric Power Systems Research, 81(10):1924-1935, 2011

[14] Raquel Gareta, Luis M. Romeo, and Antonia Gil. Forecasting of electricity prices with neural networks. Energy Conversion and Management, 47(13):1770-1778, 2006.

[15] Damien C. Sansom, Tom Downs, Tapan K. Saha, et al. Evaluation of support vector machine based forecasting tool in electricity price forecasting for Australian national electricity market participants. Journal of Electrical & Electronics Engineering, Australia, 22(3):227, 2003

[16] Jun Hua Zhao, Zhao Yang Dong, Zhao Xu, and Kit Po Wong. A statistical approach for interval forecasting of the electricity price. IEEE Transactions on Power Systems, 23(2):267-276, 2008.

[17] Jinxing Che and Jianzhou Wang. Short-term electricity prices forecasting based on support vector regression and auto-regressive integrated moving average modeling. Energy Conversion and Management, 51(10):1911-1917, 2010.

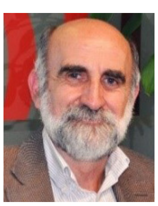
[18] Andy Liaw, Matthew Wiener, et al. Classification and regression by random forest. R news, 2(3):18-22, 2002.

[19] Jie Mei, Dawei He, Ronald Harley, Thomas Habetler, and Guannan Qu. A random forest method for real-time price forecasting in New York electricity market. In PES General Meeting| Conference & Exposition, 2014 IEEE, pages 1-5. IEEE, 2014.

[20] "OMIE" http://www.omie.es/inicio

Álvaro Romero Miralles

Álvaro Romero Miralles holds a Master's Degree in Computer Engineering and a Master's Degree in ICT Research and Innovation in Computational Intelligence a Degree in Mathematics and a Degree Computer Engineering from Universidad Autónoma de Madrid. Currently he works as Data Scientist and Project Manager at Health and Energy Predictive Analytics group of Instituto de Ingeniería del Conocimiento (IIC), a private Big Data R&D institution. He has experience in fraud detection, predictive maintenance, optimization problems among others. He collaborates as a professor in different business schools such as MBIT School and ENAE Business School.



José Dorronsoro

José Dorronsoro (PhD, Washington University in St Louis; USA) is Professor of Computer Engineering at the Universidad Autónoma de Madrid. He has authored more than 100 scientific papers in mathematical analysis, machine learning and applications and has directed a large number of research and innovation projects. Dr Dorronsoro is also a senior scientist at the Instituto de Ingeniería del Conocimiento (IIC), where he works on research and innovation on renewable energy.

Julia Díaz

Julia Díaz is Ms Degree in Mathematics, PhD in Computer Science both from Universidad Autónoma de Madrid (UAM-Spain) and General Management Program from IESE-Universidad de Navarra (Spain). At present she is Senior Innovation Manager in a private R&D+i institution named Instituto de Ingeniería del Conocimiento (IIC-UAM) dedicated to extracting knowledge on the basis of high volumes of heterogeneous data (Big Data) and optimizing business processes in areas such as healthcare and energy. She also is Part Time PhD Professor in Computer Sciences in the UAM and Professor in Big Data & Data Sciences Master in UAM and ESADE.

Biomedical Term Extraction: NLP Techniques in Computational Medicine

Antonio Moreno Sandoval[1], Julia Díaz[1], Leonardo Campillos Llanos[2], Teófilo Redondo[3]*

[1] Universidad Autónoma de Madrid (UAM)/Instituto de Ingeniería del Conocimiento (IIC) (Spain)
[2] Laboratoire d'Informatique pour la Mécanique et les Sciencies de l'Ingénieur (LIMSI-CNRS) (France)
[3] Ayming España (Spain)

Received 12 November 2017 | Accepted 4 February 2018 | Published 6 April 2018

Abstract

Artificial Intelligence (AI) and its branch Natural Language Processing (NLP) in particular are main contributors to recent advances in classifying documentation and extracting information from assorted fields, Medicine being one that has gathered a lot of attention due to the amount of information generated in public professional journals and other means of communication within the medical profession. The typical information extraction task from technical texts is performed via an automatic term recognition extractor. Automatic Term Recognition (ATR) from technical texts is applied for the identification of key concepts for information retrieval and, secondarily, for machine translation. Term recognition depends on the subject domain and the lexical patterns of a given language, in our case, Spanish, Arabic and Japanese. In this article, we present the methods and techniques for creating a biomedical corpus of validated terms, with several tools for optimal exploitation of the information therewith contained in said corpus. This paper also shows how these techniques and tools have been used in a prototype.

Keywords

Biomedical Terminology, Natural Language Processing, Term Recognition, Information Extraction.

DOI: 10.9781/ijimai.2018.04.001

I. Introduction

Terminology is a branch of Applied Linguistics whose main goal is the creation of specialized or technical language. Thematic domains are by themselves the realm of a specific sublanguage, adapted to designing the concepts in each topic or knowledge area. In this sublanguage, many exclusive terms coexist with those that have acquired meanings other than those common to the general language. Elaborating a terminological dictionary is a multidisciplinary task that requires contributions from both lexicographers and subject matter experts in order to define a specific term in the most precise way.

Some fields, that show a rapid evolution in the area, need to include new concepts at a very fast pace and require constant work in detecting those concepts and proceeding to normalize or standardize. Medical terminology is one such field where the sheer number of specialized terms exceeds the usual number of specialized terms in other knowledge areas, when taking into account both simple lemmas and compound forms. New terms and concepts are generated in a very dynamic fashion and this needs computing tools such as automatic recognizers (as part of the information extraction process). These applications analyze digital texts and identify candidates that can be terms of a given domain, so it can be validated by an expert (akin to a supervised learning process).

* Corresponding author.

E-mail address: teo.redondo@gmail.com

II. Basic Concepts and Techniques

A. Automatic Recognition of Terms and Concepts in Digital Texts

1) Objectives

Term Extraction or Automatic Term Recognition (ATR) is a field in language technology that involves "extraction of technical terms from domain-specific language corpora" [1], or identifying term candidates in texts of lists of words [2]. The original interest lies not in creating terminology resources, but in extracting words or expressions that identify topics in a document. This use is typical when working with medical texts [3], as a tool for information extraction and text mining [4]. Different NLP techniques are described in detail in Moreno Sandoval and Redondo, 2016 [5].

In order to detect new terms and concepts, texts that are recent and also representative are required. **Corpus Linguistics**, with an ever-growing influence in recent years due to the availability of large datasets, has the compilation of texts of a given domain as one of the main objectives. Documents must be digital, so searches or other computational handling can be performed, such as morphosyntactic annotation and statistical analysis. Once the medical corpus is created, the automatic recognizer will extract a number of candidate terms.

In Terminology there are well-established methodological traditions to enhance lexicography resources and build data banks following standard procedures [6]. However, the speed at which new terms (neologisms) are created in certain knowledge areas makes this approach extremely costly. It is at this precise point where systems for automatic extraction of terms are of great help, but always considering that the final "word" lies in the hands of the area expert.

2) Domain and Difficulties

In the classical definition of Terminology, a *term* or *terminological unit* is a linguistic expression of a concept in a specialized domain [7]. From the perspective of ATR, the task consists in identifying how a term is defined under the following lines [8]:

- *Unithood*: the degree of cohesion or stability of words in an expression.
- *Termhood*: the degree of specificity of the term with respect to the knowledge area. For instance, *hepatic* is related to a medical domain, not to aeronautics or space.

The main difficulties in *Unithood* are located in recognizing syntagmatic structures and the boundaries between words in compounds (*multiword terms*). For instance, the ATR should detect as candidate terms *infarto* (*infarct* or *heart attack*), *infarto de miocardio* (*myocardial infarct*) and *infarto agudo de miocardio* (*acute myocardial infarct*), but not *posible infarto* (*possible infarct*).

In *Termhood* it is typical to find polysemic terms that do belong in different knowledge areas. For instance, *nuclear* is a term both in Physics and in Genetics or Biology. Using resources of terms in other areas can lead to achieving wrong results.

In addition, there are two phenomena that make things more complicated in recognizing biomedical terms: *variation* and *homonymy*. In the former case, the problem appears when a knowledge area holds a great number of formal variations of the same term. This affects both simple terms (*aterosclerosis* ~ *ateroesclerosis*) and compound terms (*carcinoma microcítico de pulmón* ~ *carcinoma microcítico pulmonar*). Ananiadou and Nenadic [9] distinguish five types of terminological variation, that are basically just formal alternatives:

- Ortography: *alfa-amilasas* ~ *amilasa alfa* ~ *-amilasa*
- Morphology: *obsesiva-compulsiva* ~ *obsesivo-compulsivas*
- Lexicon: *infarto de corazón* ~ *infarto cardíaco*
- Structure: *virus del papiloma humano* ~ *papilomavirus humano*
- Acronyms and abbreviations: *SST* ~ *ST*, both referring to *somatostatina*

In addition to constant creation of neologisms in the biomedical area, foreign influence is sourcing new variations. Linguistic calques or loan translations with little or no adaptation to the new language are one such example. In biomedical texts in Spanish, terms like *bypass*, *by pass* and *baipás* appear quite naturally. Another example is the increasing inclusion of modifiers to already existing terms: *deficiencia de hexosaminidasa A* ~ *deficiencia total de hexosaminidasa A*. An essential task for both human experts and ATR is to normalize formal variations representing the same concept. The existence of multilingual ontologies and metathesaurus, such as those integrated in UMLS (*Unified Medical Language System*) [10], provide an essential contribution. This resource includes several thesaurus and terminological works: *Medical Subject Headings* (MeSH) [11], *Systematized Nomenclature of Medicine – Clinical Terms* (SNOMED-CT) [12], or version 10 of the International Classification of Diseases (ICD-10) [13]. UMLS contains unique identification codes associated to each terminology variation in different resources. For example, code C0817096 refers to *breast* or *thoracic cavity* in MeSH and also the term *thoracic* or *thorax* in SNOMED-CT.

On the other hand, term homonymy, especially acronyms, is another challenge for ATR. For instance, *IM* can refer to both *insuficiencia mitral* and *infarto de miocardio*. Without the contribution in contextual and domain knowledge from terminology experts it is very difficult to decide in which concept the acronym belongs. Some systems try to solve this by restricting the lexicon to a specific field [14], but in several cases, this presents problems since limits or boundaries between biomedical areas are rather fuzzy.

3) Approaches and Methods

Although several authors distinguish basically between linguistic techniques and statistical techniques [15], in term recognition several heterogeneous methods are combined so as to achieve the best results, as will be shown below. In a conventional way, the different approaches towards ATR are classified along four types: a) dictionary-based, b) rule-based, c) statistics-based and machine learning, d) hybrid [16].

- Dictionary-based approaches use digital resources such as grammar words without content (also known as *stop words*), as well as ontologies, glossaries and domain thesaurus. These lists allow the filtering of the text: with the former, words of no interest get eliminated and with the latter, terms are singularly identified. This approach is the most efficient and simple, but it tends to be rather incomplete and it is not available in all domains nor for all researchers. An example is detailed in Segura-Bedmar et al [17], where the UMLS metathesaurus and other name lists of generic drugs were used, with the objective of identifying and classifying pharmacological names in biomedicine texts.
- Rule-based approaches use pattern analysis of the term creation (for example, compounds by addition, hyphenated compounds, syntagmatic patterns) and grammar knowledge (morphological analysis of the terms, lists of lemmas and affixes). This approach has abundantly been used from 1990 onwards. Morphological description of lemmas and affixes, for instance, has been used to detect medical terms [18], and other researchers used concatenated category pattern-based algorithms [19]. For Spanish, noun phrases (or nominal syntagmas) have been used for medical terms extraction [20]. In general, an effective strategy can be achieved if work focuses on a language with Greek and Latin bases to create new terms. This, however, is not the case in all domains nor all languages [21].

With respect to statistics-based techniques, the foundation lies in measuring the degree of distinctiveness [22] of a word or lemma in a specialized context in contrast with their frequency in a general corpus. The two most common are the *log-likelihood ratio test* [23] and the *logDice* metrics used in The Sketch Engine [24]. The central idea of these techniques is to know which words or terms over- or under-used in the corpus for analysis when compared to the frequency of the same words in a reference corpus. In our case we take a corpus of medical terms (*MultiMedica*) and compare it to the *Reference Corpus of Current Spanish* (Corpus de Referencia del Español Actual – CREA), that contains a balanced set of texts coming from different domains and linguistic registers. However, there are other statistics-based techniques, such as Mutual Information Metric [25] or the use of Distributional Semantics and lexical collocation [26]. For Spanish, the experiment for term detection has been run on a corpus of scientific texts by using n-grams and their likelihood and distribution in such corpus [27]. An algorithm to analyze lexical, morphological, syntactic features has been used to compare this with a reference corpus [28].

Machine Learning's approaches are a special type of using statistical techniques that consist in training algorithms with data from corpus that has been previously annotated by experts in the knowledge area. Machine Learning algorithms (among others, *Hidden Markov Models* – HMM, *Support Vector Machines* – SVM, or *Decision Trees*) identify features in the annotated terms and apply them to a new data set. The most basic type is called *classifier*, that divides words in a text between terms and non terms. Lastly, current advances in neural network research are yielding promising methods for sequence modeling tasks (such as PoS or NER). Biomedical entity recognition is being enhanced through *Recurrent Neural Network* (RNN) models, namely *Long-Short-Term Memory networks* [29] and hybrid architectures combining *Conditional Random Fields* (CRFs) [30], attention mechanisms and language modelling [31], among others. These kinds of approaches use vector representation of words along with their occurrence context or frequency distribution (word embeddings) [32] [33].

Hybrid techniques combine two or more techniques mentioned above. The most usual case uses a linguistic approach (dictionaries and rules of term formation) and a statistical metric, a hybrid method already developed for Spanish [34].

III. BIOmedical NLP Use Case – MultiMedica

MultiMedica (Multilingual Information Extraction in Health Domain and its Application to Scientific and Informative Documents) was a coordinated project between the LABDA research group (UC3M), the GSI group (UPM) and the LLI (UAM), the latter group being in charge of the following tasks:

- Compilation of a specialized corpus of texts about health topics. The corpus gathers documents in three languages with different genetic and typological features: Arabic, Japanese and Spanish
- Morpho-syntactic tagging of the corpora,
- Contrastive research on term formation,
- Development of an automatic term extractor,
- Design of a web-based search tool.

A. The Corpus

The initial experiment used a corpus of text in Spanish, a corpus that was later extended to include text in Japanese and Arabic. The subcorpus consists of 4,200 documents with a total of 4 million words. The textual typology covers from general articles written by doctors with a no-specialist audience in mind (typically reviewed and edited by journalists) up to scientific texts for a specialized audience (i.e. healthcare professionals). Technical/specialized texts prevail over general content (more than 80% correspond to technical texts), with most of the medical specialties represented in a balanced number. This qualifies the corpus as a reliable source to produce a list of valuable candidate terms. As an interesting addition, the corpus was morphosyntactically annotated (category and lemma), in order to allow for searches and agreement [35].

The MultiMedica corpus has gathered 51,476 biomedical texts in different genres (popular and technical texts) written in Spanish, Japanese and Arabic. The tool enables two main functions: queries in the medical corpus and medical term extraction of an input text. The tool presents a web interface for ease of use.

Table I outlines the composition of the corpus (number of texts and words/characters):

TABLE I. Summary of the MultiMedica Corpus Data

Subcorpus	Documents	Word or characters
Japanese	3,746	1,131,304
Arabic	43,526	2,559,323
Spanish	4,204	4,031,174
TOTAL	51,476	7,721,801

The Spanish corpus is made up of three subcollections: The *Harrison* subcorpus assembles professional and scientific texts written by medical doctors; the *OCU-Salud* subcollection gathers journalistic texts written by medical doctors and edited by journalists; and finally, the *Tu otro médico* subcorpus collects popularized texts from encyclopaedic articles written by professional doctors for non-specialists. Regarding the Arabic corpus, gathering documents was made difficult by the fact that most medical doctors in the Arabic-speaking world write articles in English. Most documents in this subcorpus were articles and popularized news collected from *Altibbi*, a Jordanian medical website equivalent to *Healthline* in the United States. The remaining texts were drawn from the health sections of the following journals: *Al-Awsat* (from Saudi Arabia), *Youm7* (from Egypt), and *El Khabar* (from Algeria).

In relation to the Japanese corpus, only abstracts of five medical journals were collected, due, again, to the lack of availability of data. However, the texts gather contents on different specialties: Oriental medicine in Japan (from the journal *Kampo Medicine*), infectious diseases (*Kansenshogaku Zasshi*), liver diseases (*Kanzo*), otolaryngology, (*ORLTokyo*), and obstetrics (*Sanfujinka no shinpo*).

B. Methodology and Pipeline

We summarize some experiments carried out on ATR of medical terms (full details are explained in another paper) [36]. For the initial experiment only identifying simple terms (those with one single word, such as *aspirina* or *ADN*) or words as part of a compound (*ascórbico* in *ácido ascórbico*, or *Down* in *síndrome de Down*) was considered. The objective was to evaluate which of the previous strategies would provide the best results. The process followed three steps (see Fig. 1):

1. Preselect candidates by means of one of the three methods
2. Filtering of term candidates by means of a list of biomedical lemmas and affixes
3. Manual check of each candidate term by consulting bibliography or other resources

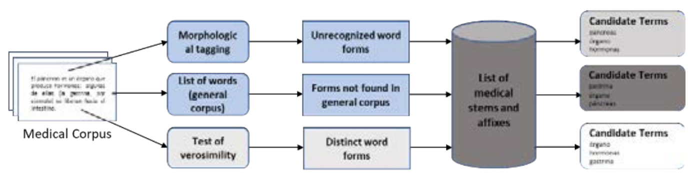


Fig. 1. Phases of the term extractor [36].

1) Preselect Terms Following each Method

Each method for term candidate extraction is not based on a similar strategy, and consequently the list obtained from each has a different size, although it is applied to the same data set. However, obtaining more candidates does not mean that the rate of success increases.

The first method uses a morphological tagger. It is an example of the rule-based type: the analyzer contains a set of recognition rules and analysis of words in Spanish. Here only words with the tag "unknown" (*desconocido*) are of interest, because medical terms are assumed to have a morphological structure not included in the analyzer used: GRAMPAL [37] covers a lexicon with more than 50,000 lemmas of general use and is capable of analyzing more than 500,000 inflection forms. Obviously, GRAMPAL contains a large number of medical terms that have found their way into the common lexicon, as would be collected in any reference dictionary (DRAE or Maria Moliner being the most typical ones). But similarly, most of the specific and technical terms of the domain are not included (i.e, *ADN* or *distal*). After an initial run over the corpus with 4 million words, a total of 22,413 "unknowns" were produced, which then were listed as term candidates.

The second method uses a corpus-based strategy: words in MultiMedica are compared with those in the Spanish general corpus (CREA). Given that it is a large and balanced corpus, it can be considered as a reliable reference of general use of words in Spanish. CREA contains no less than 150 million words and around 700,000 different forms. However, this list presents around 50% of *noisy* words for the experiment: foreign words, orthographic and typographic mistakes as well as proper nouns. A task for cleaning up the list reduced the total number to 350,000 distinct forms. A lot of medical terms of general use (as opposed to technical or professional use) appear on this list, and, additionally, proper nouns such as Down or Alzheimer, that are part of compound terms, were removted. However, when reviewing the number of proper nouns that are not relevant, we chose to eliminate all of them. After this process, only a total of 23,239 candidate terms were included in the list, which are words that are not in the reviewed list in CREA. To provide additional context to the relative size that has been handled, a lexicon like GRAMPAL with 50,000 lemmas generates around 150,000 different forms more than those in a corpus like CREA with more than 150 million words.

The third method uses a purely statistical technique: the Log-Likelihood (LLH) is applied to identify distinct words in the medical corpus [38]. This test is always used in programs checking agreement (such as, Wordsmith or AntConc) to extract keywords in a text. The process performs a comparison of the occurrence frequency between the words in a given corpus with those in a reference corpus. In this case, MultiMedica was compared with the CREA version already pre-processed (see above). To achieve 99.9% of confidence rate, we applied a threshold of significance in 10.83. As a result, the list of candidate terms contains only words with a test value above 10, which renders a list of just 8,667 candidate terms.

Several natural language processing (NLP) techniques were utilized. First, each collection was processed and tags for part-of-speech were included. The Spanish subcorpus was tagged by using GRAMPAL [39], already mentioned. The tagging process is semisupervised, as it requires manual revision to ensure annotation quality. A random sample representing 5% of the popularized texts in Spanish was revised twice to compute the inter-annotator agreement (IAA) value. This was assessed by computing the F-measure, as exposed in Hripcsak and Rothschild (2005) [40], and it was found that both annotators agreed in about 98 per cent of the texts.

Herrero et al. (2014) [41] explain the methodology followed in the creation of the morphological tagging for the Japanese corpus. After considering three different taggers (ChaSen, Mecab and Juman), Juman was chosen, because it provides good segmentation and a wider range of morphological information. Similarly, the Arabic corpus was automatically annotated using the PoS tagger MADA+Tokan [42]. Finally, the tagged texts were indexed for all languages to enhance online queries.

2) Filtering with a List of Affixes and Lemmas

The next step was to create lists of medical terms for each language. The Spanish list was compiled semi-automatically, combining rule-based, tagger-based and statistical approaches [43], as already described in the section above. A gold standard list included terms that appeared in leading medical dictionaries (e.g., RANM 2011, Dorland 2005). A silver-standard list gathered terms that were found only in biomedical books and journals.

Regarding Japanese, a single list was compiled with terms from several medical dictionaries: *Online Life Science Dictionary* [44] and *Japanese-English-Chinese Dictionary* (1994). As for Arabic, the final list is a combination of full terms translated from English resources (SNOMED and UMLS) and a list of Arabic words equivalent to Spanish prefixes and suffixes, such as *-itis*, *cardio-*, etc. [45].

An initial review of the candidate terms shows that some kind of filter must be applied to the list since it contains words not included in the lexicon of the morphological analyzer nor in the CREA list, but that are words of common usage (i.e. *tabúes* or *vinculador*). To further enhance the precision of the selected terms a program was applied for identifying affixes and lemmas of medical terms. The program contains 2,128 items, including orthographic variations such as *aden-* or *adeno-*:

- Greek and Latin affixes in the medical knowledge area (i.e. *cardio-*, *-itis*) and frequent medical lemmas (i.e. *pancrea-*), collected from several sources of medical terms [46]. To avoid false positives, highly frequent affixes were removed from the list, because they are not restricted to the biomedical domain (such as *pre-* or *-able*).
- Lemmas and affixes for identifying pharmacological compounds (*-cavir*) and biochemical substances (*but-* or *-sterol*). All of them have been compiled from lists proposed and approved by the *World Health Organization* (WHO) [47], as well as lists approved by the *American Medical Association* (AMA) [48] for clinical compounds official denominations. As most of scientific English affixes have a unique correspondence with equivalent Spanish affixes, the adaptation was direct with a minimal effort, especially for those ending in vowels such as *-ine* > *-ina* (*creatine* > *creatina*).

In order to obtain the final list, all possible variations of each affix and lemma have been generated. On one side, graphic variations due to diacritics (i.e. tilde), such as *próst-* (as in *próstata*) and *prost-* (as in *prostático*). On the other hand, variations due to an epenthetic vowel: *escoli- scoli-*. And finally, variations due to gender and number inflection, such as the suffix *-génico* can have four different forms: *-génico*, *-génica*, *-génicos* and *-génicas*.

The program that compares affixes with the candidate terms first compares each candidate with all affixes appearing in two different lists (prefixes and suffixes). When a candidate term contains a biomedical affix or lemma, it is considered a potential term. Fig. 1 above displays the whole process.

3) Manual Verification of each Proposed Term

The last phase performs a manual review of all the candidate terms, by confirming or rejecting each term. The final result can be called a *gold standard* or set of reference terms with all validated forms. For a term to be validated, it must appear in a well-known and accepted medical source. In order to avoid subjectivity, the decision is based on consulting the following reference works, and in this order:

- Diccionario de Términos Médicos [49]: with almost 52,000 terms
- Diccionario Médico Enciclopédico Dorland [50]: more than 112,000 terms
- Diccionario Espasa Medicina [51]: 18,000 terms (collected by medical professionals in the Universidad de Navarra)
- Dicciomed [52]: around 7,000 terms (with a historic and etymological approach).

Similarly, terms found regularly in journals and books of biomedical research have been validated and included in the list. Table II is a summary of the classification criteria followed in order to accept or reject a term.

TABLE II. Four Types of Terms

	Term classification	Examples
Accept	List 1 – terms with an entry in a medical reference dictionary List 2 – terms without an entry in a medical reference dictionary, but found in books and scientific articles	*páncreas, ADN …* *RAS, cisteínico …*
Reject	List 3 – terms rejected by specialists, due to orthographic or typographic errors or poor adaptation into Spanish List 4 – non-biomedical terms	**perirenal, *croup…* *Aragón, Pfizer …*

Biomedicine is an extremely wide area for research, and establishing clear-cut boundaries to the domain is almost impossible. The terms of the golden standard come in such fields as Anatomy (*hígado* > *liver*, *nefrona* > *nephron*), Microbiology (*cilio*, "*Escherichia*"), Genetics (*transcripción*, *ARN*), Oncology (*oncogén*, *leukemia*), Biochemistry (*fosforilación*, *amina*), Pharmacology (*aspirina*, *prozac*), History of Medicine (*frenología*, *miasma*), or Surgery and other medical techniques or procedures (*tomografía*, *maniobra*), among others.

Terms from other knowledge areas not strictly related to biomedicine, but common in medical texts were also accepted. For instance, concepts referring to statistical metrics (*variable*, *significance*), agents involved in a disease, like poisonous animals or environmental conditions (*anopheles*, *vipéridos*, *contaminación*) or plants producing pharmacological substances (*Vinca*, *cornezuelo*). In total, the list contains 24,639 terms.

4) Developing a Term Extractor for Each Language

Each language required a different approach in order to build the term extractor. The Spanish extractor uses lists of terms, medical roots and affixes, the GRAMPAL tagger, and rules for multi-words and context patterns. The processing of the input text to detect candidate terms is as follows. First, a dictionary-based method that relies on pattern matching is applied. Each item found in the gold standard list is marked as a highly reliable candidate term (e.g., *pulmón*, 'lung'). Likewise, each term found in the silver standard list is selected as a medium reliable candidate term (e.g., *secundario*, 'secondary'). In the third stage, those words that were not found in any list are POS-tagged through the GRAMPAL tagger. Unrecognized items (i.e., words not included in the lexicon of the tagger, which was designed for the general language) are then filtered using a list of biomedical roots and affixes (e.g.,*hemat(o)-*, an affix related to blood). In this way, for example, an adverb such as *hematológicamente* ('hematologically') may be recognized as a term and highlighted with medium reliability. The last stage involves applying multi-word formation rules to the previous list of candidate terms. If any element of the multi-word candidate term has medium reliability, the whole unit is highlighted as such. For example, if the term *complejo* ('complex,' medium reliability) and *amigdalino* ('tonsillar,' high reliability) are recognized, a multi-word rule will join both terms in *complejo amigdalino* ('tonsillar complex') and mark it as a medium reliability candidate term. Fig. 2 outlines the architecture of the system.

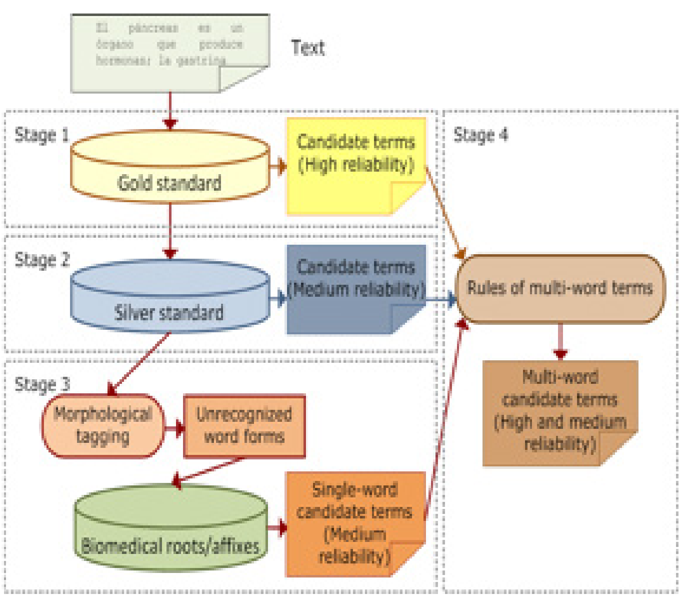


Fig. 2. Phased architecture of the Spanish term extractor [53].

The extractors for Japanese and Arabic follow a simpler procedure. The Japanese extractor performs an initial pattern matching throughout the dictionary, identifying those terms as highly reliable. Secondly, a series of rules are applied bearing in mind the agglutinative nature of the language. For example, if two dictionary terms are joined with a connective particle, it will be considered as a single multi-word term; also, if additional *kanji* characters are added to the initial or final part of a dictionary term, the extractor recognizes the whole string of characters as a single term. The terms detected using this rule-based procedure are classified as medium reliable ones. The Arabic language is mainly a dictionary-based extractor that recovers terms from the medical list created for this purpose.

Improvement in the term extraction in the future includes adding more medical terms, or codes from the *International Classification of Diseases version 10* (ICD-10) [54], the *Unified Medical Language System* (UMLS) and the *Systematized Nomenclature of Medicine – Clinical Terms* (SNOMED-CT) [55].

5) Interaction with the MultiMedica Corpus

Users can perform queries in the corpus in two ways: *simple word search* ("Search" tab, "Consulta" in the Spanish version) and *medical term search* ("Medical Term Search" tab, "Consulta de Términos Médicos" in Spanish). In addition, users can input a free text to detect and extract candidate terms in the domain ("Medical Term Extractor," "Extractor de Términos Médicos").

a) Word Search

Any word in the corpus can be searched according to form, lemma or part-of-speech (POS). For example, if the user inputs the lemma *cáncer*, the results may be *cáncer* or *cánceres* (respectively, 'cancer' or 'cancers'). The user has the option of looking up the collocations of the word as well as its frequency and log-likelihood value.

In the search results, frequency values are normalized per million words (hereafter, *pmw*). Counts are also compared to the frequencies in the *Corpus de la Real Academia Española* (CREA) corpus. This makes it possible to know the *distinctiveness* of the searched word in a specialized corpus and in relation to a general language corpus. For example, when the word *hepatitis* is searched, the normalized frequency in the MultiMedica corpus is 385.8 pmw, and 6.1 pmw in the CREA corpus. This shows that this token is highly related to this specialized genre. In contrast, if *corazón* ('heart') is searched, the normalized frequency in the MultiMedica corpus drops to 140.8 pmw, which is close to the normal frequency in the CREA corpus (125.3 pmw). This indicates that *corazón* appears with a similar frequency in a health and a general corpus. Since this is a polysemous word, other senses beyond the anatomical context are used in the general language (e.g., related to feelings, or as a synonym of 'nucleus' or 'core').

The word search for Spanish, Arabic, and Japanese are shown in Fig. 3, 4 and 5, respectively.

Fig. 3. Search medical terms in Spanish [53].

Fig. 4. Search medical terms in Arabic [53].

Fig. 5. Search medical terms in Japanese [53].

The search tool for the Spanish corpus also provides information about word distribution (i.e., its frequency in each type of text). This feature makes it possible to compare different text genres (popular vs. technical documents). If we search for *dolor de espalda* ('upper back pain'), the results show that this term is more frequent in popularized texts than in technical texts. However, when we search for *dorsalgia* (the technical synonym of 'dolor de espalda'), the results reveal that this term is restricted to academic documents.

b) Medical Term Search

The medical term search allows users to look up the most frequent medical terms in the corpus. An autocomplete function provides a list of all the possible terms that contain the typed letters introduced by the user. The list is based on the 5,000 more frequent terms in the corpus.

c) Medical Term Extractor

The medical term extractor detects candidate terms from an input text (Fig. 6 and 7). The tool highlights medical terms according to their level of reliability: high (terms included in the gold standard list) and medium (terms in the silver list). The user may also download the term list in text format for further use. In addition, terms that are found in the *BabelNet* dictionary [56] contain a hyperlink to this resource, which provides their translation in many languages.

IV. Future Work

Biomedical Natural Language Processing (BioNLP) is receiving a growing interest from both academia and industrial specialized applications. The specific field of biomedical text mining is one of the most mature domains. Biomedical text mining, of which term extraction is just one area, is providing great advances in terms of widespread availability of expert-annotated text resources, biomedical term banks, and a great number of information extraction components.

Fig. 6. The medical term extractor for Spanish texts [53].

Fig. 7. A screenshot of the Japanese term extractor [53].

Biomedical text processing components have been published, covering various aspects, from tokenization approaches [57] to the creation of specialized tokenizers for biomedical texts [58]. Equally important are special linguistic and NLP tools for biomedical texts, such as POS taggers [59] or dependency-based parsers [60] for pure syntactic analysis (Enju/Mogura [61], GDep), which present biomedical domain models to create graphic representations of syntactic dependency relations. These syntactic relations are used to express bioentity relationships present in the text (such as protein-protein interactions [62]) in combination with recent machine learning techniques.

Current and future promising trends biomedical natural language processing include the following: to rank a classification of topics of relevance in a text after term identification [63]; detection of different types of bioterms applying semantic roles; indexing of documents to terms and concepts from controlled vocabularies and corpora, as in the case of Multimedica, which may build bioontologies [64] to be applied in other domains, and extracting relationships between biomedical terms (protein or gene relations [65]). Another area of biomedical term extraction research field is the detection of associations between disease concepts and actual disease areas [66], like in the bioontologies mentioned above.

As already covered in the present paper, the first step or phase in most biomedical term identification is to locate mentions of biological entities of interest or terms, in the sense used here. Work in biomedical natural language processing is very much dependent on research in the biomedical sciences, which have recently focused on the study of a set of concepts, like genes, proteins, chemicals, drugs or certain diseases. Tools, like the term extractor and search engine presented here, can be a great help for a more efficient way of finding information in documents, that build up the corpora, and then characterize those concepts so researchers can reach deeper insights into their own domains.

One example of the importance given to this topic are initiatives like BioASQ [67]. This is a European Commission-funded project under the FP7 programme, whose goal is to organize challenges on biomedical semantic indexing and question answering (QA). The challenges include tasks relevant to hierarchical text classification, machine learning, information retrieval, QA from texts and structured data, multi-document summarization and many other areas.

In the last couple of years, the work in biomedical NLP was dominated by applications of deep learning to: punctuation restoration [68], text classification [69], relation extraction [70] [71] [72] [73], information retrieval [74], and similarity judgments [75], among other exciting progress in biomedical language processing. For a more detailed exploration of recent topics, the BioNLP Annual Workshop [76] covers the most researched and debatable areas.

Term extraction has other applications beyond BioNLP, as is the case with chemical terminology, legal texts, the engineering documentation for the oil & gas industries, or research of new drugs in the pharma industries, just to name but a few.

V. Conclusion

This paper has covered a use case of term extraction in the BioNLP domain, starting form a description of the basic techniques used to the methodology followed in the creation of a multilingual corpus of medical texts for medical term extraction, their morphological annotation and further indexation, the actual term list extraction and the development of an online tool so a user can reach the information and use it for consultation or clarification of the medical term. Three languages were selected: Spanish, Arabic and Japanese, languages so different genetically and typologically, that specific approaches and tools had to be chosen for each of them. This led to identifying several problems for the computational treatment of medical terms in these languages, for example, the lack of language resources in medical NLP for Arabic (either professional texts or electronic dictionaries). In this sense, MultiMedica is a pioneering effort in this Biomedicine domain and for this combination of languages. It has also provided an interesting typological insight into how languages behave within the medical domain. Each of the three languages presented different challenges when developing the extractor: the variation in inflection of Spanish terms, variation in the Arabic writing system or word segmentation in Japanese due to the lack of white spaces between words. Even though the initial steps of creating the corpus, tagging, and development of a medical term list was approximately equal in the three languages, the processing of the texts and the creation of the extractor had to be adapted to the specificities of each language.

Looking into the future it is reasonable to expect that the corpus and online tools may provide the users with a good amount of data for future linguistic research into biomedical discourse and may be used for many other use cases. The term extractor may fulfil terminologists' and translators' needs by helping them identify term candidates and finding their equivalents in other languages. In addition, health professionals, in the broad sense, including clinical, pharma or chemical professionals, and medical students could make use of this interface to seek and translate biomedical information online.

References

[1] Z. Zhang, J. Iria, C. Brewster, and F. Ciravegna. "A comparative evaluation of term recognition algorithms", in *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08*), Marrakech, Morocco. European Language Resources Association (ELRA). http://www.lrecconf.org/proceedings/lrec2008/. 2008

[2] K. Kageura and B. Umino, "Methods of automatic term recognition: A review", Terminology, 3(2) (Ámsterdam, 1996), págs. 259-289; and M. Krauthammer y G. Nenadic, "Term identification in the biomedical literature", *Journal of Biomedical Informatics, 37*, pp. 512-526, Ámsterdam, 2004

[3] S. Ananiadou and J. Mcnaught (eds.), *Text Mining for Biology and Biomedicine*, Artech House, Boston, MA, 2006.

[4] K. B. Cohen, "Biomedical Text Mining", in N. Indurkhya and F. J. Damerau (eds.), *Handbook of natural language processing, 2ª ed*., Chapman and Hall, Boca Raton, pp. 605-625, 2010.

[5] A. Moreno Sandoval and T. Redondo. "Text Analytics: the convergence of Big Data and Artificial Intelligence". *International Journal of Artificial Intelligence and Interactive Multimedia*, Vol. 3-6. 2016.

[6] J. Vivaldi, *Extracción de candidatos a término mediante la combinación de estrategias heterogéneas*, PhD Thesis, Universidad Politécnica de Cataluña, 2001.

[7] M. T Cabré, *Terminology: Theory, methods and applications*, John Benjamins, Ámsterdam, 1999.

[8] K. Kageura and B. Umino, op. cit.

[9] S. Ananiadou and G. Nenadic, "Automatic terminology management in biomedicine", in S. Ananiadou and J. McNaught (eds.), op. cit, 2006.

[10] O. Bodenreider, "The Unified Medical Language System (UMLS): integrating biomedical terminology", *Nucleic Acids Research, 3*2 (Database issue), Oxford, 2004.

[11] MeSH (Medical Subject Headings) is the National Library of Medicine controlled vocabulary thesaurus used for indexing articles for PubMed (https://www.ncbi.nlm.nih.gov/mesh).

[12] SNOMED CT (Systematized Nomenclature of Medicine – Clinical Terms) is the most comprehensive and precise clinical health terminology product in the world (https://www.snomed.org/snomed-ct/).

[13] ICD10Data.com is a free reference website designed for the fast lookup of all current American ICD-10-CM (diagnosis) and ICD-10-PCS (procedure) medical billing codes (http://www.icd10data.com/).

[14] A. Ballester, Á. Martín Municio, F. Pardos, J. Porta, R. J. Ruiz and F. Sánchez, "Combining statistics on n-grams for automatic term recognition", in *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*. Universidad de Las Palmas de Gran Canaria, 2002.

[15] K. Kageura and B. Umino, op. cit.

[16] M. Krauthammer and G. Nenadic, 2004, op. cit.; in S. Ananiadou and G. Nenadic, op. cit.

[17] I. Segura-Bedmar, P. Martínez Fernández and D. Samy, "Detección de fármacos genéricos en textos biomédicos", *Procesamiento del Lenguaje Natural, 40*, Jaén, 2008.

[18] S. Ananiadou, "A methodology for Automatic Term Recognition", COLING'94 – *Proceedings of the 15th Int. Conf. on Computational Linguistics*, pp. 1034-1038, 1994.

[19] I. Dagan and K. Church, "TERMIGHT: Identifying and Translating Technical Terminology", in *4th Conference on Applied Natural Language Processing*, 1994; and J. S. Justeson and S. M. Katz, "Technical terminology: some linguistic properties and an algorithm for identification in text", *Natural Language Engineering, 1(1)* Cambridge, 1995.

[20] W. Koza, Z. Solana, M. DA S. Conrado, S. O. Rezende, T. A. Pardo, J. Díaz-Labrador and J. Abaitua, "Extracción terminológica en el dominio médico a partir del reconocimiento de sintagmas nominales", *INFOSUR, 5*, Rosario, Argentina, 2011.

[21] For a comparison between Greek-Latin suffixes in English and Japanese, two languages belonging to two very distinct language families, please review C. Herrero Zorita, C. Molina and A. Moreno Sandoval, "Medical term formation in English and Japanese: A study of the suffixes -gram, -graph and -graphy", *Review of Cognitive Linguistics, 13(1),* Amsterdam, 2015.

[22] A. Moreno Sandoval and J. M. Guirao, "Frecuencia y distintividad en el uso lingüístico: casos tomados de la lematización verbal de corpus de distintos registros", in *Actas del I Congreso Intl. de Lingüística de Corpus*, Universidad de Murcia, Murcia, 2009.

[23] T. Dunning, "Accurate methods for the statistics of surprise and coincidence", *Computational Linguistics, 19(1),* Cambridge, MA, 1993.

[24] A. Kilgarriff, P. Rychly, P. Smrz and D. Tugwell, "The Sketch Engine", in *Proceedings of EURALEX 2004*, Lorient, France, 2004.

[25] H. Nakagawa and T. Mori, "Automatic term recognition based on statistics of compound nouns and their components", *Terminology, 9(2*), Amsterdam, 2003.

[26] R. Nazar, J. Vivaldi and L. Wanner, "Automatic taxonomy extraction for specialized domains using distributional semantics", *Terminology, 18(1)*, Amsterdam, 2012.

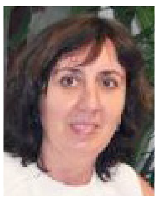
[27] A. Ballester Á. Martín Municio, F. Pardos, J. Porta, R. J. Ruiz and

F. Sánchez, "Combining statistics on n-grams for automatic term recognition", in *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*. Universidad de Las Palmas de Gran Canaria, 2002.
[28] R. Nazar and M. T. Cabré, "Un experimento de extracción de terminología utilizando algoritmos estadísticos supervisados", *Debate Terminológico, 7*, 2010.
[29] S. Hochreiter and J. Schmidhuber. "Long short-term memory". *Neural computation*, vol. 9, no 8, pp. 1735-1780, 1997.
[30] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer. "Neural architectures for named entity recognition", in *Proceedings of NAACL 2016*, San Diego, CA, 2016.
[31] M. Rei. "Semi-supervised Multitask Learning for Sequence Labeling", *Proceedings of ACL 2017*, Vancouver, Canada, 2017.
[32] T. Mikolov, K. Chen, G. Corrado, and J. Dean. "Efficient estimation of word representations in vector space". 2013. https://arxiv.org/abs/1301.3781.
[33] J. Pennington, R. Socher, and C. D. Manning. "Glove: Global vectors for word representation". in *EMNLP*. volume 14, pages 1532– 1543, 2014.
[34] A. Barrón Cedeño, G. Sierra, P. Drouin and S. Ananiadou, "An Improved Automatic Term Recognition Method for Spanish", in A. Gelbukh (ed.), *CICLing2009 LNCS 5449*. Springer, Berlín, 2009.
[35] A. Moreno Sandoval and L. Campillos Llanos, "Design and annotation of MultiMedica - a multilingual text corpus of the biomedical domain", in C. Vargas-Sierra (ed.), *Procedia, 95*, Elsevier, Berlín, 2013.
[36] A. Moreno Sandoval and L. Campillos Llanos, "Combined strategies for automatic term recognition and its application to a Spanish corpus of medicine". *Lingüística española actual*, 37(2), pp. 173-197. 2015.
[37] A. Moreno Sandoval and J. M. Guirao Miras, "Morpho-syntactic Tagging of the Spanish C-ORAL-ROM Corpus: Methodology, Tools and Evaluation", in Y. Kawaguchi, S. Zaima and T. Takagaki (eds.), *Spoken Language Corpus and Linguistic Informatics*, John Benjamins, Amsterdam, 2006.
[38] T. Dunning, "Accurate methods for the statistics of surprise and coincidence", *Computational Linguistics, 19(1),* Cambridge, MA, 1993.
[39] A. Moreno Sandoval and J. M. Guirao Miras, op. cit.
[40] G. Hripcsak, and A. S. Rothschild. "Agreement, the F-measure, and reliability in information retrieval." *Journal of the American Medical Association 12*: 296-298, 2005.
[41] C. Herrero Zorita, L. Campillos Llanos and A. Moreno Sandoval. "Collecting a POS-Tagging a Lexical Resource of Japanese Biomedical Terms from a Corpus." *Procesamiento del Lenguaje Natural 52*: 29-36. 2014.
[42] N. Habash, O. Rambow and R. Roth. "Mada+Tokan: A Toolkit for Arabic Tokenization, Diacritization, Morphological Disambiguation, POS Tagging, Stemming and Lemmatization." *Proceedings of the 2nd International Conference on Arabic Language Resources and Tools (MEDAR)*. Cairo, Egypt: 242-245. 2009.
[43] A. Moreno Sandoval and L. Campillos Llanos, op. cit.
[44] Online Life Science Dictionary. Available at https://lsd-project.jp/cgi-bin/lsdproj/ejlookup04.pl [30/12/2017]
[45] D. Samy, A. Moreno Sandoval, C. Bueno-Díaz, M. Garrote-Salazar and J.M. Guirao. "Medical Term Extraction in an Arabic Medical Corpus." *Proceedings of the 8th Language Resources and Evaluation Conference,* pp. 640-645. Istanbul: LREC, 2012
[46] J. M. López Piñero and M. L. Terrada Ferrandis, *Introducción a la terminología médica*, Masson, Barcelona, 2005; and M. E. Jiménez, "Afijos grecolatinos y de otra procedencia en términos médicos", *MEDISAN*, 16(6) (Santiago de Cuba, 2012), pp. 1005-1021; M. A. Sánchez González, *Historia de la medicina y humanidades médicas*, 2 ed., Elsevier/Masson, Barcelona, 2012
[47] WHO, "The use of stems in the selection of International Nonproprietary Names (INN) for pharmaceutical substances", 2013, http://www.who.int/medicines/services/inn/StemBook_2013_Final.pdf and http://www.who.int/medicines/services/inn/Addendum_StemBook2013_201506.pdf [01/12/2017]; OMS, "International Nonproprietary Names (INN) for biological and biotechnological substances (a review)", 2016, < http://www.who.int/medicines/services/inn/BioReview2016.pdf> [01/12/2017].
[48] AMA: < https://www.ama-assn.org/sites/default/files/media-browser/public/usan/stem-list-cumulative_0.xlsx > [01/12/2017]. Michael Quinion's list of affixes was also used: < www.affixes.org > [01/12/2017]
[49] Real Academia Nacional de Medicina, *Diccionario de términos médicos*, Editorial Médica Panamericana, Madrid, 2011.
[50] Dorland, *Diccionario enciclopédico ilustrado de medicina* Dorland, Elsevier, Madrid, 30 edition, 2005; online: < https://www.dorlandsonline.com/dorland/home> [01/12/2017]
[51] L. M. Gonzalo Sanz (coord.), *Diccionario Espasa Medicina*, Espasa S.L., Madrid, 1999.
[52] F. Cortés Gabaudán (coord.), *Dicciomed*, 2007-2013. http://dicciomed.eusal.es [01/12/2017]
[53] A. Moreno SandovaL, L. Campillos Llanos, C. Herrero-Zorita, J. M. Guirao Miras, A. González Martínez, D. Samy y E. Takamori "An online tool for enhancing NLP of a biomedical corpus", *6th International Conference on Corpus Linguistics (CILC 2014)*, Las Palmas de Gran Canaria, 2014
[54] A Spanish version of the ICD-10 is accessible through the web of the Spanish Ministry of Health (http://eciemaps.msssi.gob.es/ecieMaps/browser/index_10_mc_old.html) [6/01/2018]
[55] Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT): https://www.snomed.org/snomed-ct/ [6/01/2018]
[56] BabelNet: http://babelnet.org/ [6/01/2018]
[57] Y. He, and M. Kayaalp. "A Comparison of 13 Tokenizers on MEDLINE"; *Technical Report LHNCBC-TR-2006-003*; The Lister Hill National Center for Biomedical Communications: Bethesda, MD, December 2006
[58] N. Barrett, and J. Weber-Jahnke, "Building a Biomedical Tokenizer Using the Token Lattice Design Pattern and the Adapted Viterbi Algorithm". *BMC Bioinf. 2011*, 12 (Suppl 3), S1, 2011
[59] Y. Tsuruoka,, Y. Tateishi, J. D. Kim, T. Ohta, J. McNaught, S. Ananiadou, and J. Tsujii, "Developing a Robust Part-of-Speech Tagger for Biomedical Text" in *Advances in Informatics*; Bozanis, P., Houstis, E. N., Eds.; Vol. 3746 Springer Berlin Heidelberg: Berlin, Heidelberg, 2005.
[60] K. Sagae, and J. Tsujii, "Dependency Parsing and Domain Adaptation with LR Models and Parser Ensembles". *Proceedings of the CoNLL Shared Task Session of EMNLP-CoNLL*; Prague, Czech Republic, June, 2007.
[61] Y. Miyao and J. Tsujii. "Feature Forest Models for Probabilistic HPSG Parsing". *Computational Linguistics*. 34 (1), 2008.
[62] M. Miwa, R. Sætre, Y. Miyaq, and J. Tsujii, "Protein-Protein Interaction Extraction by Leveraging Multiple Kernels and Parsers". *International Journal of Medical Informatics*, 78 (12), 2009.
[63] J-F. Fontaine, A. Barbosa-Silva, M. Schaefer, M.R. Huska, E. M. Muro, M.A. Andrade-Navarro, "MedlineRanker: Flexible Ranking of Biomedical Literature". *Nucleic Acids Res. 37*, 2009.
[64] I. Spasic, S. Ananiadou, J. Mcnaught, and A. Kumar, "Text Mining and Ontologies in Biomedicine: Making Sense of Raw Text", *Briefings in Bioinformatics.* 6 (3), 2005.
[65] F. Leitner, S.A. Mardis, M. Krallinger, G. Cesareni, L.A. Hirschman, A. Valencia, "An Overview of BioCreative II.5". *IEEE/ ACM Transactions in Computational Biological Bioinformatics*. 7 (3), 2010.
[66] M. vazquez, M. krallinger, F. leitner, and A. Valencia, "Text Mining for Drugs and Chemical Compounds: Methods, Tools and Applications". *Molecular Informatics*. 30 (6−7), 2011.
[67] Project BioASQ: http://bioasq.org/ [6/01/2018]
[68] W. Salloum, G. Finley, E. Edwards, M. Miller and D. Suendermann-Oeft *Deep Learning for Punctuation Restoration in Medical Reports*, 2017
[69] S. Baker and A. Korhonen, *Initializing neural networks for hierarchical multi-label text classification*, 2017
[70] C. Lin, T. Miller, D. Dligach, S. Bethard and G. Savova, *Representations of Time Expressions for Temporal Relation Extraction with Convolutional Neural Networks*, 2017
[71] M. Asada, M. Miwa and Y. Sasaki, *Extracting Drug-Drug Interactions with Attention CNNs*, 2017
[72] Y. Peng and Z. Lu, *Deep learning for extracting protein-protein interactions from biomedical literature*, 2017
[73] J. Tourille, O. Ferret, A. Neveol and X. Tannier. "Neural Architecture for Temporal Relation Extraction: A Bi-LSTM Approach for Detecting Narrative Containers.". *Proceedings of ACL 2017*, pp. 224-230. 2017.
[74] S. Mohan, N. Fiorini, S. Kim and Z. Lu, *Deep Learning for Biomedical Information Retrieval: Learning Textual Relevance from Click Logs*, 2017.
[75] B. McInnes and T. Pedersen, *Improving Correlation with Human Judgments by Integrating Semantic Similarity with Second–Order Vectors*, 2017
[76] K. Bretonnel Cohen, D. Demner-Fushman, S. Ananiadou, and J.-I.

Tsujii. *Biomedical natural language processing in 2017: The view from computational linguistics* (http://aclweb.org/anthology/W17-23) [6/01/2018]

Antonio Moreno-Sandoval

Antonio Moreno-Sandoval (BArts 1986, MArts 1988, PhD 1991, Universidad Autónoma de Madrid, UAM) is Professor of Linguistics and Director of the Computational Linguistics Lab at UAM. He is a former Fulbright postdoc scholar at the Computer Science Dept., New York University (1991-1992) and a former DAAD scholar at Augsburg Universität (1998). His training in Computational Linguistics began as a research assistant in the Eurotra Machine Translation Project (EU FP-2) and then at IBM Scientific Center in Madrid (1989-1990). He was the principal researcher of the Spanish team in the C-ORAL-ROM Project (EU FP-5). He has managed over 15 projects (national, regional-funded) as well as industry contracts. Since 2010 he is Senior Researcher at the Instituto de Ingeniería del Conocimiento (IIC-UAM) in the Social Business Analytics group. Moreno-Sandoval has supervised 9 theses to completion. He is author or co-author of 4 books and over 80 scientific papers.



Julia Díaz

Julia Díaz received a MSci Degree in Mathematics, PhD in Computer Science (both from Universidad Autónoma de Madrid – UAM-Spain) and General Management Program (IESE-Universidad de Navarra – Spain). At present she is Senior Innovation Manager at the Instituto de Ingeniería del Conocimiento (IIC-UAM), a private R&D&i institution dedicated to extracting knowledge from high volumes of heterogeneous data (Big Data) and optimizing business processes in areas such as Healthcare and Energy. She also is Part Time PhD Professor in Computer Sciences in the UAM and Professor in the Big Data & Data Sciences Master in UAM and ESADE.

Leonardo Campillos-Llanos

Leonardo Campillos-Llanos (BArts 2004, Universidad Complutense de Madrid-UCM; MArts 2006, PhD 2012, Universidad Autónoma de Madrid-UAM) is a postdoctoral researcher at the Laboratoire d'Informatique pour la Mécanique et les Sciencies de l'Ingénieur (LIMSI-CNRS) at Orsay (France). He has been working in the area of BioNLP since 2011 in different projects (MULTIMEDICA, PatientGenesys). Currently, Dr. Campillos is in charge of the natural language interaction and terminology modules of a dialogue system simulating a virtual patient in a medical consultation (PVDial project).

Teófilo Redondo

Teófilo Redondo (BArts -1985, MArts - 1986; Universidad Complutense de Madrid - UCM) is Senior Consultant – ICT at Ayming España, with a special focus on Artificial Intelligence, Cognitive Robotics and all things NLP. Prior to this he was Project Portfolio Coordinator at Zed Worldwide, in the Deparment of Innovation. He was before Technology Architect & Director of Innovation Projects at Universidad Internacional de La Rioja (UNIR). Previously he developed a career at IBM covering several areas like Cloud Computing and Big Data Architectures, Enterprise Solutions Architect (SAP, Oracle Solutions, Dassault Systèmes), and as SOA Architect. He started in the IBM Research Division (IBM Scientific Center in Madrid) with several projects on Machine Translation, during which he produced a number of articles on this subject. He was Visiting Scholar at Stanford University (1987). He is affiliated with SEPLN (Sociedad Española para el Procesamiento del Lenguaje Natural) since almost the beginning.

Evaluation of a Diagnostic Decision Support System for the Triage of Patients in a Hospital Emergency Department

JC Nazario Arancibia[1]*, FJ Martín Sanchez[2], AL Del rey Mejías[1], J Gonzalez del Castillo[2], J Chafer Vila-plana[1], MA García Briñon[2], MM Suárez-Cadenas[1], J Mayol Martínez[1], G Seara Aguilar[1]

[1] Innovation Unit, Hospital Clínico San Carlos, IdISSC, Madrid (Spain)
[2] Emergency Department, Hospital Clínico San Carlos, IdISSC, Madrid (Spain)

Received 13 December 2017 | Accepted 27 March 2018 | Published 20 April 2018

Abstract

One of the biggest challenges for the management of the emergency department (ED) is to expedite the management of patients since their arrival for those with low priority pathologies selected by the classification systems, generating unnecessary saturation of the ED. Diagnostic decision support systems (DDSS) can be a powerful tool to guide diagnosis, facilitate correct classification and improve patient safety. Patients who attended the ED of a tertiary hospital with the preconditions of Manchester Triage system level of low priority (levels 3, 4 and 5), and with one of the five most frequent causes for consultation: dyspnea, chest pain, gastrointestinal bleeding, general discomfort and abdominal pain, were interviewed by an independent researcher with a DDSS, the Mediktor system. After the interview, we compare the Manchester triage and the final diagnoses made by the ED with the triage and diagnostic possibilities ordered by probability obtained by the Mediktor system, respectively. In a final sample of 214 patients, the urgency assignment made by both systems does not match exactly, which could indicate a different classification model, but there were no statistically significant differences between the assigned levels (S = 0.059, p = 0.442). The diagnostic accuracy between the final diagnosis and any of the first 10 Mediktor diagnoses was of 76.5%, for the first five diagnoses was 65.4%, for the first three diagnoses was 58%, and the exact match with the first diagnosis was 37.9%. The classification of Mediktor in this segment of patients shows that a higher level of severity corresponds to a greater number of hospital admissions, hospital readmissions and emergency screenings at 30 days, although without statistical significance. It is expected that this type of applications may be useful as a complement to the triage, to accelerate the diagnostic approach, to improve the request for appropriate complementary tests in a protocolized action model and to reduce waiting times in the ED.

Keywords

Computer Aided Diagnostic Decision Support System, Emergency Triage, Emergency Service.

DOI: 10.9781/ijimai.2018.04.006

I. Introduction

Triage is the first evaluation and classification process used to prioritize the serious patients that arrive at the Hospital Emergency Departments (HED), its usefulness is accentuated by the considerable demand for care, with frequent congestions of services, in a context of limited resources.

In Spain, with data referring to all public and private hospitals, visits to the HED grew from 17.9 million in 1997 (449.5 visits per 1,000 inhabitants/year) to 24.4 million in 2005 (553, 1 visits per 1,000 inhabitants/year), with an average annual increase of 2.6% during that period [1]. Most of this increase is attributed to a disproportionate increase in patients who use Emergency Department (ED) inappropriately. This often results in saturation of HED [2]. According to the most recent data published by the Ministry of Health, Social Services and Equality, the number of emergencies attended in 2016 in the hospitals of the National Health System was 21.5 million, of which 11.3% required hospital admission [3]. These voluminous figures justify the development of different triage scales to offer safer, easier and more organized assistance.

* Corresponding author.

E-mail address: jcnazario@shealth.eu

Currently, there are several classification models adapted to the emergency area and validated for adults, such as: the Australian classification scale "Australasian Triage Scale (ATS)", the Canadian classification and severity scale of the emergency department (CTA), the system triage of Manchester "Manchester Triage System (MTS)", the Emergency Severity Index (ESI) and the Andorra Triage Model (MAT). The latter was adopted as a standard reference for the Spanish Triage Model [4].

It is required that classification systems have the ability to define the most appropriate initial location and reasonable attention times, but they are also expected to help predict the clinical evolution and the patient's resource needs, so that they can contribute to a more efficient management of the emergency service [5].

Some studies have documented a good correlation between the level of classification and the hospital admission index, the duration of the stay in emergencies, the requirements in the consumption of diagnostic resources and even with the survival at 6 months [6-9].

Regarding hospital admissions, research has focused mainly on the development of scales with high negative predictive value at non-urgent levels, compromising the positive predictive value for more severe cases and resulting in hospital admission rates that vary widely, with percentages as different as 15% -90% in the urgent levels of the scale. Therefore, the correct classification of high complexity cases continues to be a challenge [5]. But so is the case of less complex cases, responsible for the frequent saturation of emergency services.

Most authors have tried to calculate, with limited results, the probability of hospital admission based on individual variables such as the subjective feeling of the doctor and/or nurse [7-9], the hemodynamic situation of the patient [10] or the levels of severity assigned by the classification systems [11]. One of these models was developed in our hospital after the retrospective analysis of 2476 visits, depending on age, sex, reason for consultation, initial location and the existence of prior treatment or the need to perform complementary tests in the emergency services [12].

On the other hand, misdiagnosis is more likely in emergency departments, and this can involve serious injury or death [14-16]. Almost half of the main diagnostic errors are the result of an incorrect evaluation by the medical staff, either due to erroneous collection of medical history (10%), errors in the physical examination (10%) and decision logic with the patient (30%) [17]. More research is needed to develop tools to prevent misdiagnosis [18].

Along with this approach, there has been a proliferation of more sophisticated programs, known as symptom analysers, that attempt to provide patients with a potential diagnosis and direct them to the appropriate care setting. With the use of computerized algorithms, through a series of questions or the introduction of symptoms, these programs give the user a list of possible diagnoses ordered by probability [19].

The symptom analysers have two functions: pre-diagnosis and classification. The prediagnosis presents the patient with a range of diagnoses that can be adjusted to their symptoms. The classification function tells the patient what kind of attention should be sought and with what level of urgency [20].

Although this validation focuses on the field of external hospital emergency, symptom analysers also provide access to the medical diagnostic approach in low-resource settings, such as in rural or developing countries. Active medical cooperation with symptom analysers can help improve public health [21]. These tools can also help people who can not use telephone triage or who are too insecure to access a care center without an appointment. A study showed, as early as 1976, that people are more honest about their drinking habits compared to a computer than with a doctor, and in a 1994 trial, pregnant women shared more problems with a computer than with an obstetrician [22]. Diagnostic support tools could be useful to prevent misdiagnosis and improve patient safety. General decision support systems have been developed in a technically rigorous way in computer laboratories, but, to a large extent, they are not used because they have paid little attention to the workflow of the place where they will be used [23]. Although there are few studies on the specific success factors of diagnostic decision support systems, a recent systematic review found that the most predictive feature of their success was "the automatic provision of decision support as part of the clinical workflow". In the analysis of our use case, we have used an approach that can offer a greater chance of success in this integration in the workflow in the emergency services, collecting the data directly from the patients before their meeting with the doctor [24-25].

A. Description of the Application Used

Mediktor[1] is an application designed to facilitate and speed up the medical diagnostic process. Through artificial intelligence algorithms that guide the patient's interrogation in a similar way to how a doctor would do it, Mediktor obtains diagnostic forecasts, establishes a level of urgency, and suggests the medical specialty indicated for each patient. The system collects demographic data of the patient (gender, age), geolocation, season of the year, vital signs, previous illnesses and medication if applicable. Using sophisticated natural language interpretation algorithms, it recognizes the reason for consultation as expressed by the user without the need to use specialized language and initiates a series of questions about symptoms and signs, until a list of pre-diagnoses is presented and a level of priority is assigned.

Considering the results of the aforementioned study carried out in the HED of our hospital, both on prediction of hospital admission, and on detection of high risk of poor outcomes [12], we added some variables to the original Mediktor record in order to evaluate its applicability to the patient segment included in this use case. These variables include the patient's need for help in basic activities, and if these needs have increased during the acute process that leads to emergencies, their cognitive and sensory situation (memory and vision), the previous use of medication, recent hospital admissions, the reason for consultation, the location in the ED, and the need for additional tests or to establish treatment in the ED.

B. Objective and Context of the Study

The main objective of this study was to determine, in a group of urgent patients of low complexity, the level of coincidence between the pre-diagnoses generated by Mediktor and the diagnosis at discharge or admission issued by the physician in the ED. The secondary objectives were to evaluate the effect of Mediktor's diagnostic concordance on hospital admissions; to compare the allocation recommended by Mediktor and by Manchester, and to ascertain any relationship with the frequency of admissions and discharges, readmissions, post-emergency visits, and mortality during the 30 days following the evaluation. In addition, to evaluate the relation of the triage with the possible predictive variables of admission and to describe the duration of the different phases of the stay in the emergency room.

The study was developed as a collaboration between the Emergency Department and the Innovation Support Unit of Hospital Clínico San Carlos (HCSC) in Madrid. Teckel Medical made available the Mediktor tool and its maintenance during the study, as well as the results data of its algorithms necessary for the analysis.

II. Methodology

A. Study Design

The validation was designed as a prospective observational cohort study performed in the Emergency Department of the HCSC, a reference hospital with 960 beds of high complexity. The study was approved by the HCSC Research Ethics Committee.[2]

In the ED, the triage of hospital emergencies is carried out by nurses following the Manchester triage method. A series of consecutive patients attending the Emergency Department between 3/15/2017 and 4/10/2017 were selected.

B. Inclusion/Exclusion Criteria

To participate in the study, the subjects had to fulfil all the following requirements: Assignment of levels 3, 4 or 5 in the triage according to the Manchester assessment scale; access to the Emergency Service without documented prior medical evaluations; Age 18 or older; ability to give informed consent and to grant it and reason for consultation reported in the triage evaluation among the 5 most frequent: dyspnoea, chest pain, gastrointestinal haemorrhage, general discomfort or abdominal pain.

1 Mediktor is a registered trademark of Teckel Medical; Mediktor Corp, Barcelona, Spain.

2 Opinion CI 16/508-E, November 16, 2016.

To be excluded from the study, subjects had to meet the following requirements: patients assigned to levels 1 and 2 in the triage, according to the Manchester assessment scale and participation in a pharmaceutical research study during the previous 30 days.

C. Methods for Data Acquisition and Coding

After the acceptance of the informed consent, the Mediktor software generated a randomized code with 6 digits, which was registered manually in a separate file, for its later relation with data of the patient's clinical history. This register was kept under custody by clerical personnel of the Emergency Department, without access for the research team.

The results of the Mediktor triage for each patient were stored in a coded electronic database, without any personal identification data except the random code. Independently, in a different physical location, and without access to Mediktor, the outcome variables of the clinical history established in the protocol were collected, and later matched to the Mediktor data, using the random code mentioned above.

In the analysis stage, subjects without an explicit record of diagnosis at discharge were excluded, as were those in whom the final diagnosis was described as a symptom or sign (Diagnostics not admitted in the ICD-10), those in which by mistake the same code was identified for two patients, and cases with diagnoses not included in the Mediktor database at the time of the study. After filtering data from the 307 patients registered in the system, a final valid sample of 214 patients was obtained (Fig. 1).

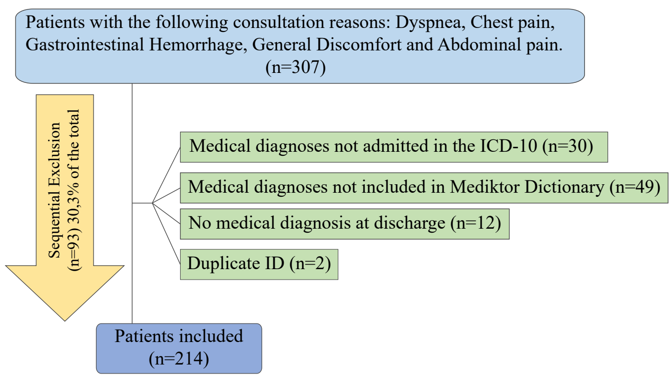


Fig. 1. Flow diagram of the methodology to identify the included patients.

Given the paucity of studies, to obtain a precision of 5.0% in the estimation of a proportion by means of a normal 95% bilateral asymptotic confidence interval, assuming a percentage of losses of 10% and with the proportion estimated in the literature of the 58% (26), a necessary sample size of 289 patients was estimated.

D. Variable Results

The main outcome variable was the agreement between the diagnosis established by the Mediktor software and the final diagnosis of the patient. The final diagnosis of the patient was defined as the one that would appear first in the patient's hospital discharge report, the Emergency discharge report, if this occurs, or the hospital discharge report for those patients who required hospital admission.

All patients were followed 30 days after the event to determine the secondary variables, that is, hospital readmission, emergency visit or death.

All the patients were in low complexity (First Assistance Unit, FAU) of the Emergency Department and the triage level awarded by the Manchester system was collected.

The length of stay was recorded from admission until the first medical evaluation was received, from admission to medical diagnosis until discharge.

E. Methods

The study was conducted over a period of three weeks, from March 15 to April 10, 2017, until the recruitment period was completed (n = 307 patients). All the patients who went to the Emergency Service of the San Carlos Clinical Hospital were selected as possible study subjects, from 8 am to 10 pm, for the indicated reasons for consultation. The patients were previously triads (Manchester 3, 4 or 5 levels) and located in the UPA following the usual process. After corroborating the inclusion/exclusion criteria, a nurse from the Innovation Unit specifically assigned to the study, was presented to the patients while they waited in the waiting room to be attended, and they were accompanied to an appropriate office for the interview. He explained clearly and without technicalities the objective of the study and its observational nature, emphasizing to the patient how his participation in the study would not interfere or influence the process, the length of stay, the diagnosis, the complementary tests or the treatment. He was offered the possibility to participate in the study by asking the patient and/or guardian for his informed consent to participate and was given a copy of it.

Once included, they were interviewed by the Mediktor software, in a computer exclusively dedicated to the study. The nurse performed the suggested questions and recorded the patient's own answers, clarifying any doubts that might arise, but without interfering with their answers. The interview guided by Mediktor began with the variables and the main reason for consultation recorded in the Manchester triage, followed by the sequence of questions until generating a level of complexity of the urgency and a list of possible pre-diagnoses. At the end of the interview, the patient was thanked for his participation without showing him the possible pre-diagnoses so as not to interfere in the usual care process, and he was asked to wait again in the waiting room to receive the usual assistance. The results collected by the research nurse were not shared with any healthcare professional during the data collection phase and were only shared with the rest of the team's researchers in the analysis phase after the closing of the data collection period.

After 30 days of the interview of the last patient recruited in the study, the information system records of the Emergency Service (SISU) of the HCSC were reviewed to corroborate the diagnosis of discharge or hospitalization, the level of triage, the revisits to emergencies and the times of stay in the emergency room. The HCSC Admission database (HP HIS) was reviewed.

F. Statistical Analysis

For descriptive purposes, quantitative variables will be expressed as means and standard deviation or median and interquartile range; and the qualitative variables will be expressed as number and frequencies. To check the main hypothesis, we have evaluated by means of a binomial test that compares the agreement of the Mediktor diagnosis against the possibility of random success for the 548 diagnoses included in the Mediktor database.

For the secondary hypotheses we use the Chi-square test or the Fisher exact test (in case more than 25% of the expected frequencies were less than 5). To evaluate the relationship between ordinal quantitative variables we have used the Spearman Correlation.

In all hypothesis contrasts, the null hypothesis with an error of less than 0.05 will be rejected. The statistical analysis will be carried out with the help of the statistical package SPSS 20.0®.

III. Results

In the descriptive analysis, we show the socio-demographic characteristics, reason for consultation, emergency room and follow-up of HUS patients (Table I).

TABLE I. Description of the Sample

		n (214 Patients)
Sex	Male	88 (41, 1)
	Female	126 (58,9)
Age		52, 13 (19, 41)
Reason for Consultation	Abdominal Pain	94 (43, 9)
	General Discomfort	60 (28,)
	Chest Pain	34 (15,9)
	Dyspnea	26 (12,1)
Destiny	Hospital Discharge	193 (90,2)
	Hospital Admission	21 (9,8)
Length of stay (minutes)	Length of Stay until first medical examination	45,00(34,75 – 66,00)
	Length of the stay until the medical diagnosis	215 (159 – 314,75)
	Length of stay at emergency department	283,50 (200,75 – 433,25)
	- Length of stay at emergency department/ Hospital Admission Patients	471 (351 – 1000,50)
	- Length of stay at emergency department /Hospital Discharge Patients	260 (196,50 - 389)
Follow-up	Re-admission/Hospital Discharge up to 30 days	6 (3,1%)
	New visit to ED/Hospital Discharge up to 30 days	28(14,5%)
	Mortality up to 30 days	0 (0%)

A. Diagnostic Agreement

The main objective of the study was to evaluate the concordance of the Mediktor diagnoses with the final diagnosis of the HUS patients, as a possible tool to support the decision. Because Mediktor offers a total of 10 pre-diagnoses classified by probability, we have divided our results into 4 different levels of agreement.

If we took any of the 10 pre-diagnoses (Top10), the level of agreement was 76.5%. Considering any of the first 5 pre-diagnoses (Top5), the level of accuracy was 65.4% and for any of the first 3 pre-diagnoses (Top3) it was 58%. The exact correctness in the first pre-diagnosis (Top1) was obtained in 37.9%. All of them were statistically significant p <0.001 (Table II).

TABLE II. Diagnostic Agreement

Position of the right diagnose at Mediktor´s Ranking	Marginal % of Accuracy	Aggregated % of Accuracy
1º	37,9	**37,9**
2º	13,1	51
3º	7	**58**
4º	5,1	63,1
5º	2,3	**65,4**
6º	3,7	69,1
7º	2,3	71,4
8º	1,4	72,8
9º	2,3	75,1
10º	1,4	**76,5**

As secondary objectives, there was no statistically significant relationship between Mediktor's diagnostic concordance and the percentage of hospital admissions [Top1- 9.9% (χ^2 = 0.001; p = 0.981); Top3- 8.9% (χ^2 = 0.296; p = 0.587); Top5- 8.6% (χ^2 = 0.705, p = 0.401) and Top10- 9.1% (χ^2 = 0.353, p = 0.553)].

There were no statistically significant differences in other secondary objectives, such as the relationship between diagnostic concordance and readmissions in the emergency department [Top1- 2.5% (χ^2 = 1907, p = 0.167); Top 3- 4.8% (χ^2 = 0.055, p = 0.815); Top 5- 5.0% (χ^2 = 0.016, p = 0.898); and Top 10- 4.9% (χ^2 = 0.099, p = 0.753)].

Because we did not have any deaths in the sample, we can´t determine the relationship between the number of emergency visits and mortality during the first 30 days after discharge.

B. Comparison between Triage Manchester and Mediktor

We determined the relationship between the level assigned by the Manchester triage and that of Mediktor, and there was no statistically significant relationship between them (s = 0.059, p = 0.442), so it is necessary to deduce that they perform a different triage. We show the correspondence of the triages in Table III.

Table III. Classification Correspondences between Triage Manchester and Mediktor

		Triage Mediktor					
		1	*2*	*3*	*4*	*5*	*Total*
Triage Manchester	*3*	4 (5,60%)	11 (15,50%)	33 (46,50%)	20 (28,20%)	3 (4,20%)	71
	4	1 (0,70%)	26 (19,30)	66 (48,90%)	31 (23%)	11 (8,10%)	135
	5	1 (12,50%)	0 (0%)	2 (25%)	3 (37,50%)	2 (25%)	8
	Total	6 (2,8%)	37 (17,30%)	101 (47,20%)	54 (25,20%)	16 (7,50%)	214

Regarding the correspondence between levels 3, 4 and 5 of the Manchester and Mediktor systems, more than half of the patients (61.1%) classified by the Manchester system in Level 4 correspond to Level 3 of Mediktor, while 28.7% are in Level 4. More than half of the patients classified in Level 3 by the Manchester system (58.9%) are classified by Mediktor in Level 3, 35.7% in level 4 and 5.4% in level 5. The correspondences found (Fig. 2) do not reach statistical significance. (p = 0.211).

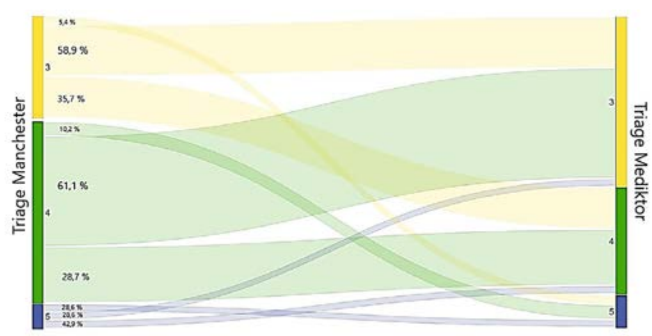


Fig. 2. Classification correspondences in levels 3, 4 and 5.

In the group of patients discharged (Fig. 3), the patients classified with the greatest severity (Level 3) by Mediktor presented a higher percentage of hospital admission (76.5%) with respect to Levels 4 and 5 (11.8%), respectively. The Manchester system presents a statistically significant higher percentage of hospital admission in Level 4 (52.4%) than in Level 3 (42.9%). In neither of the two cases is there a significant relationship (p = 0.034).

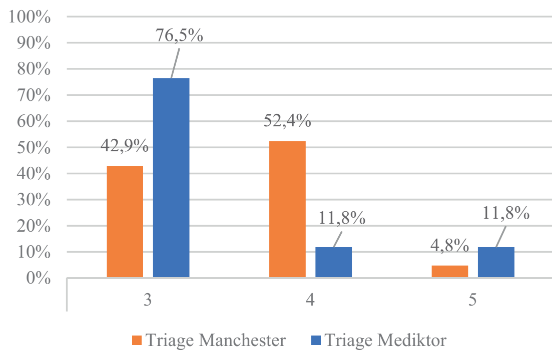


Fig. 3. Hospital Admission according to classification of triage.

The stay in the emergency room from admission until discharge or hospital admission of the patient had a median of 285.5 min. (RIC; 200-434.25). The stay in the emergency room until the first consultation by the emergency doctor registered a median of 45 min. (RIC, 33,75-66,00), and the stay until the diagnosis a median of 218 min. (RIC, 159.75-324.25). The stay in emergency was greater in the patients who were admitted [476 min. (RIC, 356-1302, 75)] compared to patients discharged [median 263 min. (RIC; 199.25-389). Significant statistical difference was found between these groups. ($p < 0.001$).

C. Possible Predictive Variables of Hospital Admission

We determined the difference between patients who presented one or more possible predictive variables of admission in previous studies and those who did not, without finding significant differences ($\chi^2 = 3.91$, $p = 0.272$). There was no statistically significant relationship between the possible predictive variables of admission and the patient's hospital admission:

- Take 3 or more drugs per day: 11.1% of hospital admission (IH) ($\chi^2 = 0.303$, $p = 0.582$).
- Need for help with basic activities: 0% of IH ($\chi^2 = 1.032$, $p = 0.310$).
- Need for more help than usual after the start of the acute process that he visits in the Emergency Room: 40% of IH ($\chi^2 = 4.01$, $p = 0.405$).
- Memory problems: 0% of IH ($\chi^2 = 0.137$, $p = 0.712$).
- Good vision: 16.7% of IH ($\chi^2 = 0.303$, $p = 0.582$).
- Hospital admission in the last 6 months: 16% of IH ($\chi^2 = 0.569$, $p = 0.451$).

D. Tracking

The follow-up variables were measured in the group of discharged patients. In Fig. 4 it is shown that the Mediktor classification presents the highest number of hospital readings at 30 days in level 3 (83.3%) and decreases in the following levels (Level 4: 16.70% and in Level 5: 0%). In contrast to the Manchester Triage system, which presents a higher level of readmission, Level 4 (66.70%), followed by 33.30% at level 3. In any case, there were no statistically significant differences ($p = 0.361$) between the classification scales and the patient's hospital readmission.

In the 30-day urgency follow-up (Fig. 5), the Mediktor triage behaviour indicates that the greater the urgency in the classification, the higher the percentage of emergency department visits (Level 3: 50%, Level 4: 45.8% and Level 5: 4.2%). Unlike the Manchester System, which presents a higher percentage of revisits to emergencies at 30 days in Level 4 (75%), followed by 21.4% in Level 3 and finally 3.6% in Level 5. No there were statistically significant differences ($p = 0.585$) between the classification scales and the 30-day patient's emergency visit.

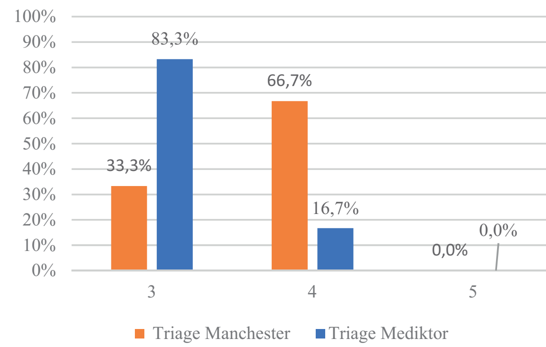


Fig. 4. Hospital readmissions at 30 days in discharged patients.

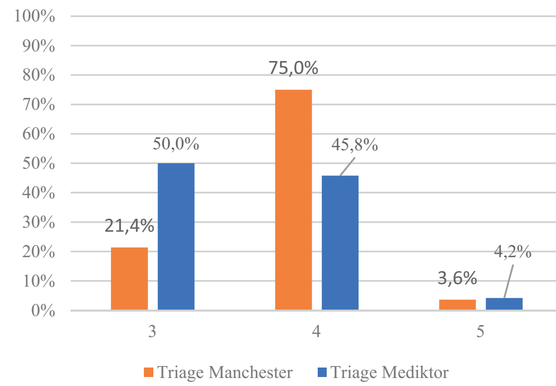


Fig. 5. New visit to ED in 30 days in patients discharged.

IV. Discussion

The present work did not include all emergency episodes, but only those initially classified in Manchester triage as low priority and treated in the First Assistance Unit, which excluded specialties such as paediatrics, traumatology, gynaecology and obstetrics, psychiatry, ophthalmology, otorhinolaryngology and all those patients who were initially located directly in the Acute Ward, corresponding to Levels I and II of triage. This means that, as indicated in the presentation of the project, the conclusions cannot be generalized to any emergency patient.

In our study, the percentage of hospital admission was 9.8%, and severity levels according to the Manchester triage system were not directly proportional to the percentage of hospital admission (Yellow - Level 3: 42.9%, Green - Level 4: 52.4% and Blue - Level 5: 4.8%). It must be considered that triage with the Manchester System is a process and not a result and its main objective is to reduce the variability of urgent care by assigning the most appropriate level for that care, and that its validity is susceptible to changes related to the age and other circumstances [27]. The complexity of the patient does not depend exclusively on the level of urgency of the triage, but on other variables, such as age, reason for consultation, comorbidity, baseline and cognitive capacity. In addition, it uses exclusively clinical discriminators; For example, severe pain is a subjective discriminator that classifies the patient at a high level of severity, even if hospitalization is not necessary. The behaviour of Mediktor in the classification of severity levels is close to what is expected with the percentage of hospital admission (Level 3: 76.5%, Level 4: 11.8% and Level 5: 11.8%). The scope of Mediktor in the collection of data is greater, so a different classification model is expected, although without statistically significant differences in this sample size based on a study that used case simulations [26] when not finding similar studies.

In the development of this pilot, the actual validated sample corresponds to 69.7% of the sample calculated a priori, with a percentage of loss that rose to 30.3% due to different causes (Fig. 1). Future studies with a larger number of patients will be necessary, considering the percentage of losses from this study and previous studies [29].

Regarding the diagnostic agreement, previous research suggests that, at the present time, symptom analysers may be less effective than physicians in terms of diagnostic accuracy [29], to conclude that the superiority of physicians over symptom analysers could be, in fact, excessive, especially if what we are considering as "gold standard" is the emergency medical diagnosis that, according to some studies, can pick up errors in 18% of patients [30] and could be total or partially true, compared to the final diagnosis of the patient, in only 71.4% of the cases [31]. The implantation of electronic medical records could imply a significant improvement in diagnostic accuracy and safety, although their results vary depending on the scenarios in which it is applied and the profile of the professionals [32, 33], especially at a time when the use of new data extraction technologies for natural language processing in the health field allows us to better understand our health actions [34].

This is because the diagnostic accuracy of the symptom analysers can be improved after appropriate feedback. One reason that the diagnostic accuracy is not optimal is that there is no accurate database to continuously feedback the diseases in the database. When continuous feedback is obtained from diagnoses and symptoms validated by physicians, the symptom analysers will update the probabilistic weights of their database, making the diagnostic agreement continuously improve. Soon, symptom analysers will be able to provide useful information for the performance of clinical work (e.g., regarding sensitivity and specificity of symptoms for diseases, prevalence of diseases) closing the circle of collaboration between professionals-new technologies. Therefore, symptom analysers could be useful tools for medical personnel and other health professionals [20].

The variables that we use from the hospital admission prediction model have not undergone a multicentre validation process that includes all patients who come to the emergency room. On the other hand, in the original publication [12] it was not valued independently in the different levels of classification, so we did not have previous data of its possible applicability to the group of less complexity studied and, therefore, with lower forecast of need for hospital admission

In our project we used a nurse to feed the system with the patient's answers and answer their questions, without interfering with the symptoms that the patient wanted to expose. Herrick et al (26), in a study on the usability of a self-administered questionnaire for emergency diagnosis, found that 86% did not require assistance. Those who needed help were older (54 ± 19) years. In our sample, 30% are in the 50-69 age group. The errors in the participants of the study occurred during the interviews using a pencil instead of a touch screen. Despite this, in our project we opted for the help in the registration of the questionnaire.

In recent years, powerful initiatives have emerged to help physicians make decisions, in both diagnosis and therapeutics [35]. The results in this paper indicate that an improvement could be made in the Mediktor diagnostic agreement, as in similar devices, with the feedback of medical comments after its use, for its use as a diagnostic aid tool in the emergency department. With this improvement of easily expected agreement, the possibility of its use as a tool for the streamlining of the care process is opened. In a process with a high level of basic protocolization, such as emergency care, a rapid initial approach to probable syndromic diagnoses would allow the suggestion of basic protocolized complementary tests, especially interesting in a healthcare setting in which the presence of personnel in training is frequent. This could represent a contribution of a tool such as Mediktor as a complementary element to the triage, to improve the adequacy of complementary tests requested in the emergency, to improve the initial diagnostic orientation, to generate a time saving in the urgency process and, finally, to improve the safety in the assistance of the external hospital emergency.

V. Conclusions

The level of concordance of the first ten diagnoses suggested by Mediktor with respect to the final diagnosis of the patients was 76.5%, higher than that published in similar programs. Its accuracy, like that of other comparable programs, is amenable to improvement with feedback from actual final diagnoses and medical comments.

The assignment of Mediktor levels, in this group of urgent patients of low complexity, does not coincide with that assigned by the Manchester system, which could indicate a different classification model. The classification of Mediktor in this segment of patients shows that a higher level of severity corresponds to a greater number of hospital admissions, hospital readmissions and revisits to emergencies at 30 days, although without statistical significance.

It is expected that the application has utility as a complement to the triage, to streamline the diagnostic approach, improve the adequacy of request for complementary tests and reduce waiting times in the emergency services in a protocolized action model.

References

[1] Peiró S, Librero J, Ridao M, Bernal-Delgado E. Variabilidad en la utilización de los servicios de urgencias hospitalarios del Sistema Nacional de Salud. Gac Sanit. January 2010;24(1):6-12.

[2] Sarría Santamera A, Prado Galbarro J, Quintana Díaz M, Martínez Virto A. [Determinants of utilization of emergency departments in Spain]. Aten Primaria. July 2015;47(6):386-8.

[3] Informe Anual del Sistema Nacional de Salud 2016. https://www.msssi.gob.es/estadEstudios/estadisticas/sisInfSanSNS/tablasEstadisticas/InfAnSNS.htm Access 25/03/2018.

[4] Jiménez JG. Clasificación de pacientes en los servicios de urgencias y emergencias: Hacia un modelo de triaje estructurado de urgencias y emergencias. Emergencias. 2003;15:165-74.

[5] Miró Ò, Salgado E, Bragulat E, Junyent M, Asenjo MA, Sánchez M. Estimación de la actividad en urgencias y su relación con la provisión de camas de hospitalización. Med Clínica. 2006;127(3):86-9.

[6] Martín-Sánchez FJ, González-Del Castillo J, Zamorano J, Candel FJ, González-Armengol JJ, Villarroel P, et al. El facultativo, un elemento necesario en el triaje de un Servicio de Urgencias en un hospital terciario. Emergencias. 2008;20:41-7.

[7] Dent AW, Weiland TJ, Vallender L, Oettel NE. Can medical admission and length of stay be accurately predicted by emergency staff, patients or relatives? Aust Health Rev. 2007;31(4):633-41.

[8] Clesham K, Mason S, Gray J, Walters S, Cooke V. Can emergency medical service staff predict the disposition of patients they are transporting? Emerg Med J EMJ. October 2008;25(10):691-4.

[9] Beardsell I, Robinson S. Can emergency department nurses performing triage predict the need for admission? Emerg Med J. 2010;emj-2010.

[10] Burch V, Tarr G, Morroni C. Modified early warning score predicts the need for hospital admission and inhospital mortality. Emerg Med J. 2008;25(10):674-8.

[11] Van der Wulp I, Schrijvers A, Van Stel H. Predicting admission and mortality with the Emergency Severity Index and the Manchester Triage System: a retrospective observational study. Emerg Med J. 2009;26(7):506-9.

[12] Elvira MC, Fernández C, González DCJ, González AJ, Villarroel P, Martín SF.Modelo predictor de ingreso hospitalario a la llegada al servicio de Urgencias [Prediction model for in-hospital admission in patients arriving in the Emergency Department]. An Sist Sanit Navar. 2012 May-Aug;35(2):207-17.

[13] Alonso CF, Armengol JJG, Perdigones J, Ferrer MEF, del Castillo JG, Sánchez FJM. La utilidad de la escala Identification of Seniors at Risk

(ISAR) para predecir los eventos adversos a corto plazo en los pacientes ancianos dados de alta desde una unidad de corta estancia. Emerg Rev Soc Esp Med Urgenc Emerg. 2015;27(3):181-4.

[14] Kovacs G, Croskerry P. Clinical decision making: an emergency medicine perspective. Acad Emerg Med. 1999;6(9):947-52.

[15] Schwartz LR, Overton DT. Emergency department complaints: a one-year analysis. Ann Emerg Med. 1987;16(8):857-61.

[16] Dunn JD. Malpractice in the emergency department. Ann Emerg Med. 1985;14(1):87-8.

[17] Schiff GD, Hasan O, Kim S, Abrams R, Cosby K, Lambert BL, et al. Diagnostic error in medicine: analysis of 583 physician-reported errors. Arch Intern Med. 2009;169(20):1881-7.

[18] Newman-Toker DE, Pronovost PJ. Diagnostic errors—the next frontier for patient safety. Jama. 2009;301(10):1060-2.

[19] Farmer SEJ, Bernardotto M, Singh V. How good is Internet self-diagnosis of ENT symptoms using Boots WebMD symptom checker? Clin Otolaryngol. 2011;36(5):517-8.

[20] Semigran HL, Levine DM, Nundy S, Mehrotra A. Comparison of Physician and Computer Diagnostic Accuracy. JAMA Intern Med. 2016;176(12):1860.

[21] Morita T, Rahman A, Hasegawa T, Ozaki A, Tanimoto T. The Potential Possibility of Symptom Checker. Int J Health Policy Manag. 05 de 2017;6(10):615-6.

[22] Wyatt JC. Fifty million people use computerised self-triage. BMJ. 8 de julio de 2015;351:h3727.

[23] Miller RA. Medical diagnostic decision support systems—past, present, and future: a threaded bibliography and brief commentary. J Am Med Inform Assoc. 1994;1(1):8-27.

[24] Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. Bmj. 2005;330(7494):765.

[25] Benaroia M, Elinson R, Zarnke K. Patient-directed intelligent and interactive computer medical history-gathering systems: a utility and feasibility study in the emergency department. Int J Med Inf. 2007;76(4):283-8.

[26] Herrick DB, Nakhasi A, Nelson B, Rice S, Abbott PA, Saber Tehrani AS, et al. Usability characteristics of self-administered computer-assisted interviewing in the emergency department: Factors affecting ease of use, efficiency, and entry error. Appl Clin Inform. 2013;4(2):276-92.

[27] Zachariasse JM, Seiger N, Rood PPM, Alves CF, Freitas P, Smit FJ, Roukema GR, and Moll HA. Validity of the Manchester Triage System in emergency care: A prospective observational study. PLoS One. 2017; 12(2): e0170811.Published online 2017 Feb 2. doi: 10.1371/journal.pone.0170811.

[28] Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. bmj. 2015;351:h3480.

[29] Barriga EM, Ferrer IP, Sánchez MS, Baranera MM, Utset JM. Experiencia de Mediktor®: un nuevo evaluador de síntomas basado en inteligencia artificial para pacientes atendidos en el servicio de urgencias. Emerg Rev Soc Esp Med Urgenc Emerg. 2017;29(6):391-6.

[30] Chiu HS, Chan KF, Chung CH, Ma K, Au KW. A comparison of emergency department admission diagnoses and discharge diagnoses: retrospective study.Hong Kong j.emerg.med. 2003;10:70-75

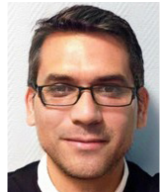
[31] Fordyce J, Blank FS, Pekow P, Smithline HA, Ritter G, Gehlbach S, Benjamin E, Henneman PLErrors in a busy emergency department.Ann Emerg Med. 2003 Sep;42(3):334-6.

[32] Ben-Assulia O, Sagib D,Leshnoc M, Ironib A, ZivImproving A. Diagnostic accuracy using EHR in emergency departments: A simulation-based study. Journal of Biomedical Informatics 55 (2015) 31–40.

[33] Rowe AK, de Savigny D, Lanata CF, Victora CG. How can we achieve and maintain high-quality performance of health workers in low-resource settings? Lancet Lond Engl. 2005;366(9490):1026-35.

[34] Medrano IH, Guijarro JT, Belda C, Ureña A, Salcedo I, Espinosa-Anke L, et al. Savana: Re-using Electronic Health Records with Artificial Intelligence. International Journal of Interactive Multimedia and Artificial Intelligence. 2018;4(Special Issue on Big Data and e-Health).

[35] González-Ferrer A, Seara G, Cháfer J, Mayol J. Generating Big Data Sets from Knowledge-based Decision Support Systems to Pursue Value-based Healthcare. International Journal of Interactive Multimedia and Artificial Intelligence. 2018;4(Special Issue on Big Data and e-Health).



Juan Carlos Nazario Arancibia

Born in Trujillo (Perú) in February 1981. Graduated in Medicine from the Universidad Antenor Orrego Medical School. Specialization in Preventive Medicine and Public Health at Hospital Universitario Clínico San Carlos in Madrid. Master´s Degree in Public Health by Escuela Nacional de Sanidad, Madrid. Master´s Degree in HealthCare Administration by Universidad Pompeu Fabra - Barcelona School of Management. He currently serves as member of Innovation Unit at Hospital Clínico San Carlos as Innovation Manager. Research interests: Innovation, Agile methodologies, Lean Healthcare, Big Data.

Francisco Javier Martín-Sánchez

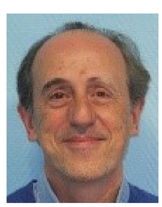
Born in Madrid, Graduated in Medicine in 1999 from Universidad Autónoma de Madrid Medical School. Specialization in Geriatrics at Hospital Universitario Clínico San Carlos 2000-2004, PhD (Universidad Complutense, Madrid). He is the member of the ICA-SEMES and GREAT group. His field of research has focused on the field of geriatric cardiology and the author of several publications in the area of acute heart failure in the elderly.

Angel Luis Del Rey-Mejías

Born in Madrid, February 1982. Superior Technician in Administration of computer systems in Arboleda (2004). Graduated in Psychology in 2010 from Universidad Complutense de Madrid (UCM) Psychology School. Master Degree in Methodology of behavior and health sciences in Universidad Autónoma de Madrid (2012) and Master Degree in Mental health research at CIBERSAM (2014). He currently serves as member of Innovation Unit at Hospital Clínico San Carlos as Innovation Manager. Publications (2017): Parellada, M., Pina-Camacho, L., Moreno, C., Aleman, Y., Krebs, M.O., Desco, M., Merchán-Naranjo, J., Del Rey-Mejías, A., Boada, L., Llorente, C. and Moreno, D., 2017. Insular pathology in young people with high-functioning autism and first-episode psychosis. Psychological medicine, 47(14), pp.2472-2482. Esguevillas, Á., Díaz-Caneja, C.M., Arango, C., Del Rey-Mejías, Á.L., Bernardo, E.G., Delgado, C., Clarkin, J.F. and Carrasco, J.L., 2017. Personality organization and its association with clinical and functional features in borderline personality disorder. Psychiatry research. Research interests: Research, Innovation, Psychiatry, Machine Learning, Big Data, Business Intelligence.

Juan González del Castillo

Born in Madrid, July 1973, Graduated in Medicine in 1997 from Universidad de Extremadura Medical School. Specialization in Internal Medicine at Hospital Universitario Clinico San Carlos (HUCSC) 1999-2004, PhD (Universidad Complutense, Madrid) in 2006. Coordinator of the Infectious Disease Group of the Spanish Emergency Medicine Society. Member of the "Revista Española de Quimioterapia" journal editorial board. Section Editor of Emergencias Journal. Board member of the Spanish Quimiotherapy Society. Senior doctor in the Emergency Department, Hospital Clinico San Carlos. One hundred and twenty-six articles published in indexed journals.



Joan Chafer Vilaplana

Born in Barcelona in 1958. Graduated in Medicine in 1958. Specialist in Internal Medicine since 1981 after an MIR program at the Hospital Clinico San Carlos. Worked in Clinical medicine and research at King's College Hospital in London, The Royal London Hospital and back in Madrid at the liver unit of the Fundación Jiménez Díaz. Since 1999 works in research management and innovation, initially at Hospital Universitario La Paz and currently at the Innovation Unit of the Hospital Clinico San Carlos. Between 2008 an 2012 was also responsible for IT and Information Systems at the latter hospital.

Miguel Angel García Briñón

Born in Salamanca, March 1970, nursing in 1994 from the Universidad de Salamanca, nursing degree from Universidad Complutense de Madrid in 2011, master's in management of health services and health companies from Universidad Complutense de Madrid in 2012.I worked as a nurse from 1996 to 2000 in different intensive care units and until 2010 in critical care of the San Carlos Hospital in Madrid. Nurse Emergency Supervisor Department of San Carlos Hospital from 2010. Member of the research group "OGS 25 pathology of urgencies and emergencies" from 2011.

María del Mar Suárez-Cadenas

Born in Almería in 1993. Graduated in 2015 by the University of Almería with a degree in Nursing. Official Master in Sexology, sexual education and sexological advice by the University Camilo José Cela in 2017. Has worked as a nurse in different emergency department since 2015 until now. Currently part of a researching group "OGS 25 pathology of urgencies and emergencies".

Germán Seara

Born in Madrid, March 1951, Graduated in Medicine in 1975 from Universidad Complutense de Madrid Medical School. Specialization in Pediatrics at Hospital Universitario Clínico San Carlos (HUCSC) 1976-80, Madrid. FIS Scholarship for Extension Studies at Northwick Park Hospital & Clinical Research Center, Harrow, UK, Top Managerial Programme in King's Fund, London/UK and ENS, Madrid, Spain. He was former Chief Medical Officer at Hospital Clinico San Carlos and in other hospitals and he is currently appointed to the Innovation Unit of HUCSC. Recent Publications: Integración clínica en el paciente crónico, Carretero L, Comes N, Borrás A, Rodríguez Balo A y Seara G. Enferm Clin. 2014;24(1):35-43. G. Seara, A. Payá, J. Mayol, Value-Based Healthcare Delivery in the Digital Era. European Psychiatry March 2016, Vol 33S pp S46. Generating Big Data Sets from Knowledge-based Decision Support Systems to Pursue Value-based Healthcare A González-Ferrer1, G Seara, J Cháfer, J Mayol. IJIMAI 2018;IV(7):42-6. Research interests: health care, innovation. Dr. Seara memberships: Asociación Española de Pediatría.

Julio Mayol

Born in Madrid, July 22nd, 1963, Graduated with honors in Medicine in 1988 and PhD summa cum laude in Medicine in Medicine in 1992, both from Universidad Complutense de Madrid Medical School. Specialization in general surgery at Hospital Clínico San Carlos 1991- 1995. Fellow in Surgery at Beth Israel Deaconess Medical Center - Harvard Medical School. Attending surgeon. Associate Professor of Surgery. Chief of the Division of Colorectal Surgery. Full Professor of Surgery at Universidad Complutense de Madrid Medical School. Visiting Professor at Wayne State University. Chief Medical Officer at Hospital Clinico San Carlos. Publications: 1. Chapman SJ, Mayol J, Brady RR. Twitter can enhance the medical conference experience. BMJ. 2016 jul 19;354: i3973. - Mayol Martinez J. Innovation in Surgery. Cir Esp. 2016 Apr;94(4):207-9. - Maeso S, Reza M, Mayol JA, Blasco JA, Guerra M, Andradas E, Plana MN. Efficacy of the Da Vinci Surgical System in Abdominal Surgery Compared with That of Laparoscopy: A Systematic Review and Meta- Analysis. Ann Surg. 2010 Aug;252(2):254-62. PMID: 20622659. Research interests: colorectal surgery, biomedical technology, innovation, Prof. Julio Mayol memberships: American Gastroenterological Association, Society for Surgery of the Alimentary Tract, Asociación Española de Cirujanos, Sociedad Española de Investigaciones Quirúrgicas.

Hybrid Algorithm for Solving the Quadratic Assignment Problem

Mohammed Essaid Riffi, Fatima Sayoti*

LAROSERI Laboratory, Dept. of Computer Science, Faculty of Sciences, University of Chouaib Doukkali, El Jadida (Morocco)

Received 1 June 2017 | Accepted 27 September 2017 | Published 20 October 2017

Abstract

The Quadratic Assignment Problem (QAP) is a combinatorial optimization problem; it belongs to the class of NP-hard problems. This problem is applied in various fields such as hospital layout, scheduling parallel production lines and analyzing chemical reactions for organic compounds. In this paper we propose an application of Golden Ball algorithm mixed with Simulated Annealing (GBSA) to solve QAP. This algorithm is based on different concepts of football. The simulated annealing search can be blocked in a local optimum due to the unacceptable movements; our proposed strategy guides the simulated annealing search to escape from the local optima and to explore in an efficient way the search space. To validate the proposed approach, numerous simulations were conducted on 64 instances of QAPLIB to compare GBSA with existing algorithms in the literature of QAP. The obtained numerical results show that the GBSA produces optimal solutions in reasonable time; it has the better computational time. This work demonstrates that our proposed adaptation is effective in solving the quadratic assignment problem.

Keywords

Combinatorial Optimization, Golden Ball Algorithm, Simulated Annealing, Quadratic Assignment Problem.

DOI: 10.9781/ijimai.2017.10.003

I. Introduction

The quadratic assignment problem (QAP) is one of the known classical combinatorial optimization problems, in 1976 Sahni and Gonzalez [1] proved that the QAP belongs to the class of NP-hard problems [1]. It was introduced for the first time by Koopmans and Beckmann in 1957 [2]; its purpose is to assign n facilities to n fixed locations with a given flow matrix of facilities and distance matrix of locations in order to minimize the total assignment cost. This problem is applied in various fields such as hospital layout [3], scheduling parallel production lines [4] and analyzing chemical reactions for organic compounds [5].

Many recent hybrid approaches have improved performance in solving QAP such as genetic algorithm hybridized with tabu search method [6], ant colony optimization mixed with local search method [7] and ant colony optimization combined with genetic algorithm and local search method [8]. Recently the hybrid algorithms are much proposed and used by many researchers to find optimal or near optimal solutions for the QAP.

In this paper we propose a new competitive approach when compared with other existing methods in the literature. The golden ball algorithm mixed with simulated annealing (GBSA) is considered here as a hybrid metaheuristic to apply in the quadratic assignment problem.

This work presents an efficient adaptation of GBSA algorithm to the quadratic assignment problem (QAP). This algorithm is based on the concept of soccer; it guides the search by simulated annealing [9] to escape from the local optima. The suggested technique has never been proposed or tested with QAP. In this research we use some small, medium and large test problems for comparing our approach to other recent methods from literature. Our approach is able to explore effectively the search space; it reaches the known optimal solutions in less time.

The rest of this paper is structured as follows: In section I, Introduction. In section II, Quadratic assignment problem formulation. In section III, Methods. In section IV, Results and discussion. In section V, Conclusion.

* Corresponding author.

E-mail address: fatimasayoti@gmail.com

II. Quadratic Assignment Problem

The QAP [1] can be defined as a problem of assigning n facilities to n locations, with given flows between the facilities and given distances between the locations (Fig.1).

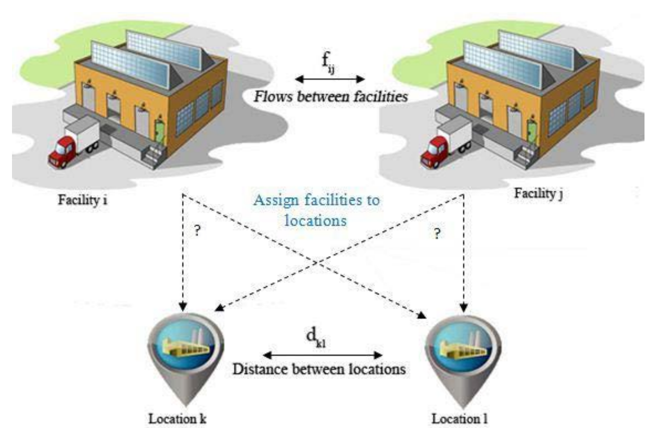


Fig. 1. Quadratic Assignment Problem.

The purpose is to assign the facilities to the locations in such a way that the total cost is minimized. Each facility must be placed just at one location.

We consider two n × n matrices, the flow matrix $F=f_{ij}$ and the distance matrix $D=d_{kl}$. The QAP formulation is given as follows (1):

$$\min_{\pi \epsilon S_n} Cost = \sum_{i=1}^{i=n} \sum_{j=1}^{j=n} f_{ij}\, d_{\pi(i)\,\pi(j)} \tag{1}$$

S_n is the set of all permutation of n elements {1, 2, …, n}.

π(i) and π(j) are respectively locations of facilities *i* and *j*, we suppose that *π(i)=k* and *π(j)=l.*

$f_{ij}\, d_{\pi(i)\,\pi(j)}$ is the cost of assigning facility *i* in location *k* and facility *j* in location *l.*

The objective function (Cost) must be minimized.

Several algorithms are usually used to solve the quadratic assignment problem:

- Exact algorithms such as branch and bound algorithm [10] and branch and cut algorithm [11].
- Metaheuristics such as genetic algorithm [12],[13],[14], tabu search method [15],[16],[17], simulated annealing algorithm [9], ant colony optimization [18] and particle swarm optimization [19].

In recent year, metaheuristic algorithms are used in solving the QAP more than the exact algorithms which are unable to solve the hard instances of QAP in a reasonable time. Many researchers compared between different metaheuristic algorithms for solving the QAP [20], [21].

III. Methods

A. Golden Ball Metaheuristic

The GB technique is a metaheuristic proposed by E.Osaba et al. [22],[23]. It uses different principles of soccer to solve combinatorial optimization problems. The quality of this technique is demonstrated applying it to four combinatorial problems [23]: Asymmetric traveling salesman problem (ATSP) [24], Vehicle Routing Problem with Backhauls (VRPB) [25],[26], n-Queen Problem (NQP) [27], One-Dimensional Bin Packing Problem (BPP) [28].This algorithm is a promising metaheuristic to solve combinatorial optimization problems [23].

In Golden Ball algorithm, groups of solutions are considered as soccer teams which are composed of a fixed number of players, the captain of team plays the rule of the best solution of the group. Each team has a coach who determines the type of training to improve the efficiency of its team. There are two types of training: conventional training and custom training. As shown in Fig. 2, the concept of this method is based on four main phases: initialization phase, training phase, competition phase and transfer phase.

In the initialization phase, we set the value of the number of teams (NT) and the number of players per team (PT). We assign randomly to each team a coach.

In the training phase, all teams must train by following a specific type of training. The conventional training is the daily training of a team. When a team becomes unable to improve its capacities, in this case, it must follow a custom training.

In the competition phase, each team must compete with other team chosen randomly. The winning team receives three points, in the case of equality; both teams receive one point. The accumulated points will be used to order the teams in descending order.

In the transfer phase, we detect three cases of transfer:

Season transfer: during the season, all teams must be sorted in the descending order according to the strength value.

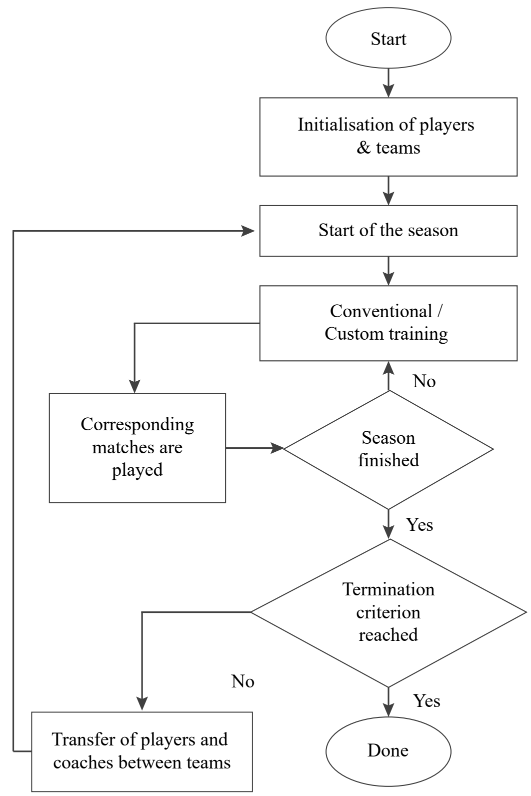


Fig. 2. Flowchart of Golden Ball algorithm.

The strength value is calculated using the following formula (2):

$$\frac{\sum_{i=1}^{PT} qij}{PT} \tag{2}$$

q_{ij} is the quality of player *i* of team *j*

All teams exchange their players in this way: the best player of the first team must be replaced by the worst player of the last team. This worst player will be replaced by this best player.

The best player of the second team must be replaced by the worst player of the penultimate team. This worst player will be replaced by this best player and so forth.

Special transfer: When a player of a given team is unable to improve after a custom training, the team must exchange it with a player of another team chosen randomly.

Cessation of coaches: after having ordered all the teams in descending order according to their accumulated point, the weaker teams must change their conventional training by another randomly selected.

The GB algorithm was tested by E.Osaba et al. with four different combinatorial optimization problems [23]. The same technique was applied on the flow shop scheduling problem [29] and the job shop scheduling problem [30].

B. Simulated Annealing Method

The simulated annealing algorithm [14] is inspired by the physical annealing process which attempt to improve the quality of the solid by using at the beginning a high temperature T_0 at which the solid is in a liquid state. With the slow decrease of the temperature T (cooling phase) the solid regains its solid form (Fig. 3). Metropolis et al. show how to generate a sequence of successive states of the solid. The new state is accepted if the energy produced by this change of state decreases; otherwise, it is accepted with a probability defined by the following equation (3).

$$p = e^{-\Delta E/(c_b \times T)} \quad (3)$$

c_b is Boltzmann constant

ΔE is the energy difference produced by this change of state

T is the temperature of the solid

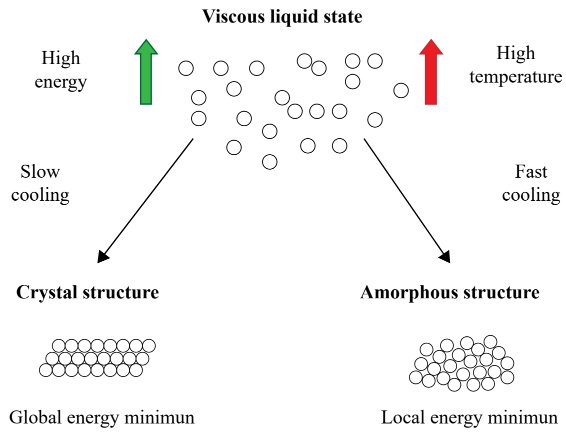


Fig. 3. Evolution of thermodynamic system.

The simulated annealing method [31] is one of the oldest algorithms; it is an iterative metaheuristic very used to solve combinatorial optimization problems in the continuous and discrete case. The strong point of this technique is to escape from the local minima and avoid the cyclic behavior. The performance of simulated annealing algorithm depends on a set of parameters which must be controlled. It means that the correct setting of the parameters produces satisfactory results.

IV. Adaptation of GBSA Algorithm

In the initialization phase we generate randomly the initial population of NT×PT solutions.

Each solution is represented in the following manner (Fig. 4):

Locations →	2	7	4	5	6	3	9	8	1
Facilities →	i=1	i=2	i=3	i=4	i=5	i=6	i=7	i=8	i=9

Fig. 4. Assign 9 facilities to 9 locations.

In the training phase, we used the following methods as conventional training functions:

2-opt [32], [33]: this iterative method is a local search algorithm, it repeatedly tries to improve the current assignment by exchanging two facilities.

Insertion method [34]: this method inserts a facility chosen randomly between two facilities.

Swapping mechanism [35]: this method swaps two parts selected randomly; the following figure (Fig. 5) explains the concept of this technique.

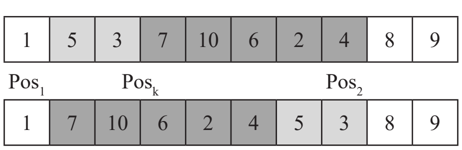


Fig. 5. Illustration of swapping mechanism.

As a custom training function the proposed adaptation used simulated annealing method [14],[31], it is used when the current solution is blocked in the local minima; it helps to accept some movement and escape from the local optimum.

Simulated annealing steps

S1:=Current solution

Generate a new solution based on the current solution. We used the swap of two random locations

S2:=New solution

f(S1):= cost of S1

f(S2):= cost of S2

if (f(S2)<f(S1))

S1:=S2

Else

Generate random number r in [0,1)

Calculate the value of p

$$p = e^{-(f(S2)-f(S1))/(T)}$$

If(r<p)

S1:=S2

Decrease the temperature value

Repeat all steps until T= 0.

V. Results and Discussion

The program was run 10 times on different instances of QAPLIB [36]. The GBSA algorithm was implemented in C language and compiled using Microsoft Visual Studio 2008, the program code was executed in computer with Genuine Intel(R) 575 @ 2.00 GHz 2.00 GHz RAM 2,00 Go.

The program uses three parameters: NT (number of groups), PT (number of schedules per group) and T (temperature).

The parameters values in the table below (Table I) produce better results during the algorithm run.

4×5 random solutions are sufficient to obtain good results.

TABLE I. Parameters values

NT	4
PT	5
T	40

At the high temperature, the simulated annealing method becomes unnecessary because proximally 50% of iterations accept decision at the high temperature [37]. In this paper we fixed the high temperature at 40 which is considered a symptom of fever in humans.

Table II represents the following information:

Optimal: Best known Solution

Best: Best permutation

NBest: The number of runs in which the algorithm reaches the best permutation

Worst: The worst permutation

Average: The average cost (= the sum of solutions cost obtained divided by 10)

The relative percentage deviation from the best known solution is calculated as follows (4):

$$\text{RPD} = \frac{\text{Average} - \text{Optimal}}{\text{Optimal}} \times 100\ \% \quad (4)$$

Time: Best time per seconds

TABLE II. Numerical Results of the GBSA Algorithm

Instance	Optimal	Best	Nbest	Worst	Average	%RPD	Time
Bur26a	5426670	5426670	10	5426670	5426670,00	0,00	0
Bur26b	3817852	3817852	10	3817852	3817852,00	0,00	0
Bur26c	5426795	5426795	10	5426795	5426795,00	0,00	1
Bur26d	3821225	3821225	10	3821225	3821225,00	0,00	2
Bur26e	5386879	5386879	10	5386879	5386879,00	0,00	1
Bur26f	3782044	3782044	10	3782044	3782044,00	0,00	1
Bur26g	10117172	10117172	10	10117172	10117172,00	0,00	0
Bur26h	7098658	7098658	10	7098658	7098658,00	0,00	1
Chr12a	9552	9552	10	9552	9552,00	0,00	0
Chr12b	9742	9742	10	9742	9742,00	0,00	0
Chr12c	11156	11156	10	11156	11156,00	0,00	0
Chr15a	9896	9896	10	9896	9896,00	0,00	0
Esc16a	68	68	10	68	68,00	0,00	0
Esc16b	292	292	10	292	292,00	0,00	0
Esc16c	160	160	10	160	160,00	0,00	0
Esc16d	16	16	10	16	16,00	0,00	0
Esc16e	28	28	10	28	28,00	0,00	0
Esc16f	0	0	10	0	0,00	0,00	0
Esc16g	26	26	10	26	26,00	0,00	0
Esc16h	996	996	10	996	996,00	0,00	0
Esc16i	14	14	10	14	14,00	0,00	0
Esc16j	8	8	10	8	8,00	0,00	0
Esc32a	130	136	01	140	139,40	7,23	240
Esc32b	168	168	10	168	168,00	0,00	0
Esc32c	642	642	10	642	642,00	0,00	0
Esc32d	200	200	10	200	200,00	0,00	0
Esc32e	2	2	10	2	2,00	0,00	0
Esc32g	6	6	10	6	6,00	0,00	0
Esc32h	438	438	10	438	438,00	0,00	0
Esc64a	116	116	10	116	116,00	0,00	0
Esc128	64	64	10	64	64,00	0,00	65
Had12	1652	1652	10	1652	1652,00	0,00	0
Had14	2724	2724	10	2724	2724,00	0,00	0
Had16	3720	3720	10	3720	3720,00	0,00	0
Had18	5358	5358	10	5358	5358,00	0,00	1
Had20	6922	6922	10	6922	6922,00	0,00	0
Nug12	578	578	10	578	578,00	0,00	0
Nug14	1014	1014	10	1014	1014,00	0,00	0
Nug15	1150	1150	10	1150	1150,00	0,00	0
Nug16a	1610	1610	10	1610	1610,00	0,00	0
Nug16b	1240	1240	10	1240	1240,00	0,00	0
Nug17	1732	1732	10	1732	1732,00	0,00	0
Nug18	1930	1930	10	1930	1930,00	0,00	0
Nug20	2570	2570	10	2570	2570,00	0,00	0
Rou12	235528	235528	10	235528	235528,00	0,00	0
Rou15	354210	354210	10	354210	354210,00	0,00	0
Rou20	725522	725522	08	725582	725534,00	0,00	1
Scr12	31410	31410	10	31410	31410,00	0,00	0
Scr15	51140	51140	10	51140	51140,00	0,00	0
Scr20	110030	110030	10	110030	110030,00	0,00	3
Sko42	15812	15880	01	16036	15969,00	0,99	240
Sko49	23386	23582	01	23736	23652,40	1,13	240
Tai12a	224416	224416	10	224416	224416,00	0,00	0
Tai15a	388214	388214	10	388214	388214,00	0,00	0
Tai15b	51765268	51765268	10	51765268	51765268,00	0,00	0
Tai17a	491812	491812	10	491812	491812,00	0,00	0
Tai20a	703482	703482	05	713260	706128,90	0,37	32
Tai20b	122455319	122455319	10	122455319	122455319,00	0,00	1
Tai25a	1167256	1181326	01	1193120	1187990,60	1,77	240
Tai30a	1818146	1841180	01	1867650	1858562,80	2,22	240
Tail40a	3139370	3215360	01	3251200	3233951,20	3,01	240
Tail50a	4938796	5084020	01	5143598	5113257,40	3,53	240
Tho30	149936	150578	01	151742	151189,20	0,83	240
Tho40	240516	243362	01	246172	244773,00	1,76	240

The program stops when the optimal solution is reached or when the execution time exceeds 240 seconds. We take two digits after the comma, for the results shown in the two columns: Average and the Relative Percentage Deviation %RPD.

As Table II shows, the proposed algorithm allows to obtain always the optimal solution of 81,25% of the instances tested in a time not exceeding three seconds. The %RPD of 93,75% of the instances does not exceed 2% and this clearly shows that the GBSA algorithm converges well to the optimal solution. According to the values shown in the Table II, when the value of %RPD is equal to 0.00%, this means that the program reaches exactly the optimal solution at least 8 times per 10 tests and in this case the best and the worst solution are often the same.

Abd El-Nasser et al. [38] presented a comparative study between Meta-heuristic algorithms: Genetic Algorithm (GA), Tabu Search (TS), and Simulated annealing (SA) for solving a real-life (QAP) and analyze their performance in terms of both runtime efficiency and solution quality [38].

The Fig. 6 compares the relative percentage deviation of some instances of QALIB for our proposed algorithm GBSA, GA, TS and SA. The result shows that GBSA has more quality than the other algorithms for solving the QAP. We can deduce that our proposed method has really improved SA's effectiveness in solving these instances which we have chosen as an example for our comparative study.

There exist two sets of problems in QAPLIB that represent a challenge for any proposed algorithm. These problems were introduced by Skorin-Kapov [39] and Taillard [40].

We selected 9 instances from Skorin-Kapov and 7 instances from Taillard. For this list of QAPLIB instances, we compared our proposed method with others recent methods such as: Memetic algorithm (BMA) [41], Breakout local search (BLS) [42] and Cooperative parallel tabu search algorithm (CPTS) [43]. The list of instances shown in Table III is a challenge for our algorithm.

We have fixed the maximum execution time of GBSA algorithm at 4 minutes. As the results depict (Table III), the GBSA algorithm needs some improvement to better solve some hard instances of QAP. But in general, the proposed algorithm seems promising to solve the quadratic assignment problem. According to the values of the relative percentage deviation from the best known solution, GBSA algorithm produces results near the global optimum in a reasonable time.

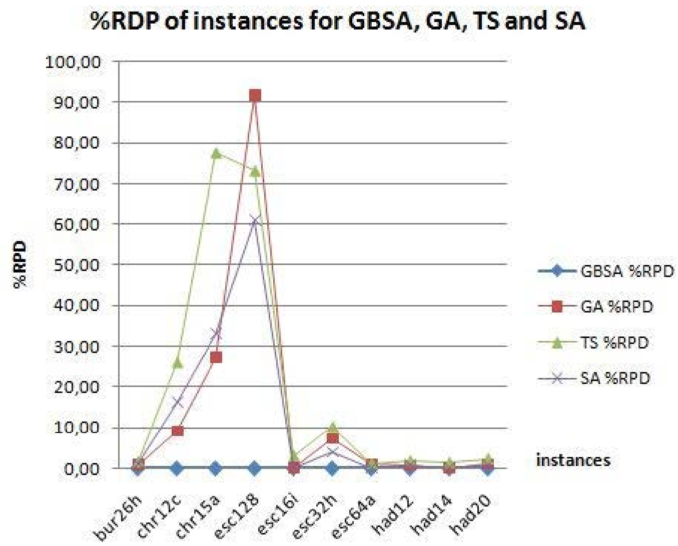


Fig. 6. %RDP of some instances for GBSA, GA, TS and SA algorithms.

VI. Conclusion

The GBSA algorithm is the result of the hybridization of two methods: golden ball metaheuristic and simulated annealing method. This new hybrid algorithm is based on soccer concepts; it incorporates and guides simulated Annealing technique to escape from the local minima and to find the global optimal solution. This method has never been proposed or tested on QAPLIB instances. In this work we proposed an adaptation of our strategy to solve the QAP. The numerical results indicate the efficiency of the proposed GBSA adaptation and its performance compared to algorithms in literature of QAP. As a result, we deduce that our proposed approach has a high convergence speed.

TABLE III. Comparison of GBSA Algorithm with Algorithms in the Literature of the QAP

		GBSA		BMA		BLS		CPTS	
Instance	BKS	%RPD	Time (m)	%RPD	Time (m)	%RPD	Time (m)	%RPD	Time (m)
Sko72	66256	0,543	4.0	0.000	3.5	0.000	4.1	0.000	69.6
Sko81	90998	0,481	4.0	0.000	4.3	0.000	13.9	0.000	121.4
Sko90	115534	0,614	4.0	0.000	15.3	0.000	16.6	0.000	193.7
Sko100a	152002	0,539	4.0	0.000	22.3	0.001	20.8	0.000	304.8
Sko100b	153890	0,679	4.0	0.000	6.5	0.000	10.8	0.000	309.6
Sko100c	147862	0,396	4.0	0.000	12.0	0.000	15.5	0.000	316.1
Sko100d	149576	0,760	4.0	0.006	20.9	0.001	38.9	0.000	309.8
Sko100e	149150	0,528	4.0	0.000	11.9	0.000	42.5	0.000	309.1
Sko100f	149036	0,704	4.0	0.000	23.0	0.000	17.3	0.003	310.3
Tai40a	3139370	3,012	4.0	0.059	8.1	0.022	38.9	0.148	3.5
Tai50a	4938796	3,532	4.0	0.131	42.0	0.157	45.1	0.440	10.3
Tai60a	7205962	2,870	4.0	0.144	67.5	0.251	47.9	0.476	26.4
Tai80a	13499184	2,965	4.0	0.426	65.8	0.517	47.3	0.691	94.8
Tai100a	21052466	2,771	4.0	0.405	44.1	0.430	39.0	0.589	261.2
Tai50b	458821517	0,285	4.0	0.000	1.2	0.000	2.8	0.000	13.8
Tai60b	608215054	0,147	4.0	0.000	5.2	0.000	5.6	0.000	30.4

Moreover, we need to ameliorate this technique even more for some hard QAPLIB instances. Finally, we plan to apply the GBSA algorithm to TSP and compare it with Random-keys Golden Ball algorithm [44]. We plan also to propose a new hybridization such as mixing Golden Ball algorithm with Tabu Search method.

References

[1] S.Sahni, & Gonzalez, T. P-complete approximation problems. Journal of the ACM (JACM), 23(3), 555-565, (1976).

[2] T.C.Koopmans, & M. Beckmann, Assignment problems and the location of economic activities. Econometrica: journal of the Econometric Society, 53-76, (1957).

[3] AN.Elshafei, Hospital layout as a quadratic assignment problem. Operational Research Quarterly, 167-179, (1977).

[4] A. M.Geoffrion, & G. W. Graves, Scheduling parallel production lines with changeover costs: Practical application of a quadratic assignment/LP approach. Operations Research, 24(4), 595-610, (1976).

[5] I. Ugi, J. Bauer, J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum, & W. Schubert, Neue anwendungsgebiete für computer in der chemie. Angewandte Chemie, 91(2), 99-111, (1979).

[6] Z. Drezner, Extensive experiments with hybrid genetic algorithms for the solution of the quadratic assignment problem. Computers & Operations Research, 35(3), 717-736, (2008).

[7] Y. Hani, L. Amodeo, F. Yalaoui, & H. Chen, Ant colony optimization for solving an industrial layout problem. European Journal of Operational Research, 183(2), 633-642, (2007).

[8] L. Y. Tseng, & S. C. Liang, A hybrid metaheuristic for the quadratic assignment problem. Computational Optimization and Applications, 34(1), 85-113, (2006).

[9] V.Černý, Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. Journal of Optimization Theory and Applications, 45(1), 41-51, (1985).

[10] R.E.Burkard, E. Cela, P.M.Pardalos, & L.S.Pitsoulis. The quadratic assignment problem. In Handbook of Combinatorial Optimization (pp. 1713-1809). Springer US, (1998).

[11] V. Kaibel, Polyhedral combinatorics of QAPs with less objects than locations. Univ., (1997).

[12] Z. Drezner, Compounded genetic algorithms for the quadratic assignment problem. Operations Research Letters, 33(5), 475-480, (2005).

[13] M. A. El-Baz, A genetic algorithm for facility layout problems of different manufacturing environments. Computers & Industrial Engineering, 47(2), 233-246, (2004).

[14] W. A. N. G. Rong-Long, & K. Okazaki, Solving facility layout problem using an improved genetic algorithm. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 88(2), 606-610, (2005).

[15] Z. Drezner, The extended concentric tabu for the quadratic assignment problem. European Journal of Operational Research, 160(2), 416-422, (2005).

[16] T. James, C. Rego, & F. Glover, A cooperative parallel tabu search algorithm for the quadratic assignment problem. European Journal of Operational Research, 195(3), 810-826, (2009).

[17] A. Misevicius, A tabu search algorithm for the quadratic assignment problem. Computational Optimization and Applications, 30(1), 95-111, (2005).

[18] H. Q. Saremi, B. Abedin, & A. M. Kermani, Website structure improvement: Quadratic assignment problem approach and ant colony meta-heuristic technique. Applied Mathematics and Computation, 195(1), 285-298, (2008).

[19] H. Liu, A. Abraham, & J. Zhang, A particle swarm approach to quadratic assignment problems. In Soft Computing in Industrial Applications (pp. 213-222). Springer Berlin Heidelberg, (2007).

[20] M. Bashiri, & H. Karimi. Effective heuristics and meta-heuristics for the quadratic assignment problem with tuned parameters and analytical comparisons. Journal of Industrial Engineering International, 8(1), 1-9, (2012).

[21] G. Paul.Comparative performance of tabu search and simulated annealing heuristics for the quadratic assignment problem. Operations Research Letters, 38(6), 577-581, (2010).

[22] E. Osaba, F. Diaz, & E. Onieva, A novel meta-heuristic based on soccer concepts to solve routing problems. In Proceedings of the 15th Annual Conference Companion on Genetic and Evolutionary Computation (pp. 1743-1744). ACM, (2013, July).

[23] E. Osaba, F. Diaz, R. Carballedo, E. Onieva, & A. Perallos, Focusing on the Golden Ball Metaheuristic: An Extended Study on a Wider Set of Problems. The Scientific World Journal, 2014, (2014).

[24] A. M. Frieze , G. Galbiati , and F. Maffioli , On the worst-case performance of some algorithms for the asymmetric traveling salesman problem, Networks, 12 (1), 23–39, (1982).

[25] B. Golden, E. Baker, J. Alfaro, and J. Schaffer, The vehicle routing problem with backhauling: two approaches, in Proceedings of the 21st Annual Meeting of SE TIMS, pp. 90–92, South Carolina, SC, USA, (1985).

[26] E.Osaba, & F. Díaz. Design and Implementation of a Combinatorial Optimization Multi-population Meta-heuristic for Solving Vehicle Routing Problems. International Journal of Interactive Multimedia and Artificial Intelligence, 4(2), 89-90, (2016).

[27] I. Rivin, I. Vardi, and P. Zimmermann, The n-queens problem, The American Mathematical Monthly, 101(7), pp. 629– 639, (1994).

[28] S. Martello, and P. Toth, Knapsack Problems: Algorithms and Computer Implementations, Wiley, New York, NY, USA, (1990).

[29] F. Sayoti, & M. E. Riffi, Golden Ball Algorithm for solving Flow Shop Scheduling Problem. International Journal of Artificial Intelligence and Interactive Multimedia, 4(1), 15-18, (2016).

[30] F. Sayoti, M. E. Riffi, & H. LabaniOptimization of Makespan in Job Shop Scheduling Problem by Golden Ball Algorithm. Indonesian Journal of Electrical Engineering and Computer Science, 4(3), (2016).

[31] A. Misevičius, A modified simulated annealing algorithm for the quadratic assignment problem. Informatica, 14(4), 497-514, (2003).

[32] G. A. Croes, A method for solving traveling-salesman problems. Operations Research, 6(6), 791-812, (1958).

[33] S. Lin, Computer solutions of the traveling salesman problem. Bell System Technical Journal, 44(10), 2245-2269, (1965).

[34] M. Fischetti, Salazar Gonzalez, J. J., & Toth, P.. A branch-and-cut algorithm for the symmetric generalized traveling salesman problem. Operations Research, 45(3), 378-394, (1997).

[35] C. D. Tarantilis, Solving the vehicle routing problem with adaptive memory programming methodology. Computers & Operations Research, 32(9), 2309-2327, (2005).

[36] R.E.Burkard,S. Karisch, & F.Rendl. QAPLIB-A quadratic assignment problem library. European Journal of Operational Research, 55(1), 115-119, (1991).

[37] W. Ellen E., C.Roger D., & F.Mark A. Parallel simulated annealing using speculative computation. IEEE Transactions on Parallel and Distributed Systems, 1991, 2(4), 483-494.

[38] G.A, El-Nasser A., A.M. Mahmoud, & E.-S. M. El-Horbaty. A comparative study of meta-heuristic algorithms for solving quadratic assignment problem. International Journal of Advanced Computer Science and Applications, 5(1), arXiv:1407.4863, (2014).

[39] J.Skorin-Kapov. Tabu search applied to the quadratic assignment problem. ORSA Journal on computing, 2(1), 33-45,(1990).

[40] E.Taillard. Robust taboo search for the quadratic assignment problem. Parallel computing, 17(4-5), 443-455, (1991).

[41] U.Benlic, & J.K.Hao. Memetic search for the quadratic assignment problem. Expert Systems with Applications, 42(1), 584-595, (2015).

[42] U. Benlic, & J.K. Hao. Breakout local search for the quadratic assignment problem. Applied Mathematics and Computation, 219(9), 4800-4815, (2013).

[43] T.James, C. Rego., & F.Glover. A cooperative parallel tabu search algorithm for the quadratic assignment problem. European Journal of Operational Research, 195(3), 810–826, (2009).

[44] F. Sayoti, & M. E. Riffi, Random-Keys Golden Ball Algorithm for Solving Traveling Salesman Problem. International Review on Modelling and Simulations (IREMOS), 8(1), 84-89, (2015).

Mohammed Essaid Riffi

Ms. Mohammed Essaid Riffi is a professor at the University of Chouaib Doukkali, El Jadida, Morocco, currently he is chief of Computer Science department and former coordinator of Mathematical and Computer Science. Its research axis is the metaheuristics.

Fatima Sayoti

Ms. Fatima Sayoti (15-02-1988) received a master's degree of software quality from University of Chouaib Doukkali, El Jadida, Morocco in 2011. She is currently pursuing his Ph.D. degree (Computer Science) at the University of Chouaib Doukkali, Faculty of Sciences, El Jadida, Morocco. She is also a temporary teacher at Polydisciplinary Faculty, El Jadida (Morocco).

Data-Aware Scheduling Strategy for Scientific Workflow Applications in IaaS Cloud Computing

Sid Ahmed Makhlouf*, Belabbas Yagoubi

L.I.O. Laboratory, Department of Computer Science, Faculty of Exact and Applied Sciences, University of Oran1 Ahmed Ben Bella, P.O. Box 1524 El M'Naouer, Oran (Algeria)

Received 1 April 2018 | Accepted 5 July 2018 | Published 20 July 2018

unir LA UNIVERSIDAD EN INTERNET

Abstract

Scientific workflows benefit from the cloud computing paradigm, which offers access to virtual resources provisioned on pay-as-you-go and on-demand basis. Minimizing resources costs to meet user's budget is very important in a cloud environment. Several optimization approaches have been proposed to improve the performance and the cost of data-intensive scientific Workflow Scheduling (DiSWS) in cloud computing. However, in the literature, the majority of the DiSWS approaches focused on the use of heuristic and meta-heuristic as an optimization method. Furthermore, the tasks hierarchy in data-intensive scientific workflows has not been extensively explored in the current literature. Specifically, in this paper, a data-intensive scientific workflow is represented as a hierarchy, which specifies hierarchical relations between workflow tasks, and an approach for data-intensive workflow scheduling applications is proposed. In this approach, first, the datasets and workflow tasks are modeled as a conditional probability matrix (CPM). Second, several data transformation and hierarchical clustering are applied to the CPM structure to determine the minimum number of virtual machines needed for the workflow execution. In this approach, the hierarchical clustering is done with respect to the budget imposed by the user. After data transformation and hierarchical clustering, the amount of data transmitted between clusters can be reduced, which can improve cost and makespan of the workflow by optimizing the use of virtual resources and network bandwidth. The performance and cost are analyzed using an extension of Cloudsim simulation tool and compared with existing multi-objective approaches. The results demonstrate that our approach reduces resources cost with respect to the user budgets.

Keywords

Cloud Computing, Workflow Data Scheduling, Clustering, Data Transformation, Clustering Quality Indexes, CloudSim.

DOI: 10.9781/ijimai.2018.07.002

I. Introduction

In recent years, cloud environments are increasingly used in the scientific field [1]. These environments are currently changing dramatically because of the integration of new technologies such as GPUs, sensors, etc.; thus providing scientists with high computing power, storage, and bandwidth. However, the drawback of this power lies in the heterogeneity of resources that makes its management more complex. Other complexities arise because of the urgent need for scale-up, reduced application response time, fault tolerance and infinite storage space, which pushes scientists to use multiple applications simultaneously, resources at the same time in a Workflow application.

Scientific workflows are used to model computationally intensive and large-scale data analysis applications [2]. In recent years, cloud computing has been evolving rapidly as a target platform for such applications [3]. As a result, several workflow-specific resource management systems have been developed by cloud providers, such as the Amazon Simple Workflow Service (SWF) [4], to enable users to dynamically provision resources.

The workflow-scheduling problem has been studied extensively over past years focusing on multiprocessor system and distributed environments like grids and clusters [5]. Workflow and directed acyclic graph (DAG) are usually interchangeable in the literature. It is a well-known research area where the programming complexity is NP-complete. [6].

The Workflow scheduling approaches can be classified according to different aspects of optimization method such as heuristic, clustering, critical path, fuzzy, greedy, market-driven, meta-heuristic, mathematical modeling, and partitioning. Majority of the Workflow scheduling approaches focus on employing heuristic and meta-heuristic as an optimization method and focusing only on the execution time [7]. However, even in these cases, communication among tasks is assumed to take zero time units. In our approach, we use Clustering scheduling to achieve a better performance regarding effectiveness and accuracy at the cost. Clustering-based scheduling is designed to optimize transmission time between data dependent tasks [8]. DAG Clustering is a mapping of all tasks onto clusters, where each cluster is a subset of Tasks, and each cluster is executed on a separate resource. The basic idea of clustering is to reduce the communication time between tasks.

Traditional techniques have examined the data sharing of workflows tasks. These techniques that investigate the scheduling of scientific workflows tasks have inspired us when developing our approach. However, the tasks hierarchy in scientific workflows has not been explored extensively. Therefore, we consider in this paper the tasks hierarchy for workflow scheduling.

* Corresponding author.

E-mail address: sidahmed.makhlouf@gmail.com

The existing scheduling techniques have been categorized into three parameters including static, dynamic, and static-dynamic. Dynamic scheduling is efficient for a cloud computing environment due to its ability to handle the arriving tasks [7]. Moreover, hierarchical scheduling cooperates static and dynamic scheduling to generate powerful solutions [8].

To address the challenge, we propose a novel approach for workflow scheduling considering the hierarchy of scientific workflows tasks. Significant contributions presented in this paper are summarized as follows:

- Conditional probability for scientific workflows tasks is computed leveraging their representation as a matrix. The conditional probability reflects the possibility that scientific workflow tasks can share or use the same data.
- Determining the exact number of workflow clusters using 13 clustering indexes. In this work, we focus on the following research question: What is the number of virtual machines required for the efficient and transparent execution of a workflow in a cloud environment?
- Do a hierarchical clustering to determine the tasks hierarchy in scientific workflows and to regroup the workflow tasks into clusters. To minimize data transfer between clusters we have measured the distance between each pair of tasks and regroup the closer tasks into the same cluster.
- Extensive experiments are conducted for evaluating the effectiveness and accuracy of our technique. The result shows that our approach reduces the resources cost and total the data transfer time.

The paper is organized as follows. Section II presents related work. Section III describes the problem statement and introduces the proposed solution. In Section IV we introduce our system model and assumptions. Section V contains a description of our proposed approach, and Section VI and VII discuss the evaluation procedure and the results by applying our approach. Finally, Section VIII outlines general conclusions and exposes our future work.

II. Related Works

In the literature, researchers classify task-scheduling strategies into two main classes: job scheduling, that focuses on scheduling a set of independent tasks to be executed sequentially or in parallel, and the workflow that map and manage the execution of interdependent tasks on shared resources. Since the advent of cloud computing, several scheduling techniques for workflow applications have been proposed. These techniques take into account many aspects of the cloud environment. For example, some techniques try to optimize costs, while others try to optimize fault tolerance.

The cost has become an important objective in workflow cloud scheduling research. The total cost incurred by running workflow can include many components such as the cost of computing resources, the cost of storage resources, and the cost of data transfer resources. A budget is often defined as the number of the maximum virtual machines to process workflow. Also in grid computing, several cost-aware scheduling techniques have been introduced [9]. In [10], the authors introduce a strategy called ADAS (Adaptive Data Scheduling) for workflow applications. This work aims to reduce the monetary cost of moving data while a workflow is running. The authors propose adaptive scheduling based on dependencies between tasks. The scheduling process is in two steps; The first one is to create initial clusters for workflow tasks. They use a matrix approach to group tasks in some data centers as initial clusters using the BEA algorithm (Bond Energy Algorithm) [11]. The second step is to group the data/task pairs using the cost-effective scheduling quality. The authors demonstrate that this strategy can improve the processing time of a workflow and the use of resources in a cloud environment. However, this strategy uses only one clustering algorithm (BEA) to create clusters; in our strategy, we use 13 clustering indexes to determine the specific clusters numbers and do a hierarchical clustering to regroups workflow tasks into clusters. In [12], the authors propose a scheduling system for a data-intensive workflow to minimize its processing time. To reduce data transfer time, the proposed system uses task clustering when submitting a workflow application. To do this, it calculates the dependencies of the tasks in the workflow according to the conditional probability of the common number of files for each pair of tasks. This clustering method is validated with the simulator WorkflowSim. We note that the authors count conditional probability of the number of common files for each pair of tasks regardless of file size. So, in this strategy, it is possible to group tasks around small files only, which involves clustering tasks with a low communication rate. In our strategy, we use the conditional probability to determine and count the possible bytes number of common files for each pair of tasks. The priority is to group only those tasks that communicate through large files. Our idea is to avoid unnecessary moving large files, which will consequently speed up the execution time of the workflow.

In addition to the makespan and cost criteria, the reliability of the workflow execution is also taken into account. This criterion ensures that the resources selected in a schedule can probably complete the scheduled tasks. The failure of the task execution is usually handled by checkpointing and replication mechanisms. On the other hand, the existence of multiple replicas introduces a challenge for maintaining the replicas consistency in the cloud environment. To solve the problem of replica coherence in the cloud, in [13], the authors propose a workflow partitioning approach that takes into account data placement by grouping tasks into a cluster based on data replication in cloud environments. The proposed approach improves the placement of data and minimizes response time. According to the authors, this is due to scheduling of tasks in data centers that contain the majority of the replicated data. Replication-based scheduling approach achieves shorter makespans. However, it makes the scheduling problem more difficult. The scheduling algorithm not only needs to observe the precedence constraints between tasks but also needs to recognize which tasks and data to duplicate. In our strategy, we used the conditional probability to determine the possible common set of data between each pair of tasks. However, both mechanisms of checkpointing and replicating the task may respectively cause the waste of time and compute resources [14, 15]. Replication and checkpointing require storage space for each task and for each file in the workflow. And probably the transfer of replicas and checkpoint across the network, which increases the makespan of the workflow. In addition, these replication and checkpointing mechanisms require a reliable central or distributed storage management system. We note that none of the works cited in this paragraph measure, in terms of data storage and transfer, the impact by using the replication and the checkpointing mechanisms.

In [16] authors propose Workflow Partition Resource Clusters algorithm for scientific workflows. In this work, scheduling is in two phases: (i) on the global level the algorithm clusters workflow and generate a set of sub-workflows to achieve high parallelism, and (ii) on the local level sub workflows generated are dispatched to selected resource clusters. The algorithm tries to minimize the cost of workflow execution and the makespan. As in [16], many algorithms in the literature suppose an unlimited number of available virtual resources. In practice, it is not possible for a system to own an unlimited number of virtual resources [17, 18, 19]. In our strategy, a limited resource number is also taken into consideration, and a mechanism is proposed to minimize the resource requirement.

Several studies have used cluster analysis to schedule workflow tasks into clusters. Authors in [20] propose an approach for dynamic resources provisioning and present a cost- and deadline-constrained workflow scheduling algorithm in a cloud environment. The work is divided into two phases: (i) in the first phase, they use the k-means clustering technique for determining the speed of VMs that would be selected in scheduling; (ii) in the second phase, they propose an approach for dynamic provisioning of VMs using a variant of "Subset Sum" problem. The results of the simulation show that the proposed approach achieves better performance with respect to the cost of execution.

K-means is one of the most known cluster analysis algorithms. However, the most important limitations of k-means are: (i) at the beginning of algorithm, the user must specify the number of clusters -k- and choosing the number -k- of clusters can be difficult; (ii) k-means is mainly limited to Euclidean distances. In our strategy, we implement an agglomerative hierarchical clustering to find the best step at each cluster fusion to determine the best cluster number k. Hierarchical clustering does not require a specific distance measure; any measure can be used.

In most studies, data transfer between workflow tasks is not considered, data uploading and downloading are assumed as part of task execution. However, this is not always the case, especially in the big data area. For a data-intensive workflow application, data movement can dominate both the execution time and cost. In [21], cloud storage resources are virtualized like Amazon S3 resources. The S3 Cloud storage is used for data availability and data broadcasting. However, in the design of the data placement strategies for resource provisioning it is important to consider the intermediate data transfer performance explicitly. Workflow tasks communicate through the transfer of intermediate files. So, the choice of locality storage system has a significant impact on workflow performance and cost. In our strategy, we take into account the intermediate data transfer cost between tasks when the VMs (Clusters) are deployed.

Based on this literature review, we find several issues that have not been sufficiently studied. These are the gaps in the reviewed work that will be directions for our works.

In this work, we try to understand the challenges of managing virtual resources when running Scientific Workflow in the Cloud. We will try to minimize the number of resources allocated to a Workflow under budget constraints. The goal was to improve the efficiency of resource provisioning in the cloud to execute the large-scale workflow better. To do this, we have developed tasks and data clustering systems. The clustering is done in relation to the underlying network load and inter-task communication rates. We have experimented our resource allocation strategy using an extension of the CloudSim [22] simulator.

III. Problem Statement

In the distributed execution paradigm, a workflow is divided into small tasks, which are assigned to different data centers for execution.

When a task requires processing of data from different data centers, data movement becomes a challenge. Some data are too large to be moved. In a cloud, data centers are geographically distant, and data movement would add monetary cost to the Workflow execution. Our work aims to reduce the monetary cost of data movements during workflow execution, to improve the use of the network in Cloud environment. The data location problem is one of the important challenges for planning a data-intensive application in Cloud Computing.

In data-intensive scientific workflows, tasks require more than one set of data to be executed. However, when these tasks are executed in different datacenters, data transfer would become inevitable. To resolve these issues, this work proposes a task dependency-based clustering method to optimize scheduling and execution of workflows. This work proposes a data-intensive workflow scheduling system to minimize data movement between data centers. In the next section the components of the proposed system are described.

IV. System Model and Assumptions

In this section, we present the main assumptions of our approach and environment model.

A. Application Model

Scientific workflows are modeled as Directed Acyclic Graph (DAG). A DAG, *G(V, E)*, consists of a set of vertices *V*, and edges, *E*. The edges represent constraints. Each edge represents a precedence constraint that indicates that the task must complete its execution before the next task begins. Each edge also represents the amount of data between tasks involved; for example, Fig. 1 shows the amount of data (in bytes) that the task should send to the next one.

A task is said to be ready if all of its parents have completed their execution and cannot begin execution until all dependencies have been satisfied. If two tasks are scheduled on the same data center, the cost of communication between them is supposed to be zero.

Applications target are Workflows. Tasks Workflow runtime is estimated, and it indicates how long it takes to execute them on a given VM type. As for execution times, we assume that the size of the files can be estimated on the basis of historical data. The files are assumed to be write-once, read-many.

A task has zero or more input files that must be fully transferred to a virtual machine before execution can begin; And has zero or more output files which can be used as inputs for another task or as the final result of a workflow. We suppose that file names are unique in a workflow.

B. Execution Model

A cloud consists of an unlimited number of virtual machines (VMs) that can be provisioned and de-provisioned on demand. A virtual machine can execute only one task at a time. A virtual machine is charged 1 $ for each interval of 60 minutes (one hour) of operation. Partial usage of a billing interval is rounded.

C. Data Transfer Model

We use a global storage model to transfer input and output files between tasks and to store the results of the Workflow. Each VM has a local cache of files. This method of transfer is widely used in cloud environments by using shared distributed file systems such as NFS [23].

To transfer a file between virtual machines, a request must be sent to the Global Storage Management System (GSMS). The performance of transferring a file to its destination depends on the dynamic state of the underlying network. The state of the network depends on the number of files being transferred, the size of the files being transferred and the presence of the files in the local cache of the VMs.

We assume that transfers between tasks always require the downloading of whole files. We also consider the usage cost of global storage. This means that the amount of data stored and transferred affects the cost of the running Workflow. In our congestion model, to transfer files to multiple tasks, starting multiple simultaneous transfers could delay the execution of these transfers. Parallel transfers do not offer benefit for a task. The bandwidth of a network link to a virtual machine is a limiting factor.

A file cannot be transferred faster than a given maximum bandwidth. Each request to the storage system is managed with latency measured in milliseconds. We assume that we can calculate precisely the time required to transfer a file. A virtual machine has a local disk that serves as a file cache, and the disc price is included in the VM cost. The files are cached in the First-In-First-Out (FIFO) policy.

V. Proposed Approach

Tasks clustering is a technique that consolidates fine-grained tasks into coarse tasks. Task clustering has proven to be an effective method of reducing overhead costs and improving the performance of scientific workflow tasks. With task clustering, Workflow execution overhead can be eliminated by grouping small tasks into a single job unit. This Workflow reduction benefits the entire cloud environment by reducing traffic between sites. For the efficient execution of a Workflow on a cloud environment, we propose in this work, a policy of allocation and management of virtual resources. This policy is based on the calculation of the dependency between workflow tasks. The dependency is based on the conditional probability of workflow tasks.

A. Task Clustering Based on Conditional Probability

The proposed system uses tasks clustering in the Workflow to reduce data transfer time. As an illustration, we use the example of the Scientific Workflow in Fig. 1. In this Workflow, there are seven tasks. In scientific workflows, tasks communicate data by sending and receiving intermediate files. In our workflow model, we assume that the edges are files. So, we have a set of tasks *T* and set of files *F*.

In our model, first, we compute the task dependencies in the workflow based on the conditional probability. Let $x_i \in T$ and $y_i \in T$ two tasks of the same Workflow *G*. It is assumed that x_i has non-zero probability. The conditional probability of a task y_i, knowing that another task x_i has finished, is the number noted $P(x_i \mid y_i)$ and defined by:

$$P(x_i \mid y_j) = \frac{P(x_i \cap y_j)}{P(x_i)}, P(x_i) > 0 \quad (1)$$

The real $P(x_i|y_i)$ is read "probability of y_i, knowing x_i" according to the common use of the sets of data between x_i and y_i.

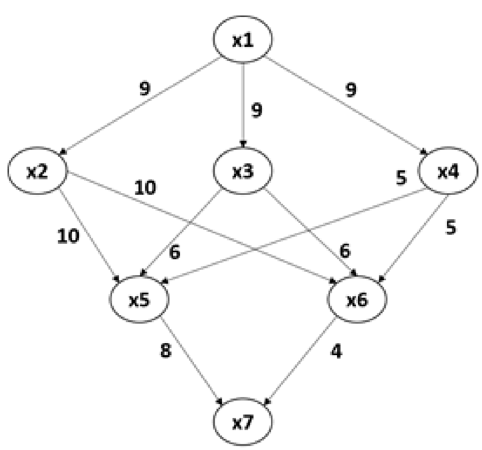


Fig. 1. Example of workflow instance.

The conditional probability imposes the creation of the conditional probability matrix $CPM[|T|][|T|]$ based on contingency table $CT[|T|+1][|T|+1]$ and the joint and marginal probability table $JMP[|T|+1][|T|+1]$ of each task pair in the workflow. First, we create contingency table as expressed in (2).

$$CT[i][j] = DataSize\left(\left\{O_{P_{i,j}}\right\} \cap \{I_i\} \cap \{I_j\}\right); 1 \le i,j \le |T| \quad (2)$$

$P_{i,j}$ is the set of common parents between each tasks x_i, y_i:

$$P_{i,j} = \left\{\{P_i\} \cap \{P_j\}\right\} \quad (3)$$

P_i and P_j are respectively the parents set of the tasks x_i, y_i. The dependency is calculated by measuring the total size of all the output files of the set $P_{i,j}$. $O_{P_{i,j}} \subset F$ is the set of the output files of $P_{i,j}$. $I_i \in F$ and $I_j \in F$ are respectively inputs files set of task $x_i \in T$ and $y_j \in T$. Each value in the contingency table is the common size of the input files for each task pair in the workflow:

$$CT[i][|T|+1] = \sum_{j=1}^{|T|} CT[i][j]; 1 \le i \le |T| \quad (4)$$

$$CT[|T|+1][j] = \sum_{i=1}^{|T|} CT[i][j]; 1 \le j \le |T| \quad (5)$$

$$CT[|T|+1][|T|+1] = \sum_{i=1}^{|T|} CT[i][|T|+1] \quad (6)$$

The joint and marginal probability table can be created from the contingency table. The joint probability of T_i and T_j is:

$$JMP[i][j] = \frac{CT[i][j]}{CT[|T|+1][|T|+1]}; 1 \le i,j \le |T| \quad (7)$$

The marginal probability of T_i is:

$$JMP[i][|T|+1] = JMP[|T|+1][i] = \frac{CM[i][|T|+1]}{CM[|T|+1][|T|+1]} \quad (8)$$

From Joint and Marginal Probability Table, the conditional probability for each task pair can be calculated as follows:

$$CPM[i][j] = \frac{JMP[i][j]}{JMP[i][|T|+1]}; 1 \le i \le |T|; 1 \le j \le |T| \quad (9)$$

To determine the number of clusters of the workflow, we apply our clustering approach on the *CPM* matrix. However, before applying any clustering method, we must first preprocess the *CPM* matrix data. Data pre-processing has to do with the steps that are required to transform the data we have in a way that allows applying further analysis algorithms (e.g., Clustering techniques).

B. Transforming Data

The original *CPM* data matrix needs to go through some modification to make it more useful for our analysis. Numeric variables sometimes have slightly different scales. This can create problems for some data analysis tools [24].

First, we identify Skewness distributions of *CPM* data matrix. Skewness is a measure of shape distribution. Negative skewness indicates that the mean of the data values is less than the median, so the data distribution is skewed to the left. Positive skewness indicates that the mean of the data values is larger than the median, so the data distribution is skewed to the right.

- If the skewness is equal to 0, then the data is symmetrical and did not need to be transformed.
- If the skewness is -1/2 and +1/2, the data is approximately

symmetric, and need to be transformed.

- If the skewness is less than -1 or greater than +1, the data is highly skewed, and need to be transformed.
- If the skewness is between -1 and -1/2 or between +1 and +1/2 then the data is moderately skewed and need to be transformed.

For example, skewness value of the workflow of Fig. 1 is 2.34, so we must do data transformation. Transforming data is used to coerce different variables to have similar distributions.

Because some measurements in nature are naturally distributed following a normal distribution, it is important to develop an approach to transforming one or more variables into a normal distribution. Other measures can be naturally logarithmic. In the literature, common data transformations include the transformation of the square root, the cube root, and the logarithm.

In data analysis, the transformation is the replacement of a variable with a function of that variable; for example, replacing a variable x with the square root of x $\left(\sqrt{x}\right)$ or the logarithm of x ($Log_b x$). A transformation changes the shape of distribution or relationship between variables.

- The logarithm transformation transforms x to $Log_{10}x$, or x to $Log_e x$, or x to $Log_2 x$. It is a strong transformation with an effect on the shape distribution. It is used to reduce the right skewness. It can not be applied to null or negative values. So we cannot apply the logarithm transformation for the *CPM* matrix data.
- The cube root transforms to $x^{\frac{1}{3}}$. It is a weaker transformation than the logarithm but with a substantial effect on the shape of the distribution. It is also used to reduce right skewness and has the advantage of being able to be applied to null and negative values.
- The square root transforms x to $x^{\frac{1}{2}}$, is a weaker transformation than the logarithm and cube root with a moderate effect on the distribution shape. It is used to reduce the right skewness, and also has the advantage of being able to be applied to null values.

The Table I shows skewness value of square and cube root transformation. We can see that the cube root transformation approximates the normal distribution. So the cube root transformation is more powerful than the square root transformation, and we will adopt it in our experiments in section VI.

TABLE I. Different Skewness Values of Fig. 1

Original skewness	square root	cube root
2.340348	1.899396	1.822186

After applying our clustering approach on the *CPM* matrix we obtain the results of Fig. 2. In the following sections, we will discuss in detail our clustering approach used for this example.

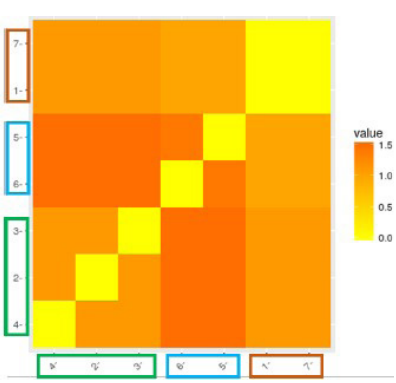


Fig. 2. Example of workflow clustering.

After the creation of the*CPM* matrix, we will fragment the workflow into clusters of tasks. Each cluster will be assigned to a virtual machine, so the number of virtual machines created depends on the number of clusters. The goal of this work is to maximize the data transfer in a cluster and minimize it between clusters. The question that arises then is: What is the minimum number of clusters needed to execute the workflow to meet user budget correctly? So, in our work, the workflow execution must be efficient, meet the workflow budget and at a minimal cost. To answer this question, we have implemented some algorithms and technics for determining the optimal number of clusters in *CPM* dataset and offering the best workflow clustering scheme from different results.

In [25], the authors identify 30 clustering quality indexes that determine the optimal number of clusters in a dataset and offers the best clustering scheme from different results to the user. We tried to apply these 30 algorithms to our *CPM* matrix, and we found that only 13 algorithms are compatible with the *CPM* data matrix. For the rest of the 17 algorithms, the clustering result tends to infinity.

The evaluation of the algorithms on the *CPM* matrix must deal with problems such as the quality of the clusters, the quality of the data compared to the clustering quality indexes, the degree which a clustering quality indexes fit with the data of the *CPM* matrix and the optimal number of clusters [25]. As a result, we used the following clustering quality indexes:

1. Krzanowski and Lai 1988 [26]
2. Calin'ski and Harabasz 1974 [27]
3. Hartigan 1975 [28]
4. McClain and Rao 1975 [29]
5. Baker and Hubert 1975 [30]
6. Rohlf 1974 [31] and Milligan 1981 [32]
7. Dunn 1974 [33]
8. Halkidi et al. 2000 [34]
9. Halkidi and Vazirgiannis 2001 [35]
10. Hubert and Levin 1976 [36]
11. Rousseeuw 1987 [37]
12. Ball and Hall 1965 [38]
13. Milligan 1980, 1981 [39, 32]

C. Tasks Distance Measures

The data set is represented by the *CPM* matrix. Each element in the *CPM* matrix represents the distance between two tasks in the workflow. So, the clustering is done comparing the distance between each pair of tasks of the matrix *CPM*. To measure the distance, we used the following metrics:

Euclidean distance is the distance between two tasks x and y in a R^n space and is given by (10).

$$d = \sqrt{\sum_{i=1}^{n}\left(x_i - y_i\right)^2}\,; x_i, y_i \in CPM\left[1,|T|\right]\left[1,|T|\right] \tag{10}$$

It is the length of the diagonal segment connecting x to y.

Manhattan distance is the absolute distance between tasks x and y in R^n space and is given by (11).

$$d = \sum_{i=1}^{n}\left|x_i - y_i\right|; x_i, y_i \in CPM\left[1,|T|\right]\left[1,|T|\right] \tag{11}$$

As opposed to the diagonal distance in the Euclidean distance, the distance between two tasks in a grid is based on a strictly horizontal/vertical path. The Manhattan distance is the sum of the horizontal and vertical components.

Minkowski distance: is the p^{th} root of the sum of the p^{th} powers of the differences between the components. For two tasks x and y in R^n space, Minkowski distance is given by (12).

$$d = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{\frac{1}{p}} ; x_i, y_i \in CPM\left[\overline{1,|T|}\right]\left[\overline{1,|T|}\right] \quad (12)$$

Minkowski distance can be considered as a generalization of both the Euclidean and the Manhattan distance [40]. $p=1$ corresponds to the Manhattan distance and $p=2$, to the Euclidean distance. For p reaching infinity, we obtain the **Chebyshev** distance.

D. Cluster Analysis Method and VMs Number Interval

As mentioned above, we used 13 partitioning algorithms to determine the exact number of clusters needed. In addition to the *CPM* matrix and the distance metric, these algorithms require other input parameters, namely the cluster analysis method, the minimum and maximum cluster interval. In our works, we have used the following cluster analysis methods:

Single: The distance D_{ij} between two clusters C_i and C_j is the minimum distance between two tasks x and y, with $x \in C_i, y \in C_j$.

$$D_{ij} = \min_{x \in C_i, y \in C_j} d(x, y) \quad (13)$$

Complete: The distance D_{ij} between two clusters C_i and C_j is the maximum distance between two tasks x and y, with $x \in C_i, y \in C_j$.

$$D_{ij} = \max_{x \in C_i, y \in C_j} d(x, y) \quad (14)$$

We have set the cluster interval calculation between 2 and 20 clusters. This interval choice is based on the Amazon EC2 resource allocation policy [17, 18, 19]. Amazon EC2 allows the possibility of reserving only 20 virtual machines. This is why we cannot create more than 20 clusters. So, we have fixed the user budget to 20 virtual machines.

Algorithm 1 models the first step of our approach, determining the cluster number. We start by creating the *CPM* matrix in line 2. Then the matrix created will be one of the input parameters to the 13 clustering quality indexes cited above. In line 6 each algorithm will calculate the possible cluster number for the *CPM* matrix. At the end of algorithm 1, we calculate the average number of the clusters from all the proposed numbers obtained from the 13 clustering quality indexes in line 8.

Algorithm 1: Determining the cluster number

```
Data: T, F
/* Tasks and files set                                          */
Result: avgClustersNb
/* Return average number of clusters                            */
1 foreach x_i, y_j ∈ T; i, j = 1,|T| do
2 |  CPM[i][j] ← P(x_i|y_j);
3 end
4 TCNb ← 0;
/* Total Clusters number                                        */
5 foreach cm = 1,13 do
  |  /* cm : Clustering Method index                             */
6 |  TCNb ←
  |  TCNb + ClusteringMethodd_cm(CPM, DistMe, minCNb, maxCNb);
  |  /* DistMe: distance measures(Euclidean/Manhattan).
  |     minCNb,maxCNb: minimum and maximum cluster number
  |     */
7 end
8 avgClustersNb ← TCNb/13;
```

E. Hierarchical Clustering Method

The average obtained will be used to do a hierarchical clustering of the *CPM* matrix. Hierarchical clustering is one of the domains of automatic data analysis and data classification. Strategies for hierarchical clustering are generally divided into two types:

Agglomerative: This is a "bottom-up" approach; this method starts from a situation where all tasks are alone in a separate cluster, then pairs of clusters are successively agglomerated until all clusters have been merged into one single cluster that contains all tasks.

Divisive: This is a "top-down" approach, in which all tasks are in a single cluster; we divide this cluster into two sub-clusters which are, in turn, divided into two sub-clusters and so on. At each step, each cluster is divided into two new clusters.

In our work, we have used agglomerative clustering of a set of tasks T of n individuals. Our goal is to distribute these tasks in a certain number of clusters, where each cluster represents a virtual machine.

The agglomerative hierarchical clustering assumes that there is a measure of dissimilarity between tasks; in our case, we use *CPM* matrix as a measure for dissimilarity calculation. The dissimilarity between tasks x and y will be noted $dissim_{cpm}(x, y)$.

The agglomerative hierarchical clustering produces a hierarchy H of tasks. H it is the set of clusters at all the steps of the clustering approach and checks the following properties:

1. $T \in H$: at the top of the hierarchy, when grouping clusters to obtain a single cluster, all tasks are grouped;
2. $\forall x \in T$, $\{x\} \in H$: at the bottom of the hierarchy, all tasks are alone;
3. $\forall (h,h') \in H^2$, $h \cap h' = \emptyset$ or $h \subset h'$ or $h' \subset h$

Initially, in our approach, each task forms a cluster. We try to reduce the number of clusters to the average calculated previously by the 13 clustering quality indexes; this is done iteratively. At each iteration, two clusters are merged, which involves reducing the number of total clusters.

The two clusters chosen to be merged are the most "similar", in other words, those whose dissimilarity is minimal (or maximal). This dissimilarity value is called aggregation index. Since we first merge the closest tasks, the first iteration has a low aggregation index, but it will increase from iteration to another iteration.

For agglomerative clustering, and to decide which clusters should be merged; a measure of dissimilarity between sets of clusters is required. This is achieved by measuring a distance between pairs of clusters discussed in section V.D.

The dissimilarity of two clusters $C_i = \{x\}, C_j = \{y\}; 1 \le i, j \le n$, each containing one task, is defined by the dissimilarity between its tasks $dissim(C_i, C_j) = dissim(x, y); 1 \le i, j \le n$.

When clusters have several tasks, there are multiple criteria for calculating dissimilarity. We used the following criteria:

Single link: the minimum distance between tasks of C_i and C_j:

$$dissim(C_i, C_j) = \min_{x \in C_i, y \in C_j} \left(dissim_{cpm}(x, y) \right); 1 \le i, j \le |T| \quad (15)$$

We have used this method to minimize data transfer between clusters and maximize the data transfer inside a cluster.

Complete link: the maximum distance between tasks of C_i and C_j:

$$dissim(C_i, C_j) = \max_{x \in C_i, y \in C_j} \left(dissim_{cpm}(x, y) \right); 1 \le i, j \le |T| \quad (16)$$

Unlike the previous method, this one increases data transfer between clusters. We suppose that this method will not give good results, but we have used it to prove the performance of our workflow clustering policy.

Algorithm 2 models the agglomerative hierarchical clustering of the *CPM* matrix. It receives as input parameters the *CPM* matrix and the cluster number that we want to create. The cluster number is the average of all cluster numbers obtained from the 13 clustering quality indexes. The result of hierarchical clustering is presented in a dendrogram. Fig. 3 is the result of clustering, a LIGO workflow of 50 tasks into 11 clusters using the Euclidean distance measurement and the Single Link agglomeration method.

Algorithm 2: Clusters creation

```
Data: T, CPM, avgClustersNb
/* Tasks set, CPM matrix and clusters number                          */
Result: ClustersList
/* Return clusters list                                                */
ClustersList ← φ;
foreach x_i ∈ T; i = 1,|T| do
    Cluster ← {x_i};
    ClustersList ← ClustersList ∪ Cluster;
end
while |ClustersList| > avgClustersNb do
    foreach i = 1,|ClustersList| do
        foreach j = 1,|ClustersList| do
            DissimMat [i] [j] ←
            dissim_cpm (ClustersList.get (i) , ClustersList.get (j));
        end
    end
    {Cluster′, Cluster″} ← get2SimularClusters (DissimMat);
    NewCluster ← Cluster′ ∪ Cluster″;
    ClustersList ← ClustersList − {Cluster′, Cluster″};
    ClustersList ← ClustersList ∪ NewCluster;
end
```

F. Objective Function

We have introduced an objective function that allows us to know the granularity of our resource allocation system. The objective function depends on the workflow resources allocation cost. When a task x_i is assigned to a VM V_k with price p_k, we refer to this as a Resources Allocation RA_i. For each RA_j its cost value is computed :

$$Cost\left(RA_i\right) = \cos t\left(x_i, v_k\right) = \omega_{i,k} \cdot p_k \tag{17}$$

Where p_k is the monetary cost per hour to execute a task x_i on the VM V_k. The execution takes $\omega_{i,k}$ time units. The execution time includes the transfer time of the input tasks data set to the virtual machine from task y_j to the task x_i.

$$\omega_{i,k} = E_{i,k} + \sum_{l=0}^{\left|I_i \cap O_j\right|} T\left(f_l\right) \tag{18}$$

$E_{i,k}$ is the execution time of task x_i on the VM v_k, I_i is the input set of files of the task x_i and O_j is the output set of files of the task y_j. $T_i^j\left(f_l\right)$ is the transfer time of file f_l of the task x_i to the VM v_k from the task y_j. For $x_i \in C_i$ and $y_j \in C_j$, $T_i^j\left(f_l\right)$ is defined as :

$$T_i^j\left(f_l\right) = \begin{cases} \frac{Size\left(f_l\right)}{Bw}; & i \neq j \\ 0; & Otherwise \end{cases} \tag{19}$$

Bw is the bandwidth between cluster C_i and cluster C_j. In this work, we aim to minimize the transfer time of the data set between the virtual machines. To do this, we try to reduce the amount of data set transferred between VM by clustering the highly connected tasks in the same cluster. By this way, we will reduce the task execution time and resource allocation costs.

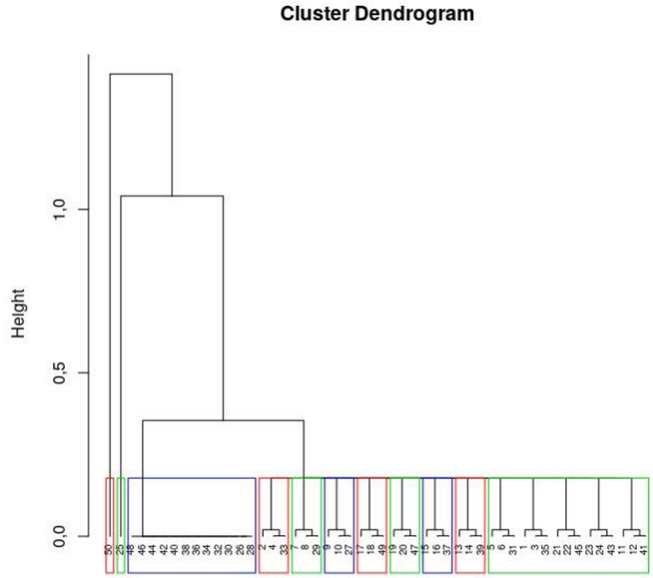


Fig. 3. Cluster dendrogram of LIGO workflow.

$$\begin{aligned} & minimize\left(DTT\left(x_i\right) = \sum_{l=0}^{\left|I_i \cap O_j\right|} T_i^j\left(f_l\right)\right) \\ & \Rightarrow minimize\left(\omega_{i,k}\right) \\ & \Rightarrow minimize\left(Cost\left(RA_i\right)\right) \end{aligned} \tag{20}$$

$DTT(x_i)$ is the data transfer time of all inputs files of the task x_i.

VI. Evaluation Methods

To validate the proposed approach, we have implemented our system in a discrete event simulator "Cloud Workflow Simulator" (CWS) [21]. The CWS is an extension of the "CloudSim" [22] simulator, has a general architecture of IaaS Cloud and supports all of the assumptions stated in the problem described in Section IV. We simulated workflow scheduling with various parameters.

We evaluated our algorithm using synthetic workflows from the Workflow Gallery [41]. We have selected workflows representing several different classes of applications.

The selected applications include LIGO [42] (Laser Interferometer Gravitational-Wave Observatory), a data-intensive application, it is a network of gravitational-wave detectors, with observatories in Livingston, LA and Hanford, WA, and MONTAGE [43], an I/O-bound workflow used by astronomers to generate mosaics of the sky. A summary of workflows used in this work and their characteristics is presented in Table II. In our work, we simulated workflows whose size does not exceed 200 tasks. Because, according to our simulations, the execution of the workflows whose size is greater than 300 tasks will exceed our budget, which is fixed to 20 virtual machines.

TABLE II. Simulated Workflows Characteristics

Size/ Type	Total Data Read (Gb)		Total Data Read (Gb)		Total Data Read (Gb)	
	MONTAGE	LIGO	MONTAGE	LIGO	MONTAGE	LIGO
50	0,68	1,43	0,24	0,02	0,92	1,49
100	1,39	2,86	0,43	0,04	1,82	2,9
200	2,83	5,57	0,79	0,79	3,61	5,64

The experiments model cloud environments with an infinitely NFS-like file system storage.

Table III shows the skewness of each *CPM* workflow, after and before data transformation. We note that the cube root transformation reduces the skewness of the two workflows, especially on the LIGO workflow where it is significantly closer to the normal distribution.

TABLE III. Normalized Conditional Probability Matrix

Size/ Type	Original skewness		Square root		Cube root	
	MONTAGE	LIGO	MONTAGE	LIGO	MONTAGE	LIGO
50	10,34	15,15	3,01	0,94	1,88	0,089
100	15,33	30,88	4,00	1,42	2,72	0,092
200	22,29	33,85	5,43	2,44	3,84	0,500

We have compared our approach with the following algorithms: (i) Static Provisioning Static Scheduling (SPSS) [21]. SPSS is a static algorithm that creates provision and schedules before running workflow. The algorithm analyzes if the workflow can be completed within the cost and deadline. The workflow is scheduled if it meets the cost and deadline constraint. For a workflow, a new plan is created, and if the cost of the plan is less than the budget, the plan is accepted. The workflows are scheduled in the VM which minimizes the cost. If such VM is not available, a new VM instance is created. In this algorithm, file transfers take zero time; (ii) Storage-Aware Static Provisioning Static Scheduling (SA-SPSS) [44], it is a modified version of the original SPSS algorithm to operate in environments where file transfers take non-zero time. It handles file transfers between tasks. SA-SPSS dynamically calculates bandwidth and supports a replicas reconfiguration number.

We chose these algorithms for the following reasons: (i) These two algorithms are multi-objective. They aim to solve at both the cost and the deadline which makes its scheduling decision more complicated. Except that the SPSS supposes that the transfer time of the files is null whereas the SA-SPSS supposes that the transfer time is not null. (ii) In SPSS/SA-SPSS the user sets the cost and the deadline explicitly. On the other hand, our approach is mono-objective, and it is limited to reduce only the workflow cost. Our approach is iterative, and we suppose that its scheduling decision is not complicated. In our approach, the optimal cost to run the Workflow is calculated, and we suppose to do a right tasks clustering around the same files, to reduce the data transfer time and thus reduce the workflow execution time. By comparing our work with SPSS/SA-SPSS, we want to show that a clustering algorithm is as efficient as a multi-objective algorithm. Also, these algorithms are already programmed in the CWS simulator, so we added our approach in this simulator, and we compared it with these algorithms. This way of working ensures that we have a validated simulation environment because these algorithms and the simulator itself are already validated through publications. So, we have simulated the following algorithms:

- Static Provisioning Static Scheduling (SPSS)
- Storage-Aware Static Provisioning Static Scheduling (SA-SASS)
- Data-Aware Euclidean Complete Clustering (DA-ECC)
- Data-Aware Euclidean Single Clustering (DA-ESC)
- Data-Aware Manhattan Complete Clustering (DA-MCC)
- Data-Aware Manhattan Single Clustering (DA-MSC)

To analyze the results relating to experimentation of our approach, we measured the following metrics:

Resources costs is the number of allocated VMs to the workflow

$$wCost = \sum_{i=1}^{|T|} Cost(RA_i).$$

Total Data Transfer Time: $TDTT = \sum_{i=1}^{|T|} DTT(x_i)$; the total size of all transferred files between virtual machines.

The standard deviation of the transferred data:

$$\sigma = \sqrt{\frac{\sum_{k=1}^{|V|} \left(DT(v_k) - avgDT\right)^2}{|V|}};$$

with V is the set of VM allocated to a workflow, DT is the amount of data transferred to the VM v_k, and $avgDT$ is the average value of the data transferred to all VMs allocated to the workflow. If we get a small standard deviation, the values of transferred data to each VM are closed to the average of the data transferred to all the VMs. A large standard deviation means that the values of transferred data to each VM are farther away from the average of the data transferred to all de VMs. Our goal is to get a small standard deviation.

These metrics are used to evaluate the proposed approach compared to the SPSS and SA-SPSS approach. To do this, we simulated the execution of the synthetic workflows Montage and Ligo. We varied the size of simulated workflows between 50 and 200. In our work, we simulated workflows whose size does not exceed 200 tasks. Because, according to our simulations, the execution of the workflows whose size is greater than 300 tasks will exceed our budget, which is fixed to 20 virtual machines. This limit choice is based on the Amazon EC2 resource allocation policy [17, 18, 19]. Amazon EC2 allows the possibility of creating only 20 virtual machines. For each experiment, we measured the metrics cited above. Our objective is to study the impact of the workflow type on the metrics cited above.

VII. Performance Evaluation And Results

A. Experiment 1: Impact of the Workflow Type on the Cost

From Fig. 4, we simulated the execution of the Montage workflows and measured the execution costs in VMs number. We note that regardless of the size of the workflow, our policies give good results by reducing the number of VMs. Especially for large workflows whose size is 200 tasks; we note that the DA-MCC and DA-MCS policies use only 08 virtual machines. This result depends on the *CPM* matrix data and proves that there is not a better distance measure. The distance used depends on the data to be analyzed.

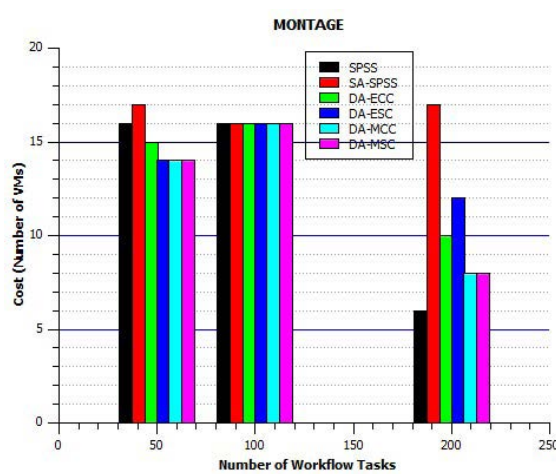


Fig. 4. Impact of the MONTAGE workflow on the cost.

From Fig. 5, we simulated the execution of LIGO workflows and measured the execution costs. We note that regardless of the size of the workflow, our policies give good results. We note that our policies allocate between 17 and 20 virtual machines. In particular, the policies DA-ECC and DA-ESC allocate 19 machines for the execution of the workflows whose size is 200 tasks. Those policies use "single" agglomeration method coupled with the Euclidian/Complete metric to measure the distances and distributes the tasks between virtual machines so that the distance between the VMs is as minimal as possible. This will naturally involve grouping highly dependent tasks into the same virtual machines (cluster) and therefore reducing file transfer between machines to a minimum.

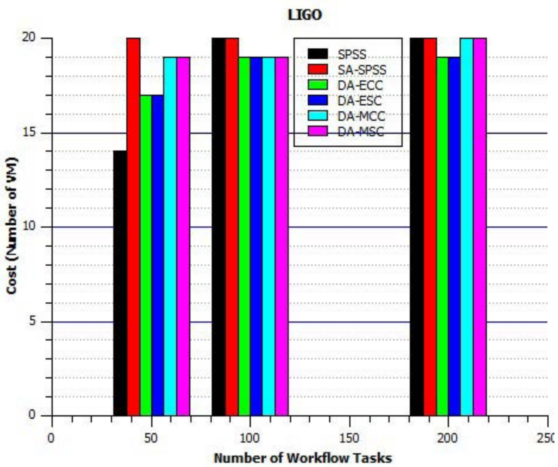


Fig. 5. Impact of the LIGO workflow on the cost.

By comparing Fig. 4 and 5, we note that the Montage workflows allocate fewer resources compared to the LIGO workflow. This confirms the information in Table II: LIGO is data-aware workflows, and Montage is processing-aware workflows. Therefore, the application of a scheduling algorithm depends on the type of the workflow.

In [45] we have several types of scientific workflows, namely, data-aware workflows, processing-aware workflows, memory-aware workflows, etc. Through the two graphs, we note that the SPSS policy gives in some cases good results. These results do not reflect reality because this policy does not support the data transfer time. Hence the importance of using a scheduling algorithm that is specific to the type of workflow [46, 47].

B. Experiment 2: Impact of the Workflow Type on the Total Data Transfer Time

From Fig. 6, we simulated the execution of the Montage workflows and measured the total data transfer time (TDTT). We note that regardless of the size of the workflow, our policies give good results. For example, for large workflows whose size is 200 tasks, we note that the DA-MCC policy completes the total data transfer of the workflow in 456 seconds. This result reinforces our supposition of previous experience according to which cost and TDTT depend on the data we are analyzing, namely the *CPM* matrix. Therefore, the distance used affects the execution time and depends on the data to be analyzed.

From Fig. 7, we simulated the execution of LIGO workflows and measured the total data transfer time. We note that regardless of the size of the workflow, our policies work well. We note that our policies give good results. In particular, the DA-MCC policy that terminates the total data transfer of workflows whose size is 200 tasks at 201 seconds. This result reinforces our previous supposition in which the choice of an agglomeration method has a direct impact on the workflow scheduling. In this case, the "Complete" agglomeration method gives good results. In addition to the previous section, the choice of an agglomeration method also depends on the analyzed data, namely the *CPM* matrix.

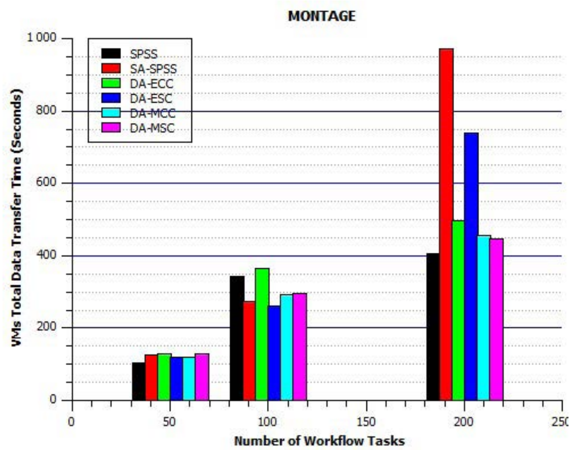


Fig. 6. Impact of the MONTAGE workflow on the total data transfer time.

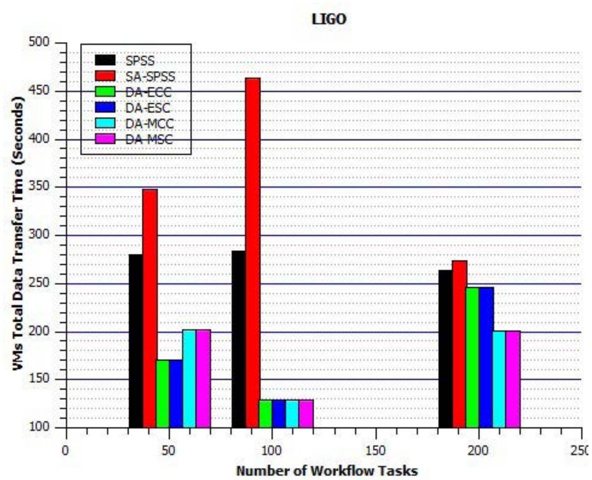


Fig. 7. Impact of the LIGO workflow on the total data transfer time.

From Fig. 6 and 7, we note that the Montage workflows has larger TDTT than the LIGO workflows. This has a relationship with the result of the previous simulation (Experiment 1), in which we noticed that the Montage workflow allocates fewer resources; unlike the LIGO workflow that allocates more resources which implies faster execution. For example, for the Montage workflow of 200 tasks, with the policy DA-MCC, it allocates 8 virtual machines and takes 456 seconds of TDTT. For the LIGO workflow of 200 tasks, with the same policy, it allocates 20 virtual machines and takes 201 seconds of TDTT.

Through Fig. 6 and 7, we note that the SA-SPSS policy gives in most cases bad results compared to our policies. The SA-SPSS workflow tasks scheduling is based on a network congestion subsystem that allows prediction of file transfer times.

The predicted duration time will be included in the overall task time. However, this subsystem does not take into consideration the dynamic and unpredictable nature of the underlying network.

C. Experiment 3: Impact of the Workflow Type on the Standard Deviation

From Fig. 8 and 9, we simulated the execution of the MONTAGE and LIGO workflows respectively and measured the standard deviation of the data transferred to the virtual machines. We note that, regardless of the size of the workflow, our policies give bad results. Unlike the SA-SPSS policy which gives excellent results by keeping a stable standard deviation and this whatever the size of the workflow. We note that the scheduling plan established by our policies is based on the *CPM* matrix in which we store information about data dependencies between tasks.

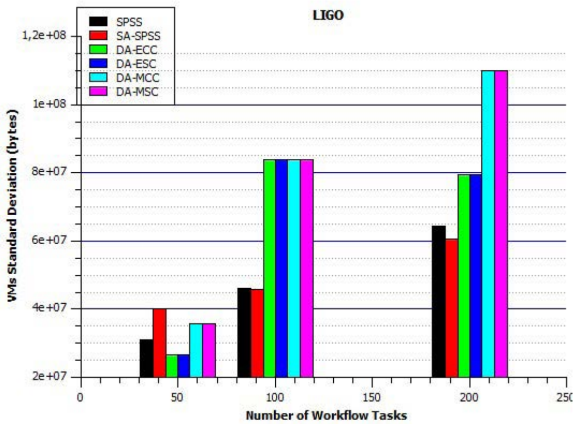


Fig. 8. Impact of the MONTAGE workflow on the standard deviation.

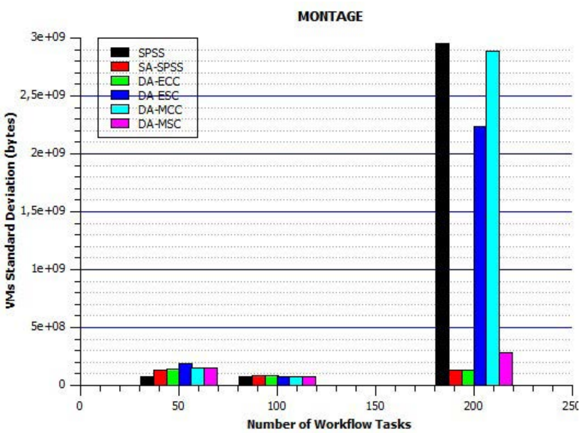


Fig. 9. Impact of the LIGO workflow on the standard deviation.

Unlike SA-SPSS policy that uses a subsystem allowing it to make dynamic scheduling with respect to network congestion.

Comparing Fig. 8 and 9 together, we note that the LIGO workflow generates a reduced standard deviation compared to the Montage workflow. This is because the Montage is processing-aware workflow and the LIGO is data-aware workflow. Through the two graphs, we also note that our policies give good results on LIGO workflow, which proves that our policies are better suited for data-driven workflows.

VIII. Conclusion

Cloud computing has gained popularity for the deployment and execution of workflow applications. Often, tasks and data workflow applications are distributed across cloud data centers on a global scale. So, workflow tasks need to be scheduled based on the data layout in the cloud. Since the resources obtained in the cloud are not free of charge, any proposed scheduling policy must respect the budget of the workflow.

In this paper, an approach was proposed to reduce the virtual machines costs in Cloud Computing. The objective of this work is to provide a scheduling strategy with low costs in cloud computing environment. In our strategy, the amount of global data movement can be reduced, which can decrease the inter-VMs communications rate and improve, therefore, the workflow makespan and network devices in the cloud. Our strategy model was built, based on the principles of communication efficiency-aware scheduling.

In this work a clustering approach was proposed, that could improve the use of resource efficiency and decrease virtual resources consumption during the workflow scheduling. Experiment results demonstrated that the proposed scheduling method could simultaneously decrease virtual resources consumption and workflow makespan.

However, some of our policies have given us unexpected results, such as policies based on the "complete" agglomeration method. We then did extensive research and found that these results have a relationship with the data to be analyzed [48]; in our case, it is the structure of the *CPM* matrix. In [49] we found that there are several types of data, namely, Interval-Scaled data, Dissimilarities, Similarities, Binary data, Nominal, Ordinal, Ratio data, Mixed data. Typically, before applying a distance measure or an agglomeration method, we first need to understand the data type of the *CPM* matrix. As future work, we will explore the field of data mining and classification to understand and define the data type in the *CPM* matrix and apply the right distance measure and the right agglomeration method.

In section V.B, we found that 17 clustering quality indexes do not match the *CPM* matrix data. As a perspective, we will try to understand the reasons why these indexes tend to infinity, and if possible to find a solution to standardize or normalize the data of the *CPM* matrices.

Conditional probability is one of the disciplines of probability theory. In this work, we automated the scheduling of a workflow by modeling the relationships between the tasks of a workflow with the concept of the conditional probability. As future work, we will implement our approach in machine learning. Machine learning is one of the domains of artificial intelligence which is based on statistics. This discipline is strong about modeling NP problems [50].

References

[1] O. Achbarou, M. A. E. kiram, and S. E. Bouanani, "Securing cloud computing from different attacks using intrusion detection systems," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 4, no. 3, pp. 61–64, 2017.

[2] R. J. Sethi and Y. Gil, "Scientific workflows in data analysis: Bridging expertise across multiple domains," Future Generation Computer Systems, vol. 75, no. Supplement C, pp. 256 – 270, 2017.

[3] D. Talia, P. Trunfio, and F. Marozzo, "Chapter 3 - models and techniques for cloud-based data analysis," in Data Analysis in the Cloud, ser. Computer Science Reviews and Trends, D. Talia, P. Trunfio, and F. Marozzo, Eds. Elsevier, 2016, pp. 45–76.

[4] "Amazon Simple Workflow Service (SWF)," https://aws.amazon.com/swf, accessed: 2017-10-29.

[5] R. Khorsand, F. Safi-Esfahani, N. Nematbakhsh, and M. Mohsenzade, "Taxonomy of workflow partitioning problems and methods in distributed environments," Journal of Systems and Software, vol. 132, pp. 253–271, 2017.

[6] M. R. Garey and D. S. Johnson, Computers and Intractability; A Guide to the Theory of NP-Completeness. New York, NY, USA: W. H. Freeman & Co., 1990.

[7] E. N. Alkhanak, S. P. Lee, R. Rezaei, and R. M. Parizi, "Cost optimization approaches for scientific workflow scheduling in cloud and grid computing: A review, classifications, and open issues," Journal of Systems and Software, vol. 113, pp. 1–26, 2016.

[8] F. Wu, Q. Wu, and Y. Tan, "Workflow scheduling in cloud: a survey," The Journal of Supercomputing, vol. 71, no. 9, pp. 3373–3418, 2015.

[9] N. M. Ndiaye, P. Sens, and O. Thiare, "Performance comparison of hierarchical checkpoint protocols grid computing," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 1, no. 5, pp. 46–53, 2012.

[10] L. Zeng, B. Veeravalli, and A. Y. Zomaya, "An integrated task computation and data management scheduling strategy for workflow applications in cloud environments," Journal of Network and Computer Applications, vol. 50, pp. 39–48, 2015.

[11] W. T. M. Jr., P. J. Schweitzer, and T. W. White, "Problem decomposition and data reorganization by a clustering technique," Operations Research, vol. 20, no. 5, pp. 993–1009, 1972.

[12] E. E. Mon, M. M. Thein, and M. T. Aung, "Clustering based on task dependency for data-intensive workflow scheduling optimization,"

in 9th Workshop on Many-Task Computing on Clouds, Grids, and Supercomputers (MTAGS2016). IEEE Computer Society, 2016, pp. 20–25.
[13] E. I. Djebbar, G. Belalem, and M. Benadda, "Task scheduling strategy based on data replication in scientific cloud workflows," Multiagent and Grid Systems, vol. 12, no. 1, pp. 55–67, 2016.
[14] L. Zhao, Y. Ren, and K. Sakurai, "Reliable workflow scheduling with less resource redundancy," Parallel Computing, vol. 39, no. 10, pp. 567–585, 2013.
[15] X. Wang, C. S. Yeo, R. Buyya, and J. Su, "Optimizing the makespan and reliability for workflow applications with reputation and a look-ahead genetic algorithm," Future Generation Computer Systems, vol. 27, no. 8, pp. 1124–1134, 2011.
[16] R. Bagheri and M. Jahanshahi, "Scheduling workflow applications on the heterogeneous cloud resources," Indian Journal of Science and Technology, vol. 8, no. 12, 2015.
[17] "Amazon AWS service limits," http://docs.aws.amazon.com/general/latest/gr/aws service limits.html, accessed: 2017-10-29.
[18] "Amazon EC2 on-demand instances limits," https://aws.amazon.com/ec2/faqs/#How many instances can I run in Amazon EC2, accessed: 2017-10-29.
[19] "Amazon EC2 reserved instance limits," http://docs.aws.amazon. com/AWSEC2/latest/UserGuide/ec2-reserved-instances.html#ri-limits, accessed: 2017-10-29.
[20] V. Singh, I. Gupta, and P. K. Jana, "A novel cost-efficient approach for deadline-constrained workflow scheduling by dynamic provisioning of resources," Future Generation Computer Systems, vol. 79, pp. 95–110, 2018.
[21] M. Malawski, G. Juve, E. Deelman, and J. Nabrzyski, "Algorithms for cost and deadline-constrained provisioning for scientific workflow ensembles in iaas clouds," Future Generation Computer Systems, vol. 48, pp. 1–18, 2015, special Section: Business and Industry Specific Cloud.
[22] R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. F. D. Rose, and R. Buyya, "Cloudsim: A toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms," Software: Practice and Experience, vol. 41, no. 1, pp. 23–50, 2011.
[23] C. Wu and R. Buyya, "Chapter 12 - cloud storage basics," in Cloud Data Centers and Cost Modeling. Morgan Kaufmann, 2015, pp. 425–495.
[24] L. Torgo, Data mining with R: learning with case studies. CRC Press, 2016.
[25] M. Charrad, N. Ghazzali, V. Boiteau, and A. Niknafs, "Nbclust: An R package for determining the relevant number of clusters in a data set," Journal of Statistical Software, vol. 61, no. 6, pp. 1–36, 2014.
[26] W. J. Krzanowski and Y. T. Lai, "A criterion for determining the number of groups in a data set using sum-of-squares clustering," Biometrics, vol. 44, no. 1, pp. 23–34, 1988.
[27] T. Cali´nski and J. Harabasz, "A dendrite method for cluster analysis," Communications in Statistics-Simulation and Computation, vol. 3, no. 1, pp. 1–27, 1974.
[28] J. A. Hartigan, Clustering Algorithms, 99th ed. John Wiley & Sons, Inc., 1975.
[29] J. O. McClain and V. R. Rao, "Clustsiz: A program to test of the quality of clustering a set of objects," Journal of Marketing Research, vol. 12, no. 4, pp. 456–460, 1975.
[30] F. B. Baker and L. J. Hubert, "Measuring the power of hierarchical cluster analysis," Journal of the American Statistical Association, vol. 70, no. 349, pp. 31–38, 1975.
[31] F. J. Rohlf, "Methods of comparing classifications," Annual Review of Ecology and Systematics, vol. 5, pp. 101–113, 1974.
[32] G. W. Milligan, "A review of monte carlo tests of cluster analysis," Multivariate Behavioral Research, vol. 16, no. 3, pp. 379–407, 1981.
[33] J. C. Dunn, "Well-separated clusters and optimal fuzzy partitions," Journal of Cybernetics, vol. 4, no. 1, pp. 95–104, 1974.
[34] M. Halkidi, M. Vazirgiannis, and Y. Batistakis, Quality Scheme Assessment in the Clustering Process. Springer Berlin Heidelberg, 2000, pp. 265–276.
[35] M. Halkidi and M. Vazirgiannis, "Clustering validity assessment: Finding the optimal partitioning of a data set," in Proceedings of the 2001 IEEE International Conference on Data Mining, 29 November - 2 December 2001, San Jose, California, USA, 2001, pp. 187–194.
[36] L. J. Hubert and J. R. Levin, "A general statistical framework for assessing categorical clustering in free recall," Psychological Bulletin, vol. 83, no. 6, pp. 1072–1080, 1976.
[37] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," Journal of Computational and Applied Mathematics, vol. 20, no. Supplement C, pp. 53–65, 1987.
[38] G. H. Ball and D. J. Hall, "Isodata, a novel method of data analysis and pattern classification," Stanford research inst Menlo Park CA, Tech. Rep., 1965.
[39] G. W. Milligan, "An examination of the effect of six types of error perturbation on fifteen clustering algorithms," Psychometrika, vol. 45, no. 3, pp. 325–342, 1980.
[40] F. Z. Filali and B. Yagoubi, "Classifying and filtering users by similarity measures for trust management in cloud environment," Scalable Computing: Practice and Experience, vol. 16, no. 3, pp. 289–302, 2015.
[41] R. F. da Silva, W. Chen, G. Juve, K. Vahi, and E. Deelman, "Community resources for enabling research in distributed scientific workflows," in eScience. IEEE Computer Society, 2014, pp. 177–184.
[42] A. C. Zhou, B. He, and S. Ibrahim, "Chapter 18 - escience and big data workflows in clouds: A taxonomy and survey," in Big Data. Morgan Kaufmann, 2016, pp. 431–455.
[43] M. A. Rodriguez and R. Buyya, "Chapter 18 - scientific workflow management system for clouds," in Software Architecture for Big Data and the Cloud. Boston: Morgan Kaufmann, 2017, pp. 367–387.
[44] P. Bryk, M. Malawski, G. Juve, and E. Deelman, "Storage-aware algorithms for scheduling of workflow ensembles in clouds," Journal of Grid Computing, vol. 14, no. 2, pp. 359–378, 2016.
[45] G. Juve, A. Chervenak, E. Deelman, S. Bharathi, G. Mehta, and K. Vahi, "Characterizing and profiling scientific workflows," Future Generation Computer Systems, vol. 29, no. 3, pp. 682–692, 2013, special Section: Recent Developments in High Performance Computing and Security.
[46] M. Masdari, S. ValiKardan, Z. Shahi, and S. I. Azar, "Towards workflow scheduling in cloud computing: A comprehensive analysis," Journal of Network and Computer Applications, vol. 66, no. Supplement C, pp. 64–82, 2016.
[47] J. Sahni and D. P. Vidyarthi, "Workflow-and-platform aware task clustering for scientific workflow execution in cloud environment," Future Generation Computer Systems, vol. 64, no. Supplement C, pp. 61–74, 2016.
[48] C. C. Aggarwal, Ed., Data Classification: Algorithms and Applications. CRC Press, 2014.
[49] L. Kaufman and P. J. Rousseeuw, Finding Groups in Data: an Introduction to Cluster Analysis. Wiley, 1990.
[50] S. I. Serengil and A. Ozpinar, "Workforce optimization for bank operation centers: A machine learning approach," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 4, no. 6, pp. 81–87, 2017.

Sid Ahmed Makhlouf

Ph.D. candidate in the University of Oran1 Ahmed Ben Bella (Algeria). His main research interests include Distributed System, Cluster, Grid & Cloud Computing, Load Balancing, Task & Workflow Scheduling, and Machine Learning.

Belabbas Yagoubi

PhD in Computer Science, is a full professor at the University of Oran1 Ahmed Ben Bella (Algeria). His field of interests is Parallel and Distributed Systems including Security, Fault Tolerance, Replication, Load Balancing and Task Scheduling. He occupies the function of Dean of the Faculty of exact and applied science since 2009.

Driver Fatigue Detection using Mean Intensity, SVM, and SIFT

Saima Naz, Sheikh Ziauddin*, Ahmad R. Shahid

Department of Computer Science, COMSATS Institute of Information Technology, Islamabad (Pakistan)

Received 13 July 2017 | Accepted 23 September 2017 | Published 6 October 2017

Abstract

Driver fatigue is one of the major causes of accidents. This has increased the need for driver fatigue detection mechanism in the vehicles to reduce human and vehicle loss during accidents. In the proposed scheme, we capture videos from a camera mounted inside the vehicle. From the captured video, we localize the eyes using Viola-Jones algorithm. Once the eyes have been localized, they are classified as open or closed using three different techniques namely mean intensity, SVM, and SIFT. If eyes are found closed for a considerable amount of time, it indicates fatigue and consequently an alarm is generated to alert the driver. Our experiments show that SIFT outperforms both mean intensity and SVM, achieving an average accuracy of 97.45% on a dataset of five videos, each having a length of two minutes.

Keywords

Driver Fatigue Detection, Eye Detection, Scale Invariant Feature Transform, Support Vector Machine, Traffic Accidents.

DOI: 10.9781/ijimai.2017.10.002

I. Introduction

Driver fatigue is one of the major causes of vehicle accidents [1],[2]. According to an estimate, up to 35-45% of all accidents are caused by driver fatigue [3]. It is reported that 57% of the fatal truck accidents are caused due to driver fatigue and 50% of the truck drivers report that driving fatigue is the major cause for heavy truck crashes [4]. It makes it imperative to devise intelligent systems that can detect driver fatigue and warn the driver accordingly.

Fatigue can be caused by many reasons such as late-night driving, sleep deprivation, alcohol usage, driving on monotonous roads, medicine intake that causes drowsiness and tiredness or sleep disorders [3], [5]. It slows down reaction time, decreases awareness and impairs judgement while driving. The motivation behind studying the driver fatigue detection problem and alerting the driver when he/she is fatigued, is to decrease the human and financial cost.

Our work has the objective of implementing different techniques and evaluating the performance of each to know which technique is the most suitable for detecting fatigue. Detection can be done based on certain indicators of fatigue. From the literature, we see that there are a few basic symptoms of drowsiness that are feasible to detect using camera and image processing techniques. These symptoms include micro-sleep, bouncing movement of the head and yawning [1], [33]. The bouncing movement of the head can be due to other reasons as well, for example if the driver is listening to music or the road is bumpy. In addition, nodding or swinging of head is apparent after the driver is almost asleep and it might be too late to prevent an accident. Detecting yawning can be a bit misleading as one's mouth may be open for more than one reasons, such as the driver might be singing or talking to his co-passenger. Therefore, bouncing movement of head and yawning are not reliable ways to detect driver fatigue. In this paper, we focus on micro-sleeps to detect fatigue and to prevent accidents. Micro-sleep is the drowsy state in which the driver closes his/her eyes for short intervals of time. He/she may wake up and then fall asleep again. Even this short duration of sleep episode may cause an accident, especially in congested areas. The most intuitive solution is to generate an alarm which will wake up the driver and hence help avoid the potential accident.

* Corresponding author.

E-mail address: sh.ziauddin@gmail.com.

In the proposed scheme, the input is in the form of driver's videos. Frames are extracted from each video and eyes of the driver are localized using Viola-Jones algorithm. After eyes have been localized, they are classified either as open or closed using three different techniques which are Mean Intensity, Support Vector Machine (SVM) and Scale Invariant Feature Transform (SIFT). When the eyes are found closed in a specified number of frames, it is considered as micro-sleep and the alarm is generated to alert the driver. For the evaluation of results, we use accuracy, specificity and sensitivity as performance measures. The results show that SIFT has better performance as compared to the mean intensity and the SVM.

The rest of this paper is organized as follows. A review of existing research literature is given in Section II. In Section III, the working of the proposed scheme is explained. Experimental results of the proposed scheme and the performance evaluation is presented in Section IV. Finally, we conclude this paper in Section V.

II. Related Work

Most of the related work in driver fatigue detection involves the use of computer vision and machine learning techniques, and their combination. The computer vision techniques used for this purpose include Template Based Matching, Feature Based Matching, Histograms of Oriented Gradients (HOG), Gabor Wavelet Transform, Circular Hough Transform, and Landmark Model Matching [6], [7], [8], [9], [10]. The machine learning techniques used in the past for fatigue detection include Markov Chain Framework, Fuzzy Logic, Support Vector Machines (SVM), Naive Bayes (NB), AdaBoost, Particle Swarm Optimization (PSO), and Neural Networks [11], [1], [12], [4]. In the following subsections, we take a brief overview of computer vision and machine learning techniques as have been used in literature.

A. Computer Vision Techniques

Use of non-invasive methods, such as making a video of the driver and alerting him/her on using cues that may help in anticipating the presence of a sleep pattern, can be a useful way to detect driver fatigue [13], [14]. PERCLOS (PERcentage of CLOSure of eyes) is a commonly used method for detection of driver fatigue [2], [15], [12]. It determines the percentage of eye closure by taking the number of frames in which driver's eyes are closed and dividing this by the total number of frames over a specified period of time. Different researchers have used different time windows for PERCLOS calculation such as 20 seconds [16], 30 seconds [36] or 3 minutes [37]. To avoid accidents, it is important to generate an alarm as early as possible when the first symptoms of fatigue are detected. In this paper, we take eye closure in consecutive frames as a measure of fatigue detection instead of PERCLOS. This helps us in generating an alarm much earlier (one second for our selected parameters) in case driver goes in micro-sleep state as compared to PERCLOS.

Localizing the eyes is the first key step towards achieving the task. Facial features in images have been localized by researchers using different techniques. These techniques can be categorized into two groups: feature-based and template-based techniques [7]. Template-based techniques use the object's shape for matching, while feature-based techniques use different geometric features and constraints for their working [1], [16].

Tock and Craw [17] used stored templates and thresholds for detection of regions of interest. For eye detection, the darkest pixel was used to reduce the computational cost. Pupil being the darkest part, is likely to contain the darkest pixel. Eriksson and Papanikolopoulos [1] used both feature-based matching and template-based matching techniques to determine the exact locations of different features. They used reduced regions of the image to detect eyes. They used the heuristic that the regions surrounding the eyes are darker than the other regions in the vicinity.

Khan and Mansoor [2] first detected the face using local Successive Mean Quantization Transform (SMQT) [18]. Next, they drew a square around the face with its center specified. To localize eyes, they divided the square into three parts, with the assumption that the eyes would lie in the top most part of the face. They assumed that the eyes would be open in the beginning and used the first frame in the real-time video to generate an on-line template for open-eyes, which could later be used to determine if the eyes were open or closed. The information regarding the state of eyes (opened or closed) in a sequence of frames was used to detect whether an alarm should be generated or not. They achieved an accuracy of up to 90%.

Brandt et al. [5] presented a visual surveillance system to monitor drivers' head motion and eye blinking pattern. Based on measured features, fatigue was detected by the system. They adopted a coarse to fine strategy to achieve their goal. First, they found the face using Haar wavelets and then they localized the eyes in the already detected face image.

Cherif et al. [19] used the measurement of gaze position to indicate the areas that attract the subject's attention in an image. A calibrated infrared light device was used to provide the horizontal and vertical eye movement. A polynomial transformation of higher order was used to model this mapping by using a mean square error criterion. This helped them to better choose the optimal order to correct the data.

Clement et al. [34] developed a fatigue detection system using output of different low pass and band pass filters on Electro-OculoGram (EOG) and Electro-Encephalogram (EEG) signals. They did not report their results in terms of alarm generation accuracy. Also, their system is not practical as drivers would resist wearing any specific hardware gear while driving on the road.

B. Machine Learning Techniques

Machine Learning techniques have also been increasingly used to detect micro-sleeps owing to their ability to identify patterns. The techniques learn the patterns of closed eyes and open eyes and separate them into two classes. In this section, we discuss machine learning techniques used in the literature to address driver fatigue detection problem.

Dong and Wu [12] combined different cues, such as PERCLOS, head nodding frequency, slouching frequency and Postural Adjustment (PA) for better performance. They used machine learning techniques such as Support Vector Machines and Naive Bayes for classification. The Landmark Model Matching (LMM) [9], [20] was used, which in turn used Particle Swarm Optimization (PSO) [11], [21], [22] to detect the face in the image. Using different cues and fusing them using fuzzy logic achieved better performance than using PERCLOS alone. The use of fused cues reduced the classification error from 15.2% to 12.7%.

Coetzer and Hancke [23] presented a driver fatigue monitoring system. They used three different techniques for classification namely AdaBoost, SVM and Artificial Neural Networks. Devi and Bajaj [4] presented a fatigue detection system. They localized eyes and then tracked the eyes. Their approach also involved yawning detection. Fuzzy inference system was used to detect yawning and micro-sleeps.

Bagci et al. [24] used Markov chain framework to determine whether the eyes are open or closed. The method can be used in monitoring the drivers' alertness as well as in applications which require non-intrusive human computer interactions. Their eye detection and eye tracking algorithms were based on the color and geometrical features of the human face.

San et al. [35] used a deep generic model (DGM) and support vector machine to detect driver fatigue. They compared the performance of power spectrum density features-based SVM and deep generic model-based SVM and found that DGM-based SVM performed better in terms of sensitivity, specifity and accuracy. One drawback of their scheme is that the data is collected using EEG signals which reduces the practicality of the system.

III. Proposed Scheme

In this paper, we use multiple techniques involving image processing/computer vision and machine learning. Before going into the details of the techniques used, first we give a high-level overview of the scheme. A schematic representation of the proposed scheme is given in Fig. 1. In Fig. 1, we show that first we read frames from the input video of the driver. For video generation, the driver is sitting on the driving seat and the video is captured using a single camera which is placed on the car's dashboard behind the steering wheel. In most of the past work, the camera is placed directly in front of the driver at eyes level which is obviously not practical as it obstructs the driver's view of the road ahead. In our dataset generation, we have placed the camera much lower than the eye level of the driver in such a way that it neither obstructs the driver's view of the scene, nor the steering wheel blocks the way between the camera and the driver. After reading the frames from the video, we detect the eyes in every tenth frame using Viola-Jones detector [25], [8]. Once the eyes have been detected, the next step is to classify the eyes into open or closed. The pseudo-code given in Fig. 2 further elaborates the high level working of the proposed scheme.

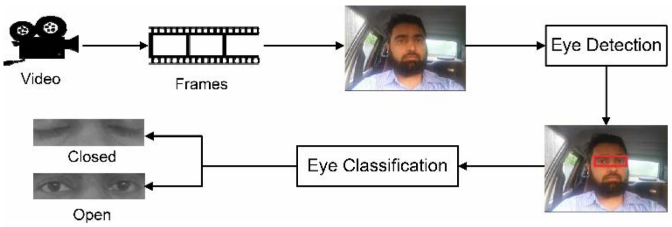


Fig. 1. A high-level overview of the proposed scheme.

```
Begin
    Input driver's video
    ClosedEyeCount := 0
    For every tenth frame in the video Do
        Localize eyes
        Check status of eyes
        If eyes are closed Then
            ClosedEyeCount := ClosedEyeCount + 1
        Else If eyes are open Then
            ClosedEyeCount := 0
        End If
        If ClosedEyeCount == 3 Then
            Generate Alarm
            ClosedEyeCount := 0
        End If
    End For
End
```

Fig. 2. Pseudocode for the proposed alarm generation algorithm.

All processing is done in every tenth frame if three consecutive observations are found where eyes are closed in each observation, and then an alarm is generated to alert the driver. On a video made at 30 frames/sec, this setting corresponds to one second of continuous eye closure which we feel is a reasonable trade-off. If an alarm is generated at a shorter eye closure interval then this, it results in many false positives as some blinks might be detected as micro-sleeps. On the other hand, if we generate alarm at a longer eye closer interval, then it might be too late and an accident might have already occurred before the driver is alerted and he/she can take some appropriate action (e.g., applying the brakes).

Next, we describe the techniques we have used in this paper. As mentioned earlier, we use three techniques namely Mean Intensity, SVM, and SIFT. In the three sub-sections, we describe each of these techniques in more detail.

A. Mean Intensity

Mean intensity is a simple statistical approach used for the classification of open and closed eyes in this paper. When an eye is open, the pupil and the iris are visible. Due to presence of these low intensity regions, the mean intensity of the closed eye image is low as compared to that for an open eye. On the other hand, when eyes are closed, pixels with higher intensity are more frequent, and when the eyes are open, due to the presence of iris, pixels with lower intensity increase.

For each input image (eye region only, detected using Viola-Jones algorithm), we find its mean intensity value and compare it with a threshold T. If mean intensity of the input image is greater than T, then we classify it as an open eye image, otherwise we consider it a closed eye image. The threshold T is selected empirically. A single threshold is selected and tried on all videos instead of selecting five separate thresholds for five videos in our dataset. Using separate thresholds would result in unrealistically higher accuracy of mean intensity results.

B. Support Vector Machine (SVM)

In literature, a number of classification techniques have been used for the task. In this paper, we use SVM for the classification task. SVM being the optimal binary classifier suits well to the problem in hand. SVM is a supervised model for classification which is widely used in Machine Learning and Pattern Recognition [38], [39], [40]. The model is first trained by providing both positive and negative training examples. As SVM is a binary classifier, all training examples belong to one of the two categories. SVM models these training examples as points in space and marks a decision surface in such a way that the gap between the decision surface and any of the two categories is as large as possible. The examples that arrive for testing are mapped to the already created model and a decision is made for each new example whether it belongs to category 1 or category 2 depending upon which side of the border that new example lies.

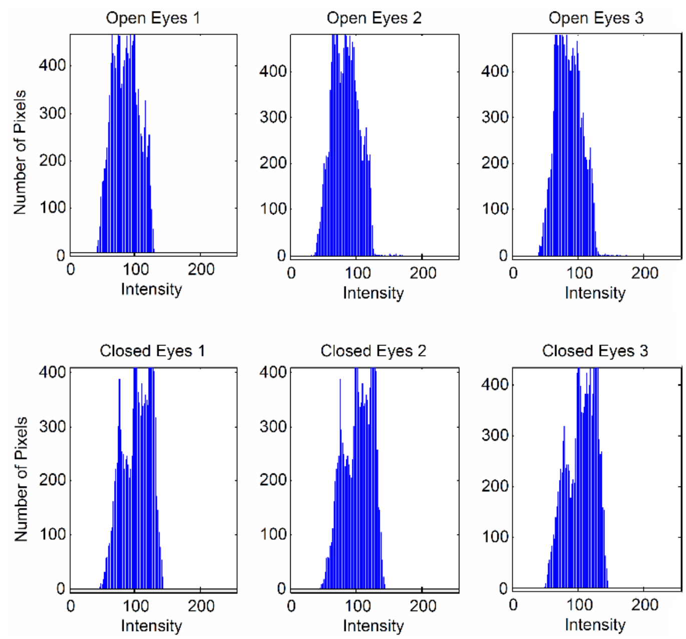


Fig. 3. Histograms of randomly picked three open and three closed eye images.

For the proper working of any classifier, the selection of appropriate features is really important. As we have already mentioned in Section 3.1, the intensity values of an eye image provide useful information regarding whether the eye is open or closed. We have used intensity histograms as features to be used with SVM. Fig. 3 displays histograms of six randomly picked eye images (three each for open and closed eyes). Though there are only subtle differences when these histograms are viewed visually but, as we will see in Section 4, SVM was able to detect these subtle differences in patterns of open and closed eye histograms and providing high classification accuracy based on the histogram features.

C. Scale Invariant Feature Transform (SIFT)

SIFT is one of the most popular algorithms in computer vision. It was presented by David Lowe in 2004 [26], and is patented by University of British Columbia. Since its birth, SIFT has widely been used for object detection and tracking, image registration, panorama stitching, robot localization and mapping, and dense correspondence across scenes. In this paper, we use SIFT to match eye images and classify them into open and closed. Next, we will briefly describe the working of SIFT algorithm. After that, we will explain how we have used SIFT in the proposed technique.

The SIFT algorithm comprises of the following four steps:

1. Scale space construction and extrema detection.
2. Keypoint localization.
3. Orientation assignment.
4. Keypoint description.

1) Scale Space Construction and Extrema Detection

The first step in SIFT is to create a scale space. This is done by creating a pyramid of octaves (set of images) by resampling the image at different sizes and smoothing the image at different scales. At each level of the pyramid, the size of the image is reduced to one-quarter of the size at lower level. For smoothing, the image is convolved with a Gaussian function as given below:

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \quad (1)$$

$$G(x,y,k\sigma) = \frac{1}{2\pi(k\sigma)^2} e^{-(x^2+y^2)/2k^2\sigma^2} \quad (2)$$

where I is the input image, x and y are the coordinates of the image, G is the Gaussian function, L is the resultant blurred image, σ is the smoothness value, and k is a constant that controls the amount of smoothing to be increased at each scale.

Next, Difference of Gaussian (DoG) matrices are generated by subtracting the Gaussian smoothed images at adjacent levels. Then potential features (keypoints) are selected by finding local extrema in a neighborhood of 3*3 around the current pixel as well as around the corresponding pixels at the adjacent upper and lower scales.

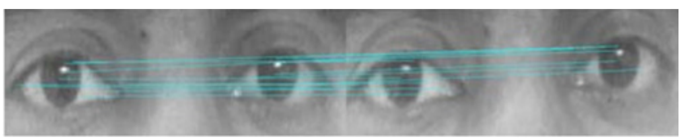


Fig 4. SIFT keypoint matching between the input image and the reference image. The figure is best viewed in color.

2) Keypoint Localization

There are too many candidate keypoints detected in the previous step. Some of these keypoints are unstable which are eliminated in this step. First, low-contrast candidates are eliminated. This is done by using the following second order Taylor series expansion:

$$\boldsymbol{D}(\boldsymbol{x}) = \boldsymbol{D} + \frac{\partial \boldsymbol{D}^T}{\partial \boldsymbol{x}}\boldsymbol{x} + \frac{1}{2}\boldsymbol{x}^T \frac{\partial^2 \boldsymbol{D}}{\partial \boldsymbol{x}^2}\boldsymbol{x} \quad (3)$$

where $\boldsymbol{D}$ is the difference of Gaussian matrix, and $\boldsymbol{x} = [x, y, \sigma]^T$ is the candidate keypoint. The derivative of the above equation is taken to get the localized keypoint $\hat{\boldsymbol{x}}$:

$$\hat{\boldsymbol{x}} = \frac{\partial^2 \boldsymbol{D}^{-1}}{\partial \boldsymbol{x}^2}\frac{\partial \boldsymbol{D}}{\partial \boldsymbol{x}} \quad (4)$$

If the value of $\boldsymbol{D}(\hat{\boldsymbol{x}})$ in (3) is below a threshold, this indicates that the candidate keypoint is a low contrast one and therefore it is discarded. In addition to low contrast candidates, candidates along the edges are also discarded and only those candidates are retained which are on the corners. This is done by computing a 2*2 Hessian matrix of DoG and applying a threshold on its trace and determinant ratio.

3) Orientation Assignment

In this step, we find the magnitude m and the direction θ of each keypoint from the smoothed image $L(x,y)$ as follows:

$$m(x,y) = \sqrt{\left(L(x+1,y) - L(x-1,y)\right)^2 + \left(L(x,y+1) - L(x,y-1)\right)^2} \quad (5)$$

$$\theta(x,y) = tan^{-1}\left(\frac{L(x+1,y)-L(x-1,y)}{L(x,y+1)-L(x,y-1)}\right) \quad (6)$$

Next, in a neighborhood of each keypoint, a weighted histogram of orientations having 36 bins is created. The bin that gets the highest weighted sum is selected as the orientation of that keypoint.

4) Keypoint Description

In this step, a 16x16 neighborhood around each keypoint is taken and, for each pixel in that neighborhood, the magnitude and orientation is calculated. Each 16x16 block is further divided into 16 sub-blocks of 4x4 size. For each sub-block, 8-bin orientation histogram is created. These values are represented as a 128-dimensional vector (concatenation of 8 bin values for 16 sub-blocks) to form the keypoint descriptor.

For object matching, the keypoints from the input image are matched against all those in the database. The matching is done using the keypoint descriptors. For each keypoint, nearest neighbor is identified from the database where nearest neighbor is that descriptor from the database which has the minimum Euclidean distance from the input keypoint descriptor.

In this paper, we perform SIFT keypoint extraction on all eye images. First, we have randomly selected two images (one each for open and closed eyes) to be used as reference images. Then for each video, SIFT keypoints from each eye image are matched with those from the two reference images. If the feature matching score is higher between the input image and the open eye reference image as compared to that between the input image and the closed eye reference image, the input image is considered as an open eye image. On the other hand, if the input image and the reference closed eye image have more keypoints in common, the input image is classified as a closed eye image.

Fig. 4 shows an example of an open eye input image where the keypoints matched with the reference open eye image are displayed. The image on left is the input image while that on right is the reference image. As it can be seen, there are many points which are matched by the SIFT keypoint matcher. On the other hand, when the same input image is matched with the sample closed eye image, only one keypoint matches between the pair of images.

IV. Experimental Evaluation

A. Video Dataset

For our experiments, we have created a dataset by placing the camera inside the vehicle on the dashboard behind the steering wheel. The camera is placed in such a way that it does not occlude driver's view of the scene. Also, the steering wheel is not hiding driver's face from the camera. We collected our data under different conditions of lighting and drivers' clothing, and with varying degrees of cloud cover and natural daylight. It contains five videos (V1-V5) of a single driver, which are captured at the rate of 30 frames per second using a two mega-pixels smart phone camera. Each video is of approximately two minutes duration.

For eye classification and alarm generation, we processed every tenth frame of each video so that the frames with normal blink are not included in the processed frames. This setting results in approximately 360 frames to be processed for each video (30/10 * 120). We generated the ground truth by visually looking at each frame and deciding whether the eyes are open or closed in that particular frame. Table I provides further description of our dataset.

B. Eye Detection Results

For eye detection, we have used the well-known Viola-Jones detector [25], [8]. Viola-Jones object detection framework uses rectangular Haar features and creates an integral image from these features. As a learning algorithm, they used a variant of AdaBoost which performs feature selection and trains the classifier. Finally, the classifier cascading is done by doing the processing in different stages. At each stage, sub-windows are classified as may be face or definitely non-face. All sub-windows which are classified as definitely non-faces are discarded and hence reducing the number of sub-windows at each stage which results in decreased computational complexity of the algorithm.

Viola-Jones detector detected eyes correctly in 1806 out of 1819 frames resulting in a correct detection rate of 99.3% as shown in Table I. We discarded those 13 frames where eyes were not correctly detected and continued our experiments for eye classification and alarm generation using the remaining 1806 frames.

TABLE I. Summary of Dataset and the Ground Truth

	V1	V2	V3	V4	V5	Total
No. of Frames	3590	3610	3640	3600	3750	18190
No. of Processed Frames	359	361	364	360	375	1819
Tracking Failures	1	0	3	3	6	13
No. of Eye Images	358	361	361	357	369	1806
No. of Instances with Closed Eyes	153	139	128	134	130	684
No. of Instances with Open Eyes	205	222	233	223	239	1122

C. Evaluation Metrics

For the task of eye classification, we have used three evaluation metrics namely *accuracy*, *specificity*, and *sensitivity*. The formulae for calculating these values are given as under:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$Specificity = \frac{TN}{TN + FP} \tag{8}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{9}$$

where TP (True Positives) are those cases where a closed eye is classified as closed and TN (True Negatives) are those when an open eye is classified as open. Similarly, FP (False Positives) corresponds to the cases where an open eye is classified as closed, and FN (False Negatives) to the cases where a closed eye is classified as open.

D. Eye Classification Results

We use mean intensity, SVM and SIFT to classify eyes as open or closed. For mean intensity and SIFT, we test the algorithm on all video frames and report our results here. On the other hand, for SVM, we use 5-fold cross validation where four of the five videos are used for training and the remaining one for testing in each iteration. Tables II, III, and IV show eye classification results for mean intensity, SVM, and SIFT, respectively. The values of False Positive (FP) Rate and False Negative (FN) Rate used in tables are calculated as follows:

$$FP\ Rate = \frac{FP}{FP + TN} \tag{10}$$

$$FN\ Rate = \frac{FN}{FN + TP} \tag{11}$$

From Tables II, III, and IV, we can see that SIFT provides highest accuracy among the three techniques.

During our experiments, we observed that the number of SIFT key points are much larger on average for the open eyes as compared to that for the closed eyes. Using this observation, we perform another experiment where we have classified the eyes based upon only the number of key points detected by the SIFT algorithm instead of matching those key points against the reference images. We have named this part of SIFT algorithm as SIFT-K. The eye classification results using SIFT-K are shown in Table V. The results of SIFT and SIFT-K are almost equal but SIFT-K has the advantage that it has better running time than that of SIFT.

Table VI compares the eye classification results obtained by mean intensity, SVM, SIFT, and SIFT-K. SIFT-K has the highest accuracy and specificity values while SIFT has highest sensitivity among all techniques.

TABLE II. Eye Classification Results using Mean Intensity

	V1	V2	V3	V4	V5	Overall
FP Rate (%)	0.49	1.35	2.58	31.84	0.42	7.31
FN Rate (%)	5.88	2.16	0.78	0.00	6.15	3.07
Accuracy (%)	97.21	98.34	98.06	80.11	97.56	94.3
Specificity (%)	99.76	99.39	98.88	82.68	99.82	92.69
Sensitivity (%)	96.08	98.44	99.39	100.0	95.2	96.93

TABLE III. Eye Classification Results using SVM

	V1	V2	V3	V4	V5	Overall
FP Rate (%)	0.49	18.47	2.15	2.69	0.42	4.81
FN Rate (%)	3.27	28.06	6.25	0.00	6.15	8.77
Accuracy (%)	98.32	77.84	96.4	98.32	97.56	93.69
Specificity (%)	99.76	90.74	99.07	98.78	99.82	95.19
Sensitivity (%)	97.84	78.09	95.05	100.0	95.2	91.23

TABLE IV. Eye Classification Results using SIFT

	V1	V2	V3	V4	V5	Overall
FP Rate (%)	2.93	4.05	4.72	3.14	2.09	3.39
FN Rate (%)	0.00	0.72	3.13	2.24	1.54	1.46
Accuracy (%)	98.32	97.23	95.84	97.2	98.1	97.34
Specificity (%)	98.55	98.13	97.92	98.57	99.11	96.61
Sensitivity (%)	100.0	99.48	97.54	98.32	98.81	98.54

E. Alarm Generation Results

In a particular frame, if the eyes are classified as closed, this can mean one of the two things: either the driver is blinking his/her eyes, or he/she is in a state of micro-sleep. To differentiate between the two conditions, we used a threshold of three i.e. if the eyes are closed in three or more consecutive processed frames, it is considered a microsleep (see Fig. 2) and results in alarm generation. All the alarms generated using the proposed techniques are then compared against those in the ground truth to find out the missed alarms as well as the false alarms.

Tables VII, VIII, IX, and X show results for alarm generation using mean intensity, SVM, SIFT, and SIFT-K, respectively.

TABLE V. Eye Classification Results using SIFT-K

	V1	V2	V3	V4	V5	Overall
FP Rate (%)	1.46	3.15	2.58	0.9	2.51	2.14
FN Rate (%)	1.96	2.16	5.47	4.48	2.31	3.22
Accuracy (%)	98.32	97.23	96.40	97.76	97.56	97.45
Specificity (%)	99.28	98.55	98.88	99.60	98.93	97.86
Sensitivity (%)	98.71	98.44	95.68	96.62	98.22	96.78

TABLE VI. Overall Eye Classification Results for Mean Intensity, SVM, SIFT and SIFT-K

Technique	Accuracy (%)	Specificity (%)	Sensitivity (%)
Mean Intensity	94.3	92.69	96.93
SVM	93.69	95.19	91.23
SIFT	97.34	96.61	**98.54**
SIFT-K	**97.45**	**97.86**	96.78

Finally, Fig. 5 shows and compares incorrect alarms generated by all techniques. As we can see, SIFT-K provides the best results here resulting in nine incorrect alarms (either missed or falsely generated). Though we have considered both false alarms and the missed alarms, it is obvious that missing an alarm is much more critical than generating a false alarm. In terms of missed alarms also, SIFT-K has the best performance followed by SIFT, SVM, and mean intensity, respectively.

TABLE VII. Alarm Generation Results for Mean Intensity

Criteria	V1	V2	V3	V4	V5	Overall
Number of Micro-sleeps	37	31	30	29	30	157
Alarms Generated	34	31	31	56	28	180
Correctly Detected Alarms	34	30	29	29	27	149
Missed Alarms (False Negatives)	3	1	1	0	3	8
False Alarms (False Positives)	0	1	2	27	1	31

TABLE VIII. Alarm Generation Results for SVM

Criteria	V1	V2	V3	V4	V5	Overall
Number of Micro-sleeps	37	31	30	29	30	157
Alarms Generated	35	31	30	31	30	157
Correctly Detected Alarms	35	30	28	29	29	151
Missed Alarms (False Negatives)	2	1	2	0	1	6
False Alarms (False Positives)	0	1	2	2	1	6

TABLE IX. Alarm Generation Results for SIFT

Criteria	V1	V2	V3	V4	V5	Overall
Number of Micro-sleeps	37	31	30	29	30	157
Alarms Generated	39	35	31	30	31	166
Correctly Detected Alarms	37	31	27	27	30	152
Missed Alarms (False Negatives)	0	0	3	2	0	5
False Alarms (False Positives)	2	4	4	3	1	14

TABLE X. Alarm Generation Results for SIFT-K

Criteria	V1	V2	V3	V4	V5	Overall
Number of Micro-sleeps	37	31	30	29	30	157
Alarms Generated	37	33	31	28	31	160
Correctly Detected Alarms	36	31	29	28	30	154
Missed Alarms (False Negatives)	1	0	1	1	0	3
False Alarms (False Positives)	1	2	2	0	1	6

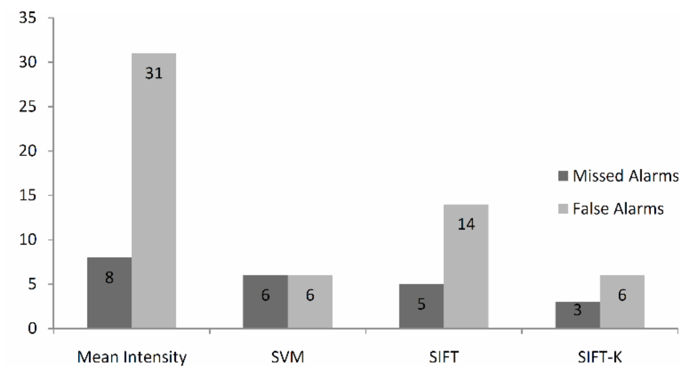


Fig. 5. Comparison of missed alarms and false alarms for different techniques.

F. Discussion on Results

In our experiments, we tried mean intensity, SVM, SIFT and SIFT-K for both eye classification and alarm generation. For eye classification, SIFT-K provided the best average accuracy (97.45%) followed by SIFT (97.34%). Though the average accuracies of mean intensity and SVM are also quite good (94.3% and 93.69%, respectively), it is the standard deviation where they are totally outperformed by SIFT and SIFT-K. SIFT and SIFT-K have standard deviations of 0.71 and 0.98, respectively. On the other hand, mean intensity has a standard deviation of 7.92 while SVM has 8.89. When we see the results of individual videos, we see that there is one video each for mean intensity (Video 4) and SVM (Video 2) where the accuracy is much lower than the other videos (see Tables II and III). SIFT and SIFT-K do not suffer from this variation. As we see from Tables IV and V, both techniques provide high accuracy for all the videos in the dataset. For SIFT, the minimum accuracy for any video is 95.84% while for SIFT-K, it is 96.4%. Overall, we can conclude that SIFT-K and SIFT perform much better as compared to the other two techniques.

V. Conclusion

In this paper, we presented a driver fatigue detection scheme. Eyes of the driver were detected in the first step and then it was analyzed whether the eyes were open or closed. If eyes remained closed for a specified amount of time then an alarm was generated to alert the driver. We used Viola-Jones algorithm to detect eyes and then three different approaches were used to classify open eyes and closed eyes i.e., mean intensity, SIFT and SVM. We saw that SIFT performed better for eye classification, alarm generation and fatigue detection as compared to the other tested techniques. In order to be more confident about our results, we need to enhance our dataset not just in terms of the number of drivers but also their characteristics. In future, we intend to create larger datasets with different subjects belonging to both the genders, with varying degrees of facial hair and other accessories such as glasses. The dataset would also be collected for the night tine. Also, different camera locations on the dashboard would be experimented with evaluation of the best positions.

References

[1] M. Eriksson and N.P. Papanikolopoulos, "Driver fatigue: a vision based approach to automatic diagnosis," Transportation Research Part C: Emerging Technologies, vol. 9, no. 6, pp. 399–413, 2001.

[2] M.I. Khan and A.B. Mansoor, "Real time eyes tracking and classification for driver fatigue detection," in Image Analysis and Recognition, pp. 729–738, Springer, 2008.

[3] S.K. Lal and A. Craig, "Driver fatigue: electroencephalography and psychological assessment," Psychophysiology, vol. 39, no. 3, pp. 313–321, 2002.

[4] M.S. Devi and P.R. Bajaj, "Fuzzy based driver fatigue detection," Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on, pp. 3139–3144, IEEE, 2010.

[5] T. Brandt, R. Stemmer, and A. Rakotonirainy, "Affordable visual driver monitoring system for fatigue and monotony," Systems, Man and Cybernetics, 2004 IEEE International Conference on, pp. 6451– 6456, IEEE, 2004.

[6] I. Craw, H. Ellis, and J.R. Lishman, "Automatic extraction of face features," Pattern Recognition Letters, vol. 5, no. 2, pp. 183–187, 1987.

[7] R. Brunelli and T. Poggio, "Face recognition: Features versus templates," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 15, no. 10, pp. 1042–1052, 1993.

[8] P. Viola and M.J. Jones, "Robust real-time face detection," International journal of computer vision, vol. 57, no. 2, pp. 137–154, 2004.

[9] R.S. Senaratne and S.K. Halgamuge, "Optimal weighting of landmarks for face recognition," Journal of Multimedia, vol. 1, no. 3, pp. 31–41, 2006.

[10] J. Chen, Y. Lin, L. Wang, Z. Jiang, and G. Wang, "An algorithm for fatigue pattern detection based on parallel Gabor and 1-nearest neighbor," Fourth International Conference on Digital Image Processing (ICDIP 2012), pp. 833410–833410, International Society for Optics and Photonics, 2012.

[11] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," Micro Machine and Human Science, 1995. MHS'95, Proceedings of the Sixth International Symposium on, pp. 39–43, IEEE, 1995.

[12] W. Dong and X. Wu, "Fatigue detection based on the distance of eyelid," VLSI Design and Video Technology, 2005. Proceedings of 2005 IEEE International Workshop on, pp. 365–368, IEEE, 2005.

[13] Q. Ji, Z. Zhu, and P. Lan, "Real-time nonintrusive monitoring and prediction of driver fatigue," Vehicular Technology, IEEE Transactions on, vol. 53, no. 4, pp. 1052–1068, 2004.

[14] P. Smith, M. Shah, and N. da Vitoria Lobo, "Determining driver visual attention with one camera," Intelligent Transportation Systems, IEEE Transactions on, vol. 4, no. 4, pp. 205–218, 2003.

[15] D.F. Dinges and R. Grace, "Perclos: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance," Federal Highway Administration. Office of motor carriers, Tech. Rep. MCRT-98-006, 1998.

[16] R. Senaratne, D. Hardy, B. Vanderaa, and S. Halgamuge, "Driver fatigue detection by fusing multiple cues," in Advances in Neural Networks– ISNN 2007, pp. 801–809, Springer, 2007.

[17] D. Tock and I. Craw, "Tracking and measuring drivers' eyes," Image and Vision Computing, vol. 14, no. 8, pp. 541–547, 1996.

[18] M. Nilsson, J. Nordberg, and I. Claesson, "Face detection using local smqt features and split up snow classifier," Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on, pp. II–589, IEEE, 2007.

[19] Z.R. Cherif, A. Nait-Ali, J. Motsch, and M. Krebs, "An adaptive calibration of an infrared light device used for gaze tracking," Instrumentation and Measurement Technology Conference, 2002. IMTC/2002. Proceedings of the 19th IEEE, pp. 1029–1033, IEEE, 2002.

[20] R. Senaratne, B. Jap, S. Lal, A. Hsu, S. Halgamuge, and P. Fischer, "Comparing two video-based techniques for driver fatigue detection: classification versus optical flow approach," Machine Vision and Applications, vol. 22, no. 4, pp. 597–618, 2011.

[21] Y. Shi and R.C. Eberhart, "Empirical study of particle swarm optimization," Evolutionary Computation, 1999. CEC 99. Proceedings of the 1999 Congress on, IEEE, 1999.

[22] A. Ratnaweera, S.K. Halgamuge, and H.C. Watson, "Self organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients," Evolutionary Computation, IEEE Transactions on, vol. 8, no. 3, pp. 240–255, 2004.

[23] R.C. Coetzer and G.P. Hancke, "Eye detection for a real-time vehicle driver fatigue monitoring system," Intelligent Vehicles Symposium (IV), 2011 IEEE, pp. 66–71, IEEE, 2011.

[24] A. Bagci, R. Ansari, A. Khokhar, and E. Cetin, "Eye tracking using markov models," Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on, pp. 818–821, IEEE, 2004.

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

[26] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.

[27] L.M. Bergasa, J. Nuevo, M.A. Sotelo, R. Barea, and M.E. Lopez, "Real-time system for monitoring driver vigilance," Intelligent Transportation Systems, IEEE Transactions on, vol. 7, no. 1, pp. 63– 77, 2006.

[28] G. Borghini, L. Astolfi, G. Vecchiato, D. Mattia, and F. Babiloni, "Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness," Neuroscience & Biobehavioral Reviews, vol. 44, pp. 58–75, 2014.

[29] A.A. Lenskiy and J.S. Lee, "Driver's eye blinking detection using novel color and texture segmentation algorithms," International Journal of Control, Automation and Systems, vol. 10, no. 2, pp. 317– 327, 2012.

[30] M. Sacco, R. Farrugia, et al., "Driver fatigue monitoring system using support vector machines," Communications Control and Signal Processing (ISCCSP), 2012 5th International Symposium on, pp. 1–5, IEEE, 2012.

[31] H. Shen and M. Xu, "Design and implementation of embedded driver fatigue monitor system," 2015 International Conference on Artificial Intelligence and Industrial Engineering, Atlantis Press, 2015.

[32] S. Singh, and N. P. Papanikolopoulos. "Monitoring driver fatigue using facial analysis techniques," Intelligent Transportation Systems, 1999. Proceedings. 1999 IEEE/IEEJ/JSAI International Conference on. IEEE, 1999.

[33] A. Shabnam, B. Hariri, and S. Shirmohammadi. "Driver drowsiness monitoring based on yawning detection," Instrumentation and Measurement Technology Conference (I2MTC), 2011 IEEE. IEEE, 2011.

[34] F. S. C. Clement, A. Vashistha, and M. E. Rane. "Driver Fatigue Detection System," International Conference on Information Processing, 2015.

[35] P. P. San, S. H. Ling, R. Chai, Y. Tran, A. Craig and H. Nguyen, "EEG-based driver fatigue detection using hybrid deep generic model," 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, 2016, pp. 800-803.

[36] D. Prokhorov. Computational intelligence in automotive applications. Springer-Verlag, 2008.

[37] S. Darshana, D. Fernando, S. Jayawardena, S. Wickramanayake and C. DeSilva, "Efficient PERCLOS and Gaze Measurement Methodologies to Estimate Driver Attention in Real Time," 2014 5th International Conference on Intelligent Systems, Modelling and Simulation, Langkawi, 2014, pp. 289-294.

[38] D. Pujari, J. R. Yakkundimath and A. S. Byadgi, "SVM and ANN Based

Classification of Plant Diseases Using Feature Reduction Technique," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 3, no. 7, pp. 6–14, 2016.

[39] F. Kurtulmuş, I. Kavdir, "Detecting Corn Tassels using Computer Vision and Support Vector Machines," Expert Systems with Applications, vol. 41, no. 16, pp. 7390-7397, 2014.

[40] É. Niaf, R. Flamary, A. Rakotomamonjy, O. Rouvière and C. Lartizien, "SVM with feature selection and smooth prediction in images: Application to CAD of prostate cancer," 2014 IEEE International Conference on Image Processing (ICIP), Paris, 2014, pp. 2246-2250.

Saima Naz

Saima Naz completed MS in Computer Science from COMSATS Institute of Technology in Islamabad, Pakistan. She is currently working as a Program Manager at iENGINEERING (Private) Limited in Islamabad. Her research interests include computer vision, image processing, pattern recognition and project management.

Sheikh Ziauddin

Sheikh Ziauddin is an assistant professor in department of computer science, COMSATS Institute of Information Technology in Islamabad, Pakistan. He did his PhD in Computer Science from Asian Institute of Technology in Bangkok, Thailand. He also completed Post-Doc from University of Nice in Nice, France. He is an author of more than 25 international conference and journal publications. His research interests include computer vision, image processing, pattern recognition, and machine learning.

Ahmad R. Shahid

Ahmad R. Shahid is an assistant professor in department of computer science, COMSATS Institute of Information Technology in Islamabad, Pakistan. He did his PhD in Computer Science from University of York in York, United Kingdom. His research interests include machine learning, pattern recognition, computer vision, and natural language processing.

PRACTICA. A Virtual Reality Platform for Specialized Training Oriented to Improve the Productivity

Juan Manuel Lombardo, Miguel Ángel López, Vicente M. García, Mabel López, Rubén Cañadas, Susana Velasco, Mónica León *

Fundación I+D del Software Libre (FIDESOL), Granada (Spain)

Received 12 December 2017 | Accepted 19 March 2018 | Published 27 April 2018

Abstract

With the proliferation of Virtual reality headset that are emerging into a consumer-oriented market for video games, it will open new possibilities for exploiting the virtual reality (VR). Therefore, the PRACTICA project is defined as a new service aimed to offering a system for creating courses based on a VR simulator for specialized training companies that allows offering to the students an experience close to reality. The general problem of creating these virtual courses derives from the need to have programmers that can generate them. Therefore, the PRACTICA project allows the creation of courses without the need to program source code. In addition, elements of virtual interaction have been incorporated that cannot be used in a real environment due to risks for the staff, such as the introduction of fictional characters or obstacles that interact with the environment. So to do this, artificial intelligence techniques have been incorporated so these elements can interact with the user, as it may be, the movement of these fictional characters on stage with a certain behavior. This feature offers the opportunity to create situations and scenarios that are even more complex and realistic.This project aims to create a service to bring virtual reality technologies closer and artificial intelligence for non-technological companies, so that they can generate (or acquire) their own content and give it the desired shape for their purposes.

Keywords

Virtual Reality, Simulator, Immersion (Virtual Reality), Occupational-Hazards, Training, Virtual Reality-Headset, Artificial Intelligence, Pathfinder.

DOI: 10.9781/ijimai.2018.04.007

I. Introduction

Nowadays, the rise has occurred in the application of virtual reality for areas outside of video games it's getting higher, and this is thanks to the impulse that this technology has had thanks to the appearance of the VR *headset*, that were originally intended for video games but have proven to offer potential in other applications. Nevertheless, the use of these technologies is not yet being exploited for the creation of simulation environments in training centers which requires performing practices with dangerous equipment dangerous equipment of high cost or that incorporates risk for other people, despite the advantages and the chances it offers that are not possible in real environments or in environments where a lot of precautions need to be taken, for example, put students at risk, in reality, most of the time it's not possible. But, in a virtual environment, where the risk doesn't really exist, can lead to even extreme situations that will help the student to have a deeper understanding of his or her profession and the risks involved.

It's true that, since many years [1], there are virtual simulators applied to the training of personnel in the industrial sector, however, its use is much lower in terms of training centers that increasingly apply technologies to teaching but the use of VR is still very scarce [2]. However, this is a trend that is changing with the impact of virtual immersion, that has increased with the release of these VR headset to the market. These devices have a relatively low cost and are compatible with different programming systems and creation of virtual environments. The combination of both technologies has boosted the proliferation of simulation systems with surprising results. Nevertheless, most of these simulators have been developed for the purpose of creating 3D environments that support specific practices and courses.

The difference between the project PRACTICA with these simulators lies in that PRACTICA has been conceived as a tool for the creation of simulation environments of free access to the community of developers and creatives, with the differential value regarding existing solutions, that the goal of PRACTICA is to provide a means for the creation of these simulators without the need to write a line of code. This facilitates access to a tool for the creation of simulators based on virtual reality for (non-ICT) professionals. PRACTICA provides a scene editing system, fully parameterize and configurable through visual elements.

This article describes how the PRACTICA project works, from its conception, design and development, to the results obtained in the creation of prototype simulators adapted to the needs of real cases of companies that carry out training for professionals that have improved the content of the courses they provide to their students, given by access to a system of practices that simulates the operation of high-cost machinery and that puts students in situations that could actually pose a risk to their integrity.

In addition to that, the capabilities of the different tools studied and the ease with which artificial intelligence techniques can be

* Corresponding author.

E-mail addresses: jmlombardo@fidesol.org (J. M. Lombardo), malopez@fidesol.org (M. Á. López), vmgarcia@fidesol.org (V. M. García), milopez@fidesol.org (M. López), rjcanadas@fidesol.org (R. Cañadas), svelasco@fidesol.org (S. Velasco), mleon@fidesol.org (M. León).

incorporated have been analyzed for the insertion of elements that function as obstacles or challenges within the exercise, that could not be used in a real environment because of the risk it would entail, such as the incorporation of fictitious people.

II. State of the Art

Since the first virtual reality devices appeared, the potential of this technology for education has been noted [3], mainly, about the access to content that may have access limitations, either by its cost or accessibility, or because of the difficulty of carrying out. Nevertheless, it's important to know if through a VR system you can achieve results similar to a real environment [4] and whether the result of learning from the use of a simulator is really effective.

A virtual reality system has the ability to offer an immersive experience in an environment similar to reality, which incorporates elements that can be viewed from a first-person perspective through self-centered vision [5], being able to even perceive a virtual body that interacts with the rest of the elements. This makes the user identify quickly with his avatar and the feeling of immersion be superior.

Thus, in the PRACTICA project, one of the elements that should not be missing in the creation of a course is the incorporation of an element or avatar that is directly related to parts of our body and that allows a better understanding of the reality that is being seen. In order to contribute to this, it has also been necessary to have a device that allows the incorporation of hand gestures by means of infrared recognition, or, through the use of physical controls, to enhance interaction with the environment. Among those technologies stand out by type: (1) virtual reality systems, which are mainly composed of headsets that allow stereoscopic vision, although some include some additional features; (2) systems of interaction between the user and the virtual environment, and no less important; (3) a flexible virtual environment creation tool that is compatible with the virtual reality systems under analysis.

A. Analysis of Virtual Reality Headset Suitable for PRACTICA

Currently there are different virtual reality systems. Which, depending on their characteristics and cost, may have one use or another. For this project, the analysis has focused on more or less affordable VR headsets for a domestic market with the following minimum characteristics: (1) allow virtual immersion; (2) have support for Cross-platform for creating virtual environments; (3) are available in a domestic market; (4) have freely usable drivers that meet the above requirements. There are several systems however, the study conducted for this project has focused on technologies that have emerged in recent years, and which are the most widely accepted in the market: Google Cardboard, Oculus Rift, HTC Vive y PlayStation VR.

1) Google Cardboard

Google Cardboard [6], is now the simplest and most economical VR system. His concept is basic: the mobile devices screen is used, in this case: smartphones based in Android o iOS, divided with two scenes so that the mobile phone is inserted into a viewfinder, which can be cardboard, plastic o ABS, that has two lenses through which you can see the screen of mobile phone and provides the basic stereoscopic vision for virtual immersion.

2) Daydream

The evolution of Cardboard is Daydream [7]. Google Daydream is an enhanced VR platform from Google for compatible mobile devices (Daydream Ready) from the version 7.x of Android (also known as Nougat). For it, they have put new VR headsets on the market called DayDream View which has a remote control that allows you to control the device and the visualization. Further, that for show VR environments, is designed for immersive reproduction of 360-degree audiovisual content.

3) Oculus Rift

Oculus Rift, they are the reference VR headsets, it was the consumption of virtual reality. In 2009, Palmer Luckey designed the first prototype, called The Rift [8] that he carried out a demonstration of its potential along with John Carmark with a modified version of Doom 3. From then on, a campaign was launched in Kickstarter for the first mass wear VR headsets that were widely accepted, which was the final takeoff of this device. Nowadays, the latest version of the Oculus Rift is composed of a series of elements, along with the VR headsets, that allow to get a much better immersion experience than the one obtained with the first device. These elements are: (1) VR headsets or viewfinder with two state-of-the-art monitors and customized optics [9]; (2) Touch controllers that allow user interaction through these controls showing some arms on the scene that provide a greater sense of reality; (3) also incorporates two sensors with an infrared mesh to capture hand movement, as do other devices such as LeapMotion [10] or Kinect [11].

4) HTC Vive

After the initial success of Oculus Rift, and other major companies in the world of technology and video games, they bet on virtual reality to offer their customers new experiences. This is the case of two large companies such as HTC y Valve (company that initially collaborated with Oculus, however, with the acquisition by Facebook of the latter, this collaboration would be broken.) who came together to compete with Oculus in this emerging market [12].

The first VR headsets that emerged from this collaboration were called VIVE PRE and were offered to video game developers. The policy of these companies is open, and the community has therefore had a great influence on the penetration of this device in the VR video game sector.

The HTC Vive has two OLED screens and its field of view is 110 degrees. They come with two controls, as well as the Oculus Rift and two laser devices for motion detection, that detect the movement of the controls and the VR headsets. Currently working on a new wireless version.

5) Sony Playstation VR

Finally, there is another great company that has bet on the creation of its own virtual reality headsets, focused on the world of entertainment and video games. This company is Sony and has launched the Playstation VR [13].

These VR headsets do not come with specific controls and motion sensors but take advantage of the technology that comes with the console to offer an immersive experience. These technological elements are the controls of the console and the second-generation camera that have taken to the market. Unlike the previous two, these VR headsets have a 5.7-inch OLED display with FullHD resolution, refresh rate of up to 120Hz and a field of view of 110 degrees. They can be connected to a PC using specific drivers.

6) Comparative

Table I includes the main characteristics of each of the above systems, considering: the type of viewfinder, refresh rate, resolution, maximum field of view and minimum area of movement of the person using the system:

TABLE I. Comparison of the Characteristics of Cardboard/Daydream, Oculus Rift, HTC Vive and Playstation VR

Name	Viewfinder type	Refresh rate	Resolution	Field of view	Area of movement
Cardboard / Daydream	Optical Lenses	Depends on the device	Depends on the device	90 degrees	It's not relevant
Oculus Rift	2 monit. OLED	90HZ	2160x1200	110 degrees	1,5mx3,3m
HTC Vive	2 monit. OLED	90Hz	2160x1200	110 degrees	5mx5m
Playstation VR	1 monit. OLED of 5,7"	120Hz	1920x1080	100 degrees	It's not relevant

As can be seen in Table I, the Google VR headsets (Cardboard / Daydream) depend on the device used, so the experience can be very variable with them. Also, the angle of view is somewhat limited. On the other hand, it can be seen that Oculus Rift and HTC Vive have a very similar technology, except for the difference in movement area between the two technologies, according to which Vive allows a larger area, which can be interesting from the point of view of creating courses that require large movement space. Finally, Sony virtual reality headsets is designed for a sitting experience in front of the Playstation console, so it's very limited in terms of the ability to use it to create courses. In addition, it's only intended for use with the Sony console.

Considering the above, the best options for use in the development of the Practica project are Oculus Rift and HTC Vive, due to their similar features. However, due to economic cost (HTC Vive have a higher cost at the time of writing) the decanting is carried out in favour of Oculus Rift.

B. Tools for Creating Virtual Environments

Another fundamental element for the development of a platform for the creation of virtual environments has been to choose the right base tool that already exists and is freely available. There are currently several *frameworks* for the creation of applications and games that recreate virtual scenarios in two and three dimensions. For this project, the incorporation of the following has been studied: (1) Unity 3D; (2) Unreal Engine; (3) Blender.

1) Unity 3D

Unity3D is a *framework* for the creation of application and video games 2D/3D, with strong support from major technology and software partners such as Facebook, Google, Intel, Microsoft, Nintendo, Samsung, Sony, etc., with a great acceptance and success stories, both by independent developers (Monument Valley, Crossy Road, Temple Run, etc.), and by large companies, such as EA, Square Enix, Ubisoft, Eidos, etc.

It has great support for virtual reality peripherals, both native and external using 3rd party SDKs such as Oculus Rift, HTC Vive/SteamVR, Ximmerse, Leap Motion, Google Cardboard/Daydream, etc.

In addition, it allows you to export to a wide range of platforms, including Windows, Linux, MacOS, Android, iOS, PlayStation 4, XBox One, etc.

Also, the IDE can be extended through a specific API for the creation and support of plugins, which allows to extend its functionality and adjust it to the needs of each developer and/or project.

The Unity3D personal license allows for free and unlimited use for personal or educational development. It is also free for commercial use as long as the annual income does not exceed $100,000.

The strength of Unity3D is its extensive documentation, a large and growing community, along with a large Asset Store where you can find all kinds of high quality resources, both free and paid.

2) Unreal Engine 4

Is a 3D game development framework with a more powerful and professional finishing engine, more geared towards AAA projects and companies, such as Square Enix, THQ, Sony Interactive Entertainment, Bandai Namco, etc, and an IDE more focused on designers than developers.

Like Unity3D, it can export to a wide range of systems, including Windows, MacOS, Android, iOS, PlayStation 4, Xbox One, etc, and also has extensive support for virtual reality peripherals such as Oculus Rift, HTC Vive/SteamVR and Daydream.

Besides, Unreal Engine 4 incorporates a VR Editor, which allows you to design 3D scenarios using virtual reality headsets and therefore design virtual reality from virtual reality.

However, both the documentation and community of Unreal Engine 4 is smaller, and its Asset Store is much smaller and more expensive than Unity3D, which is a higher handicap.

3) Blender

Blender, on the other hand, is a software designed for modeling, lighting, rendering, animation, video editing and creation of virtual scenarios that also has its own physics and scripting engine, which allows the creation of video games.

It's free software since 2002, under GPLv2 license, and is widely used by the 3D design community, and maintained by numerous volunteer developers, led by its creator Ton Roosendaal.

It has extensive documentation, a multitude of tutorials and books, but very focused on modeling and design in general, but this, in general, leaves aside the development of video games, since it is not its main purpose, and its use, with this aim, is not very widespread.

Support for virtual reality devices is in development and partially supported naively, having to resort to external plugins or blender adaptations, such as BlenderVR, to achieve somewhat more robust support. Even so, its use is primarily intended for editing and modeling in virtual reality.

TABLE II Comparative Advantages and Disadvantages of Analyzed Framework: Unity3D, Unreal Engine 4, Blender and Amazon Sumerian

Frameworks	Advantages and disadvantages
Unity 3D	It's not free software but it does have free access to the community. It's got a great community behind it. Supports all headsets studied. Exports to different platforms: Android, iOS, Windows, Linux, etc. The IDE can be extended through a specific API for plugin creation and support.
Unreal Engine 4	It has the most powerful rendering engine. Designed for video games. It incorporates a VR Editor to design the VR within the VR. Little documentation and community.
Blender	It's free software. It has smaller capabilities for application and video game development. The VR spectacle holder is still under development.
Amazon Sumerian	Applications hosted on Amazon. Pay per use. This tool has not yet been accessible as it's in beta phase.

4) Amazon Sumerian

In the final phase of the project, it has been detected that Amazon has acquired a new environment for creating and executing 3D virtual reality and augmented reality applications, called **Amazon Sumerian**, whose use for creation is free.

Applications are hosted on Amazon servers, so you pay for them from a certain size and from a certain number of plays per month. However, this platform is still in a preliminary version, and can only be used by requesting a trial of this version from Amazon.

5) Comparative

In Table II, a comparison of the different frameworks analyzed for PRACTICA is shown, referring to the advantages and disadvantages they have according to the needs of the project.

III. Conception of the Project

One of the most important decisions to carry out the PRACTICA project was to decide which Hardware and Software would eventually be used in the development of the project. To make the decision it was necessary to take into account that the development of the project required working in three-dimensional scenarios, both for the design of the training activities and for their execution or reproduction. Therefore, it was necessary to use some kind of basic 3D engine to give this three-dimensional support and create the PRACTICA tool as a layer on this engine.

To this end, as already indicated, four engines were studied which, due to their technical characteristics, acceptance in the world of 3D development and technical support, could fit into the development of the project: Unity3D, Unreal Engine, Blender and Amazon Sumerian.

Finally, it was determined that Unity3D is the one that best adapts to the needs of the project, since, despite not being open source, its use is free as long as it does not exceed a certain amount of revenue, it has the strongest community, more documentation, more assets (most of them created by the community), exports to a greater number of platforms, and supports more Hardware, mainly virtual reality, than the other two engines. NUnreal Engine has a much more professional profile and, despite being a better graphics engine, suffers from less documentation and community, and support for virtual reality devices is less. Blender is open source, however, has less support for exporting to different platforms and virtual reality devices.

Therefore, Unity3D is better suited to the needs of device support, and provides greater export possibilities to different platforms, such as Android, Windows and even PlayStation, which support Virtual Reality devices and headsets, and on which the PRACTICA activity player client could be generated.

A. Main Elements of PRACTICA. Editor and Player.

PRACTICA is divided into a scene editor and an activity Player.

1) Editor

Analyzed the characteristics of Unity3D as a development platform of this project, it was detected that one of its qualities was the creation of extensions of it. For this reason, the PRACTICA Editor has been conceived as a plugin for Unity3D, focused on the creation of training activities by trainers, whose profile is more technical.

It will be the trainers themselves who will be in charge of creating the activities, starting from a common set of resources and objects, editing the 3D environment in which the student will move, preparing the events that will arise in response to the actions of the student, deciding which of them involve a beneficial or successful action, and which are a mistake in their performance. Once the creation is complete, the trainer will be able to share the activity in an online service to make it accessible in all the PRACTICA reproduction applications. Therefore, the student can access all these shared activities online using the PRACTICA Reproductor and, through the use of VR headsets, be evaluated in them, obtaining a result of success or failure depending on their performance in the 3D world they have created.

Fig. 1 shows an example of creating an exercise with a forklift truck. This Editor allows you to clearly and easily enter all the actions that can be performed on this element, in this case, the forklift truck. Fig. 1 shows the look of Unity3D with the PRACTICA Editor plugin in activity editing mode.

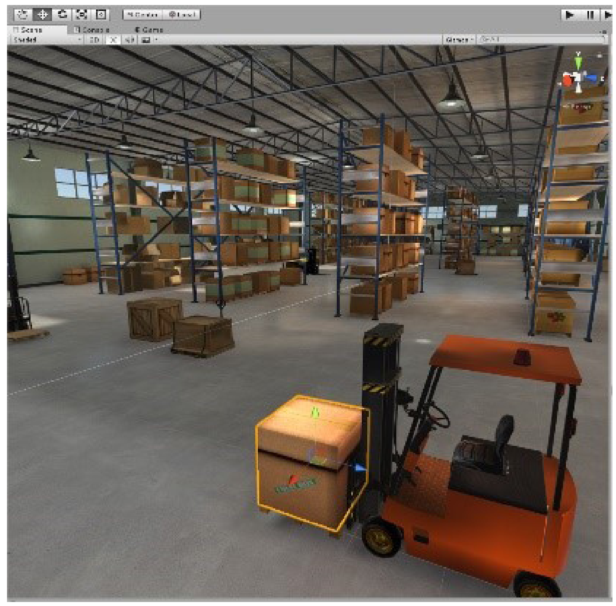


Fig. 1. Example of the creation of a scene for a forklift truck course in which all the elements of the scene can be selected and modified.

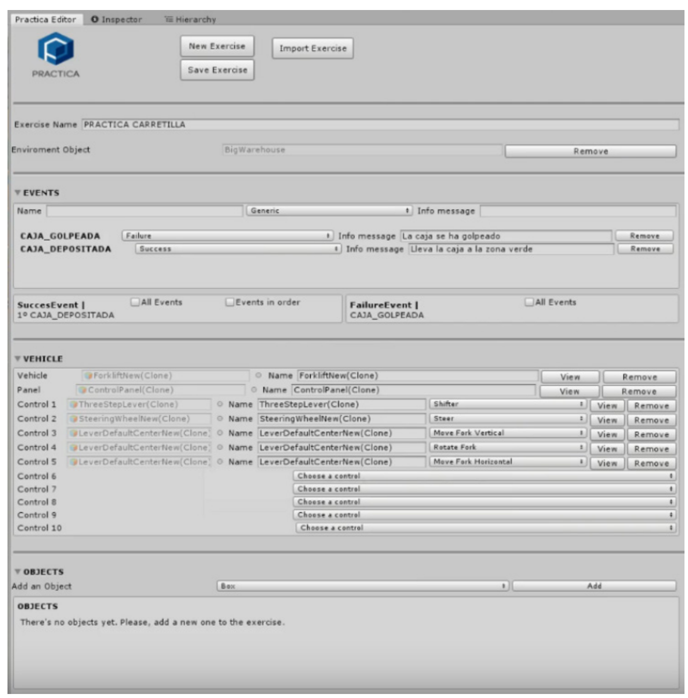


Fig. 2. Editor of elements of PRACTICA. From the editor you can configure the characteristics of the scene elements, manage events and determine the controls that the student will use during the course playback.

And in the image in Fig. 2, an example of the configuration of the Practica activity editor is shown, which shows the edition of an activity, with a scenario, success and failure events, a vehicle with its dashboard and functionalities, an object and a reactive object that triggers an event when it detects that it no longer touches the object.

2) Player

The PRACTICA Player is a stand-alone application independent of the Editor, as it shares only the resource and object package, which is common to both and accessible via a cloud service. These resource packages contain 3D objects, images, audios, environments, vehicles and activation zones, which are the basis for the creation of the activities, through the PRACTICA Editor, and with which the 3D world described in them will be created in the PRACTICA Player.

As they are hosted in an online service, they are not part of the playback application or plugin Editor, so they can be updated with more resources without having to redistribute a new version of the playback application or plugin Editor.

B. Architectural Design

Due to the 3D nature of the project and the need to modify the 3D environments in which the training activities are carried out, it has been necessary to have an Editor powerful enough to allow objects to be located, rotated and scaled in a 3D environment as faithful as possible to the final result perceived through the VR headsets. This led to the idea of taking advantage of Unity3D own scene Editor, as it is an unbeatable tool, specifically designed and developed for this purpose.

On the other hand, the fact that PRACTICA has been designed as a plugin, in addition to providing a simple tool for users who do not have technical knowledge, does not limit those users with deeper knowledge of Unity3D who want to take the editing of their activities a step further.

In addition, being an open source project, the distribution of PRACTICA as a plugin installable in Unity3D, is easier to deploy and put into operation, which can allow reaching a larger number of users and, therefore, potential partners.

Considering the above, and the fact that activity editing is a centralized and separate tool, technically and physically, from the reproduction of activities, as well as contributing to the decision to develop PRACTICA as an extension of Unity3D, led to two more technical decisions:

1) JSON as Activity Descriptor

On the one hand, the Editor should generate a description file of the training activity, which is simple to interpret and easy to share through cloud hosting.

With this premise in mind, the use of XML and JSON was evaluated, and the latter was finally adopted due to its simplicity, readability and suitability for transfer through web services.

Therefore, PRACTICA describes, using JSON, both the training activities it generates and the 3D elements available for its creation, such as scenarios, vehicles, objects, passageways, and other resources previously prepared and distributed, in the form of downloadable online packages.

That is to say, starting from a fixed, although expandable, set of 3D elements, whose functionality and properties would be described in a JSON catalog, PRACTICA had to be able to generate a JSON file describing a training activity, and its operation, being able to be shared online, making it available to the distributive part of PRACTICA in charge of the reproduction and execution of the activities.

2) 3D Resources Had to Be Accessible

Both for editing and reproduction, these being shared online, thus allowing new resources to be added without the need to update the PRACTICA software or redistribute it each time the catalog of objects available for training activities is extended. Therefore, a file format was required that would allow it to be served online, that could be readable by Unity3D, and where 3D models and audios, texts and behaviors typical of the engine could be packed.

As an additional feature, Unity3D implements a mechanism to share and update resources and content without the need to update and re-release new versions of the software. This functionality is provided by Unity3D through the AssetBundles: these are packages that can contain models, images, sounds and other Unity3D content, and can be shared online and downloaded later by the software (Fig. 3), allowing you to add or modify content to them, without having to update or modify the code.

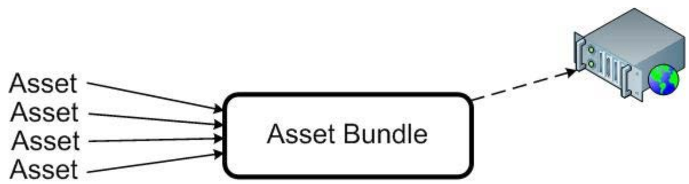


Fig. 3. Packaged resources into assets that can be easily redistributed later.

PRACTICA therefore makes use of them by adding, in addition to the resource itself, a catalog in JSON format, which describes the resources and tells the Editor what type they are and how they can be used when creating a training activity. PRACTICA provides a package (Fig. 4), common to the Editor and Player, where the elements that are part of the training activity are found, as well as the description of the type and its possible uses.

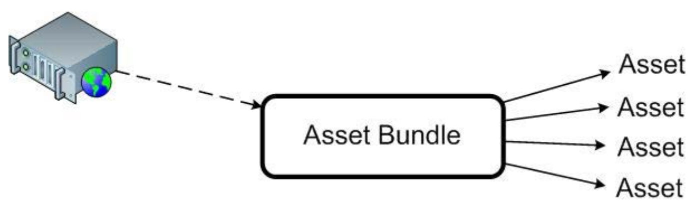


Fig. 4. PRACTICA encapsulates all the resources, descriptors and scenes to make course playback as easy as possible.

In this way, resources such as environments, objects, vehicles and controls are shared through these AssetBundles and described in the JSON catalog, where, for example, the functionalities of the vehicles, how they handle the controls and what type they are, allowing the Editor to link objects to each other in order to shoot them and make them work together, without the Editor being aware of whether he has more or less content, being able to increase or update it, without the Editor and the Player really being aware of it.

C. Search for Paths for NPC

Elements of artificial intelligence have been incorporated for the recreation of scenes that allow the incorporation of fictitious characters, also known as *non-player characters* (NPC) that allow the recreation of situations in which another person could be put at risk by doing an exercise. To achieve this, it has been necessary to analyze different techniques [19] [20] that allow these characters to move around the stage fulfilling a series of criteria or heuristics so that their movement makes sense within the scene [21].

Taking advantage of the possibilities that Unity3D currently offers, it has been decided to use a tool called *Navigation and Pathfinding* [22] as a basis for searching for paths and managing the behavior of these characters. This tool allows you to establish the criteria for a character to move intelligently within the stage using a mesh configuration that will establish the areas of the stage through which the character can move.

The first thing to define is the navigation area through which the fictitious persons will move. This is called the *NavMesh* or *navigation mesh*. We'll have to define a data structure that will contain information about the areas that will be walkable by these characters. In this way, it will be limited to our criteria the places through which they can pass. The next step will be to find the paths between the different points that make up the navigation grid. To do this, the navigation grid is divided into different polygons that, as the character progresses, will be visited. In this way, we'll know at each moment which zones the character has passed through and that he will advance to the next marked zone on the route without returning along the same path.

This tool has allowed us to configure characters with a certain behavior in the face of obstacles and to define their interaction with elements of the scene, such as opening a door. Three types of basic elements have been used: *NavMesh Agent*, *NavMesh Obstacle* and *Off-Mesh Link* [23].

On the one hand, *NavMesh Agent* is a component that will help define the behavior of an NPC or fictional person within our scene. This component allows us to configure criteria such as the character's movement, speed and behavior in the face of obstacles. In this project, basic behaviors have been programmed, such as, for example, that if a driver passes by with a forklift they will try to avoid it, or they will stop according to the distance. These behaviors may become more complex, but for the exercises defined during the prototype no higher level of behavioral complexity has been required.

On the other hand, *NavMesh* Obstacle refers to dynamic obstacles that may be present in the scene. This allows us to define more realistic situations, where we can find boxes on the floor that are hindering the passage. This type of object will have a behavior that we will determine according to its nature. *NavMesh Agents* can interact with these objects in the way that best suits the conditions of the scenario, for example, if there is a box on the ground, we can try to avoid it by changing course so as not to collide with it.

Finally, *Off-Mesh Link* allows *NavMesh Agents* to interact with each other to make movements between unconnected points within the *navigation mesh*, such as jumping over a wall or opening a door.

Therefore, we can give these characters and objects the behavior we want according to the needs of the exercise to be developed.

D. Virtual Reality Toolkit (VRTK)

Currently, VR headsets SDK, such as Oculus Rift, don't provide support for the high-level interactions required for the control of objects within a scene, and are limited to more basic interactions, leaving the programming of more elaborate interactions to the developers.

On the other hand, the SDK of the different VR headsets are not compatible with each other; that is to say, in order to give support in PRACTICA to several headsets, such as Oculus Rift, HTC Vive and Daydream, it would be necessary to make a specific programming for each one of them, with the cost that this entails taking into account the development particularities of each one of them.

Taking these two factors into account, it became necessary to use a library that would act as a unified interface for all virtual reality devices, and at the same time provide a higher level of management of user interaction in the virtual world.

The answer to these needs is the VRTK library [24], which provides unified support for different brands of virtual reality devices, such as Daydream, Oculus Rift, Steam VR and Ximmerse, and also supports much higher-level functionality than that provided by the SDK themselves.

Thanks to this, PRACTICA supports the VR Oculus Rift, HTC Vive, Samsung Gear VR, Daydream and Ximmerse devices; it also includes controls that require very detailed and precise manual movements by the user in the virtual world, which results in greater realism, and practically similar interactions to those that would be done in reality.

IV. Results Obtained During the Pilot

The results of the project have been evaluated together with a training company, *Proyectos Integrales Clave*, which helped to define the needs of simulating the driving of a forklift truck so that the students can obtain their operator's licence.

The platform allows you to learn how to use a"tool" / "machinery" and to put it in situations of risk that could never be learned in a traditional training. For example, that the forklift demo has individuals walking, or that a box falls nearby.

In PRACTICA, we are faced with the challenge of starting from an empty 3D scene and generating it in playback time following a script previously generated using another of the project's components and starting from a set of pre-defined and pre-packaged resources.

During the development of this prototype, we found that the teachers who received the courses were satisfied with the experience and requested that they continue to be offered more courses of this nature, proposing future extensions to other areas that they could offer as part of their catalog of courses and activities, either independently or as a complement to existing ones.

From a practical and application of a methodological approach, the objective of these sessions was both the practice of the exercises already carried out (Fig. 5) and that the participant build educational or training solutions to situations and/or needs of the classroom through creative proposals, through the use of the didactic potential of virtual reality, through a PC, due to its universality and adaptation to the characteristics of these potential educational tools.

The course uses specific resources for each type of training, in this case, a footswitch has been adapted to adjust the feeling of the students and give more realism to the exercise used in the proposal (pilot) and resolution of activities and tasks supervised and oriented to achieve the objectives of the course and the development of skills of participants.

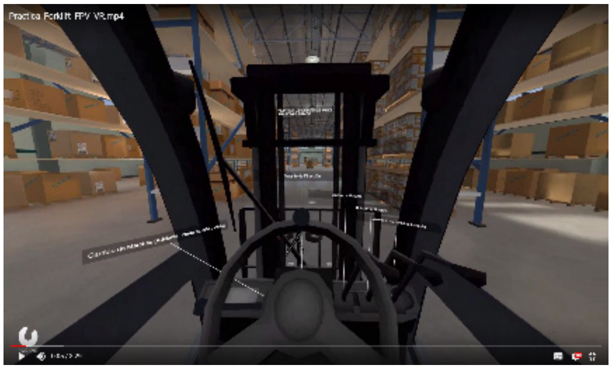


Fig. 5. Visualization of the scene from the student's point of view during the exercise.

V. Conclusion

The main objective of this project is to offer entities that require practical training for their students or workers, and that for reasons of risk or difficulty of access to materials or cost is not within their reach, the use of a simulator that will provide added value to the content already offered.

In the case of the forklift driving course, companies, when offering training and work experience to their students, cannot run the risk of putting people in a warehouse to check whether the student will be able to avoid them and not run over them. However, in a virtual simulation, these situations can be offered to better evaluate the response of students to this type of incident.

On the other hand, the realism and immersion provided by the virtual reality environment, through the use of headsets and virtual reality controls, offers an extremely realistic simulation, providing students with training comparable to the real thing. This allows students to be trained in complex skills or machinery, which are not available to trainers, are particularly dangerous to use or have very limited time for use.

Thanks to this level of realism, provided by the VR, the student who trains with PRACTICA will find a great similarity with what he has practiced when he is in a situation or manipulating real machinery, as will have acquired habits and mechanized movements that correspond one by one with the real world. This also eliminates the initial blockage, or fear of the unknown, in risky or dangerous situations, since, although in a virtual environment, they have already been in a similar situation and will know how to act.

Another added value of this platform is the way it is presented to end users, who can create exercises or share them with other users, without the need to have knowledge of software programming, since the PRACTICA Editor has been built to offer all the options regarding objects, scenes and controls, so that their configuration is as clear and intuitive for the trainer.

In addition, the components with which the exercises are prepared have been designed to encourage reuse and, above all, future expansion in number and functionality, without the need for further development or updating of the platform. Therefore, the prototype for the simulation of a mechanical forklift could be, in the future and without much effort, a crane simulator, a fire extinguisher simulator or any other training activity.

Acknowledgment

This Project has been financed thanks to the Program of Incentives to the Knowledge System Agents of the Andalusian (call 2012), which depends on the Ministry of Economy and Knowledge of the Junta de Andalucía. The project code is PI12 / 1166.

References

[1] Miller, M. (2014) Virtual Reality: The Next Big Thing in Education? *Product Lifecycle Report. PTC blog*. Available from:http://blogs.ptc.com/2014/08/13/virtual-reality-the-next-big-thing-in-education/

[2] Valentín L-, Carrasco A, Konya K, Burgos D. Emerging Technologies Landscape on Education. A review. *International Journal of Interactive Multimedia and Artificial Intelligence*. 2013;2 (Special Issue on Improvements in Information Systems and Technologies): 55-70.

[3] Greenwald, S., Kulik, A., Kunert, A. (2017) Technology and applications for collaborative learning in virtual reality. *In CSCL, 12th International Conference on Computer Supported Collaborative Learning (CSCL), Pennsylvania, USA, 18 - 22 June 2017*. Philadelphia, PA, USA: International Society of the Learning Sciences, pp. 719-726.

[4] Hu, B.; Ma, LA., Zhang, W., Salvendy, G., Chablat, D., Bennis, F. (2011) Predicting real-world ergonomic measurements by simulation in a virtual environment. *International Journal of Industrial Ergonomics*. Volume: 41 Issue: 1 Pages: 64-71.

[5] Interrante, V., Ries, B., Lindquist, J., Kaeding, M., and Anderson, L. (2008). Elucidating factors that can facilitate veridical spatial perception in immersive virtual environments. *Presence: Teleoperators and Virtual Environments*, 17(2):176–198.

[6] Google VR Cardboard. Available in: https://vr.google.com/cardboard/

[7] DayDream. Available in: https://vr.google.com/daydream/

[8] A short history of Oculus Rift. Available in: http://www.tomshardware.com/reviews/ar-vr-technology-discussion,3811-5.html

[9] Oculus Rift. Available in: https://www.oculus.com/rift

[10] Leap Motion. Available in: https://www.leapmotion.com/

[11] Kinect para Windows. Available in: https://developer.microsoft.com/es-es/windows/kinect

[12] How HTC and Valve built the Vive. Available in: https://www.engadget.com/2016/03/18/htc-vive-an-oral-history/

[13] Playstation VR Technical Specifications. Available in: https://www.playstation.com/es-es/explore/playstation-vr/tech-specs/

[14] Unity3D Documentation. Available in: https://docs.unity3d.com/es/current/Manual/index.html

[15] Features of Unreal Engine 4. Available in: https://www.unrealengine.com/en-US/features

[16] Features of Blender. Available in: https://www.blender.org/features/

[17] Amazon Sumerian. Available in: https://aws.amazon.com/es/blogs/aws/launch-presenting-amazon-sumerian/

[18] Asset Bundles de Unity3D. Available in: https://docs.unity3d.com/es/current/Manual/AssetBundlesIntro.html

[19] S. M. Lucas, ‘Computational intelligence and AI in games: A New IEEE transactions’, *IEEE Trans. Comput. Intell. AI Games*, vol. 1, no. 1, pp. 1–3, 2009.

[20] Pandit, D. (2017) 3D Pathfinding and Collision Avoidance Using Uneven Search-space Quantization and Visual Cone Search. ResearchGate, article. Available on: https://www.researchgate.net/publication/317356810_3D_Pathfinding_and_Collision_Avoidance_Using_Uneven_Search-space_Quantization_and_Visual_Cone_Search

[21] Khantanapoka, Khammapun & Chinnasarn, Krisana. (2009). Pathfinding of 2D & 3D Game Real-Time Strategy with Depth Direction A*Algorithm for Multi-Layer. 184 - 188. 10.1109/SNLP.2009.5340922.

[22] Unity3D. Navigation and Pathfinding. Available in: https://docs.unity3d.com/es/current/Manual/Navigation.html

[23] Unity3D. Navigation and Pathfinding. Navigation reference: Available in:https://docs.unity3d.com/es/current/Manual/class-NavMeshAgent.html

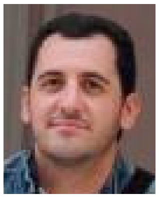
[24] VRTK – Virtual Reality Toolkit. Available in: https://vrtoolkit.readme.io/

Juan Manuel Lombardo

PhD in Computer Science from the Pontifical University of Salamanca, was graduated in Economics and Business Administration in the University of Granada, Spain, Diploma of Advanced Studies (DEA) in Economics from UNED, Research Sufficiency in Business Science from the Complutense University of Madrid and Diploma of Advanced Studies (DEA) in Sociology from the Pontifical University of Salamanca. He is CEO at Fidesol and Professor at Andalusia Business School. Dr. Lombardo is the author of numerous articles and research papers published in journals and books of national and international conferences. Visiting Professor at the Private Technical University of Loja (UTPL Ecuador), The National University of the Northeast (Argentina), University Francisco José de Caldas (Colombia), Catholic University of Colombia, Catholic University of Ibarra (Ecuador), University of Lisbon (Portugal) and National Engineering University (Peru). Member of the Knowledge Management committee of AEC (Spanish Association for Quality) and the Institute CICTES (Ibero-American Centre on Science, Technology and Society).

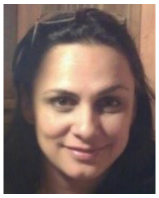
Miguel Ángel López

He has a degree in Engineering in Technical Engineering in Computer Systems from the University of Almería, graduates in Computer Engineering and Master in Soft computing and Intelligent Systems from the University of Granada. Currently he is CTO. at Fidesol where performs different roles on the projects. His research focuses on distributed systems, management, integration and analysis of data, robotics, fuzzy logic systems, and the development of virtual reality environments for different purposes.



Vicente M. García

He has a degree in Technical Engineering in Computer Systems from the University of Catalunya. Currently he is a researcher at Fidesol where performs the role of programmer analyst on frontend and fullstack software projects. He is an experienced Unity3D and VR developer. He has developed various games in Unity3D and his final degree thesis was a VR Cinema and 360 degrees Video Player for Android smartphones using Google Cardboard technology.



Mabel López

She has a degree of Computer Science Engineering. She is Knowledge Manager at Fidesol. Participates in the research and development strategy of this entity, technology transfer and analysis of technological trends, such as big data, internet of things, virtual reality, cognitive engines, machine learning, etc. Currently, she is involved in several R & D projects related to the mentioned technologies.

Ruben Jose Cañadas Sastre

Master's degree after graduating from the University of Almería in computer science. Currently hold the position of Project Manager in FIDESOL where is in charge of the fit gap analysis, project planning, budgeting, customer requirements, client advise and offshore training for several public and private projects.

Susana Velasco

She has a Technical Engineer in Computer from the University of Granada. In the past, she worked in manufacturing, financial and service sector enterprises as software engineer and analyst programmer. Her research interests include quality assurance, software quality management systems, Ambient Intelligence (AmI) systems and devices, and new generation of ICT technologies.

Mónica León

Degree in Documentation from the University of Extremadura and Diploma in Library Science from the University of Granada. He is currently a technician in the R & D Transfer Office of the Free Software R & D Foundation, performing tasks of technology surveillance, coordination and management of proposals for public incentives for R & D projects, throughout the chain of development of the projects -from the call to the final justification-, and the search, constitution and coordination of partners and consortiums for collaborative projects.

Multilevel Thresholding for Image Segmentation Using an Improved Electromagnetism Optimization Algorithm

Ashraf M. Hemeida[1] *, Radwa Mansour[2], M. E. Hussein[2]

[1] Electrical Engineering Department, Faculty of Energy Engineering, Aswan University (Egypt)
[2] Computer Science Division, Department of Mathematics Faculty of Science, Aswan University (Egypt)

Received 11 March 2018 | Accepted 3 September 2018 | Published 22 September 2018

Abstract

Image segmentation is considered one of the most important tasks in image processing, which has several applications in different areas such as; industry agriculture, medicine, etc. In this paper, we develop the electromagnetic optimization (EMO) algorithm based on levy function, EMO-levy, to enhance the EMO performance for determining the optimal multi-level thresholding of image segmentation. In general, EMO simulates the mechanism of attraction and repulsion between charges to develop the individuals of a population. EMO takes random samples from search space within the histogram of image, where, each sample represents each particle in EMO. The quality of each particle is assessed based on Otsu's or Kapur objective function value. The solutions are updated using EMO operators until determine the optimal objective functions. Finally, this approach produces segmented images with optimal values for the threshold and a few number of iterations. The proposed technique is validated using different standard test images. Experimental results prove the effectiveness and superiority of the proposed algorithm for image segmentation compared with well-known optimization methods.

Keywords

Image Segmentation, Multilevel Thresholding, Otsu's Entropy, Electromagnetic Optimization, Levy Function.

DOI: 10.9781/ijimai.2018.09.001

I. Introduction

Recently, image processing has several applications in different areas such as; industry agriculture, medicine, etc. Image segmentation is considered the most important step in image processing [1]. This process is classifying the pixel in the image depending on its intensity value. In literature, different methods have been proposed for image segmentation, including edge based method [2] neural network based method [3],watershed based method [4], clustering based method [5], and artificial threshold based method [6].

Thresholding method is a simple and effective tool to isolate objects of interest from the background. Its applications include several classics such as document image analysis, whose goal is to extract printed characters [7] logos, graphical content, or musical scores; also it is used for map processing which aims to locate lines, legends, and characters [8]. It is also used for scene processing, aiming for object detection and marking [9];; Similarly, it has been employed to quality inspection for materials discarding defective parts[10].

Thresholding techniques are used for segmenting the image into two (bi-level) or more classes (RGB). The binary level thresholding is taking only one threshold value (t) and then testing every pixel with specific intensity value, if it is higher, the threshold value (t) classified as the first class and the other pixel with a different intensity value are classified as second class. In multilevel thresholding, the pixels in the image are divided into more than one class, where, every class is taken a specific threshold value [11-13]. Basically, two approaches called parametric and nonparametric can be used to determine the optimal threshold value [12]. In parametric approach, some parameters of a probability density function should be estimated for classifying the classes of image. But this approach is computationally expensive and time consuming whereas that non parametric approach optimizes several criteria such as; the error rate, the entropy, etc, in order to determine the optimal threshold values However, there are two methods can be used for binary level thresholding; Otsu's and Kapur. Otsu method uses the maximization of between classes variance [14].

* Corresponding author.
E-mail address: ashraf@aswu.edu.eg

Kapur method maximizes the entropy to measure the homogeneity of the classes [15]. The computational complexity of the two methods for multilevel thresholding is increasing for each new threshold [16].

Many optimization techniques deal with multilevel thresholding (MT) for image segmentation such as; Particle Swarm Optimization (PSO) [17] , Moth Flame Optimization (MFO) [18], Genetic Algorithm (GA) [19], Whale Optimization Algorithm (WOA) [18], Ant Colony Optimization (ACO) [20].

Recently, EMO contributed to solving many engineering problems such as; control systems [22], array pattern optimization in circuits [23],neural network training [24], vehicle routing [25], communications [26], flow-shop scheduling [27], image processing [28] Although EMO algorithm is consistent with characteristics of other approaches, EMO's performance generates a better precision and computation time in most of the studied cases when it is compared with other metaheuristic optimization techniques such as; Cuckoo Search (CS), Sine Cosine algorithm (SCA) and MFO, and WOA. However, EMO algorithm simulates the mechanism of attraction-repulsion which subject to the electromagnetic law of physics to develop the individuals of population based on their objective function values [21]. The main idea of EMO is to move a particle through the space following the force exerted by the rest of the population. Using the charge of each particle depending on its objective function value, the force is calculated.

This paper proposes an improved EMO algorithm based on levy function (EMO-levy) for multilevel thresholding of image segmentation. Different standard images are used to validate the proposed technique compared with other well-known optimization techniques such as; original EMO, CS, SCA, MFO and WOA. The developed approach produces a multilevel segmentation of digital images with few number of iterations and fast computational time.

The organization of the paper is as follows. In Section II, presents the developed Electromagnetism Optimization Algorithm with levy function (EMO.levy). Section III illustrates the problem definition of multilevel thresholding (MT) and the application of developed optimization algorithm for solving it. Section IV presents comprehensive results obtained by the developed algorithm, comparing with other techniques. Section V provides the outstanding features of the proposed algorithm. Finally, Section VI gives the final conclusions of the paper.

II. Improved Electromagnetism Optimization Algorithm

A. EMO

The EMO formulation has been developed to find a global solution of a nonlinear optimization problem with satisfying the operating constraints as [12]:

$$\max f(x), x = (x_1, x_2, x_3, ..x_k) \in Z \tag{1}$$

Subject to $x \in X$

where, f: Z→Z is a nonlinear function, $X = \{x \in Z | l_k \le x_k \le u_k, k = 1, \dots, n\}$, l_k is the lower limit and u_k is the upper limit.

EMO technique uses some variable such as:

m → number of population inside one iteration

n → dimensional control variable $x_{j,t}$ and t refers to the iteration number of the algorithm.

$E_t=(E_{1,t},E_{2,t},E_{3,t},\dots\dots\dots\dots E_{m,t})$→is the initial population in the same iteration t which is taken by random samples from search space X of image histogram. After the initialization of E_t, EMO continues its iterative process until reaching to the maximum number of iterations [30].

B. Lévy Electromagnetism Optimization Algorithm (EMO.levy)

Lévy flight distribution is integrated in EMO technique to enhance the searching capability and exploration ability of this optimization algorithm by increasing its probability of producing new solutions to avoid stagnation of algorithm and to avoid trapping in local minima. Levy flight is a random process for generating a new solution based on random walk where its steps are captured from a Lévy distribution. The new population position that is based on Lévy distribution can be found as follows [35]:

$$X_i^{new} = X_i + \propto \oplus Levy(\beta) \tag{2}$$

where, ∝ represents a random step size parameter. ⊕ is the entry wise multiplication and β is a Lévy Flight distribution parameter. The step size can be found as:

$$\propto \oplus Levy(\beta) \sim 0.01 \frac{u}{|v|^{1/\beta}} \left(X_i^t - X_{best}^t \right) \tag{3}$$

where, u and v are normally variables produced by normal distribution where,

$$u \sim N\left(0, \phi_u^2\right), \; v \sim N\left(0, \phi_v^2\right) \tag{4}$$

$$\phi_u = \left[\frac{\Gamma(1+\beta) \times \sin(\pi \times \beta / 2)}{\Gamma[(1+\beta)/2] \times \beta} \right]^{1/\beta}, \phi_v = 1 \tag{5}$$

where Γ, is the standard gamma function and, $0 \le \beta \le 2$. To enhance the exploitation of EMO the best search agent is updated by using variable bandwidth as follows:

$$X_{best}^{new} = X_{best}^t \pm C_5 \times K_w \tag{6}$$

where, C_5 is a random number in [0, 1]. K_w is a variable bandwidth that decreases dynamically as:

$$K_w = K_{\max} e^{(s \times t)} \tag{7}$$

$$s = \left(\frac{\ln\left(\frac{K_{\min}}{K_{\max}}\right)}{T_{\max}} \right) \tag{8}$$

where, K_{max} and K_{min} are the maximum and the minimum limits. t is the current iteration and T_{max} is the maximum number of iterations.

However, the solution process of EMO technique is going through basic steps:

Step1- Input variable $iter_{max}$, $iter_{local}$, α local search parameter, m number of population, n dimensional of x.

Step2- The algorithm in this step makes initialization by taking random particles from the search space between lower and upper bound at same iteration=1 ,then the objective function values

f $(x_{i,t})$ are calculated over all population $E_{i,t}$, (where i=1,….,m) from the results, the best value of $E_{i,t}$ is E^b that represents the optimum objective function value $f(x_{i,t})$ which is generated from the best point x_t^b.

where,

$$E_{i,t} = f(x_{i,t}) \tag{9}$$

$$x_t^b = \arg\max f(x_{i,t}) \tag{10}$$

$$E_t^b = f(x_t^b) \tag{11}$$

Step3- The force value must be calculated for each individual of the population in E_t, It is known that each individual of the population has a value of objective function $f(x_{i,t})$ therefore a value the electromagnetic charge $(q_{i,t})$ can be calculated based on the value of its function as:

$$q_{i,t} = \exp\left(-n \frac{E_{i,t} - E_t^b}{\sum_{i=1}^{m} E_{i,t} - E_t^b} \right) \tag{12}$$

Simply the force between two points $x_{i,t}$, $x_{j,t}$ (where i≠j) can be calculated as:

$$F_{i,j}^{t} = \begin{cases} \left(x_{i,t} - x_{j,t}\right) \dfrac{q_{i,t} \bullet q_{j,t}}{\left\| x_{j,t} - x_{i,t} \right\|^2}, & if\ f\left(x_{i,t}\right) > f\left(x_{j,t}\right) \\ \left(x_{j,t} - x_{i,t}\right) \dfrac{q_{i,t} \bullet q_{j,t}}{\left\| x_{j,t} - x_{i,t} \right\|^2}, & if\ f\left(x_{i,t}\right) \le f\left(x_{j,t}\right) \end{cases} \quad (13)$$

The total force that affects the point xi,t can be calculated as:

$$F^T = \sum_{j=1, i\neq j}^{m} F_{i,j}^{t} \quad (14)$$

In this step, each point $x_{i,t}$ except x_b is moving from its place to another using attraction or repulsion force depending on its charge which is based on objective function value of each other. Hence, the points which have strong charge with a better objective function value will be attracted to each other, and the point $x_{i,t}$ will move to point $a_{i,t}$

$$x_{i,t} = x_{i,t} + \frac{F_i^t}{F_i^t}\left(RNG\right), i = 1, 2, , \ldots..m \quad i \neq b \quad (15)$$

where, RNG → the range of movement toward the lower or upper bound.

1 → the standard uniform distribution with minimum 0 and maximum 1.

Step4- In this step, the algorithm makes local search. After all points are moved from its place except x^b point on search space, the iteration of local search ($iter_{local}$) is generated and each point performs a local search for neighbor points of $a_{i,t}$ and generates its δ neighbor points for selecting a better point $d_{i,t}$ which generates a better objective function more than the current. Then, the local search process will be stopped when the iteration ($iter_{local}$) is finished.

After all these steps, the algorithm goes to next iteration, while the point $x_{i,t}$ becomes either $a_{i,t}$ or $d_{i,t}$ in next iteration (t+1).

III. Image Segmentation Using Improved Emo Method

The effective way for image segmentation is thresholding method. This method is used for classifying the binary level image and multilevel image based on the value of its intensity level (L). This process converts the image into (m×n) pixel. Each pixel carries an intensity level (L) value that can be classified to the class which it belongs.

If the image is grey level, the thresholding method classifies it into two classes R_1, R_2 with only a threshold value (th), hence, if the pixel has an intensity level value > th, it can be classified as the first class R_1 and otherwise classified to second class R_2.

$$\begin{aligned} R_1 &\leftarrow p \quad if \quad 0 < p < th \\ R_2 &\leftarrow p \quad if \quad th < p < L-1 \end{aligned} \quad (16)$$

For multilayer image, the image is divided into more than two classes, every class has a specific threshold value, Therefore for multilayer image TH=(th_1,th_2,th_3,……….th_{L-1}), the classes are (R_1,R_2,….R_N).

where, N is the number of classes

$$\begin{aligned} R_1 &\leftarrow p \quad if \quad 0 < p < th_1 \\ R_2 &\leftarrow p \quad if \quad p > th_1 \\ R_3 &\leftarrow p \quad if \quad p > th_2 \\ &\ldots\ldots\ldots\ldots. \\ R_N &\leftarrow p \quad if \quad p > th_{N-1} \end{aligned} \quad (17)$$

The problem for both bi-level and multilevel thresholding is to select the *th* values which correctly identify the classes. Otsu's and Kapur's methods are well-known approaches for determining such values. Both methods propose a different objective function which must be maximized in order to find optimal threshold values, just as it is discussed below.

A. Otsu's Method

Otsu method is one of the methods used to segment the image by maximizing variance value σ among classes then calculating the value of objective function as:

$$f\ (th)_{otsu} = \max(\sigma_b^{2^r}\ (\mathrm{th})) \quad (18)$$

where, $0 \leq \mathrm{th} \leq \mathrm{L} - 1$

The optimization problem is decreased to get a better intensity level (threshold) that maximizes (18). The previous objective function is used for grey level image because it contains one threshold (th). However, equation (18) can be rewritten to be used in the case of RGB images as;

$$f\ (TH)_{otsu} = \max(\sigma_b^{2^r}\ (\mathrm{TH})) \quad (19)$$

where,

$0 < th_i < \mathrm{L} - 1 = 1, .., \mathrm{k}$

TH=($th_1, th_2, \ldots.. th_{k-1}$) and k is number of class

$$\sigma_B^{2^r} = \sum_{i=1}^{k} \sigma_i^r \quad (20)$$

$$\sigma_i^r = w_i^r (m_i^r - m_t^r)^2 \quad (21)$$

where,

i→ refers to a certain class, and k is number of classes

r→ refers to a constant number equal to 1 in grey level image (r=1,2,3 in RGB image)

σ_i^r → refers to the variance among classes R (Otsu´s variance)

m_i^r →refers to the mean of a class

$$\begin{aligned} m_0^r &= \sum_{i=1}^{th_1} \frac{iph_i^r}{w_0^r(th_1)} \\ m_1^r &= \sum_{i=th_1+1}^{th_2} \frac{iph_i^r}{w_1^r(th_2)} \\ &\ldots\ldots. \\ m_{k-1}^r &= \sum_{i=th_k+1}^{L} \frac{iph_i^r}{w_1^r(th_k)} \end{aligned} \quad (22)$$

where, w_1^r → refers to the probability of occurrence

$$\begin{aligned} w_1^r(th) &= \sum_{i=1}^{th_1} ph_i^r \\ w_2^r\left(th\right) &= \sum_{i=th_1+1}^{th_2} ph_i^r \\ &\ldots.. \end{aligned}$$

$$w_{k-1}^{r}(th) = \sum_{i=th_k+1}^{L} ph_i^{r} \tag{23}$$

where, that ph_i^r is the probability distribution of the intensity level values which can be calculated as:

$$ph_i^{r} = \frac{h_i^{r}}{N}, \mathrm{r} = \begin{cases} 1 & in \quad grey\, image \\ 1,2,3 & in \quad RGBimage \end{cases} \tag{24}$$

where,

$$\sum_{i=1}^{N} ph_i^{r} = 1 \tag{25}$$

h_i^r are histogram distribution values which indicate the number of pixels corresponding to the i intensity level, N is total number of pixel in the image.

Electromagnetism optimization is used to find the optimal decision variable which is considered in the segmentation problem as threshold (TH)[20] as:

Maximize $f\ (TH)_{ostu}$

Subject to $\quad TH \in x,\ TH = (th_1, th_2, \ldots . th_k)$

where, $0 < th_i < 255$ this is lower and upper bounded of threshold i=1,..,k and here k refers to different thresholds (TH).

B. Kapur Method

Another nonparametric method that is used to determine the optimal threshold values has been proposed by Kapur [15]. It is based on the entropy and the probability distribution of the image histogram. This method aims to find the optimal *th* which maximizes the overall entropy. The entropy of an image measures the compactness and separability among classes. In this sense when the optimal *th* value appropriately separates the classes, the entropy has the maximum value. For the bi-level example the objective function of the Kapur's problem can be defined as:

$$f_{kapur}(th) = H_1^{c} + H_2^{c}, c = \begin{cases} 1,2,3 & if \;\; RGB\ image \\ 1 & if \;\; grey\, image \end{cases} \tag{26}$$

where, the entropies H_1 and H_2 are computed by the following model:

$$H_1^{C} = \sum_{i=1}^{th_1} \frac{ph_i^{c}}{w_0^{c}} \ln\left(\frac{ph_i^{c}}{w_0^{c}}\right), \quad H_2^{C} = \sum_{i=th+1}^{L} \frac{ph_i^{c}}{w_1^{c}} \ln\left(\frac{ph_i^{c}}{w_1^{c}}\right) \tag{27}$$

where, ph_i^c is the probability distribution of the intensity levels which is obtained using (24). w_0^c and w_1^c are probabilities distributions for C_1 and C_2, respectively. Similar to the Otsu's method, the entropy-based approach can be extended for multiple threshold values, for such a case it is necessary to divide the image into *k* classes using the similar number of thresholds. Under such conditions, the new objective function is defined as:

$$f_{kapur}(TH) = \sum_{i=1}^{k} H_i^{c}, \quad c = \begin{cases} 1,2,3 & if \;\; RGB\ \ Image \\ 1 & if \;\; grey\ Image \end{cases} \tag{28}$$

where,

$$\begin{aligned} H_1^{C} &= \sum_{i=1}^{th_1} \frac{ph_i^{c}}{w_0^{c}} \ln\left(\frac{ph_i^{c}}{w_0^{c}}\right) \\ H_2^{C} &= \sum_{i=th_1+1}^{th_2} \frac{ph_i^{c}}{w_1^{c}} \ln\left(\frac{ph_i^{c}}{w_1^{c}}\right) \\ &\ldots\ldots\ldots\ldots\ldots \\ H_k^{C} &= \sum_{i=th_k+1}^{L} \frac{ph_i^{c}}{w_{k-1}^{c}} \ln\left(\frac{ph_i^{c}}{w_{k-1}^{c}}\right) \end{aligned} \tag{29}$$

The solution process of image segmentation using the developed optimization algorithm can be shown in Fig. 1.

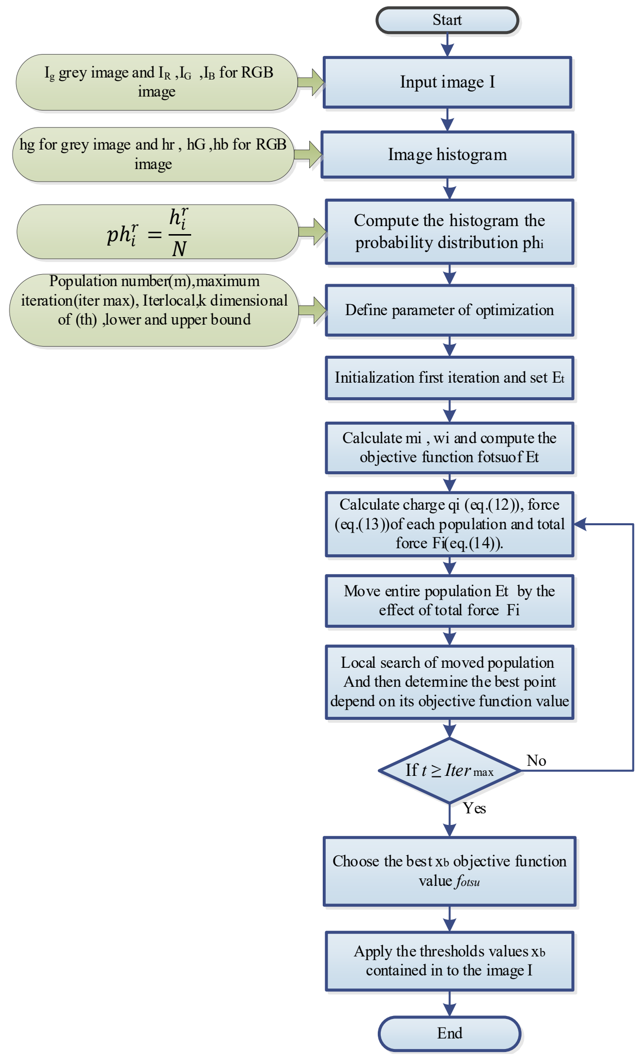


Fig. 1. Flowchart of the proposed method for image segmentation.

IV. Results and Discussion

In this section, different test images are used to validate the proposed EMO with levy algorithm

A. Standard Test Images

In this section, three test images; Baboon, Peppers, and Camera Man, are used to check the effectiveness of the optimization algorithm. These images are taken from USC-SIPI image database which are of size 512 ×512 each [36].

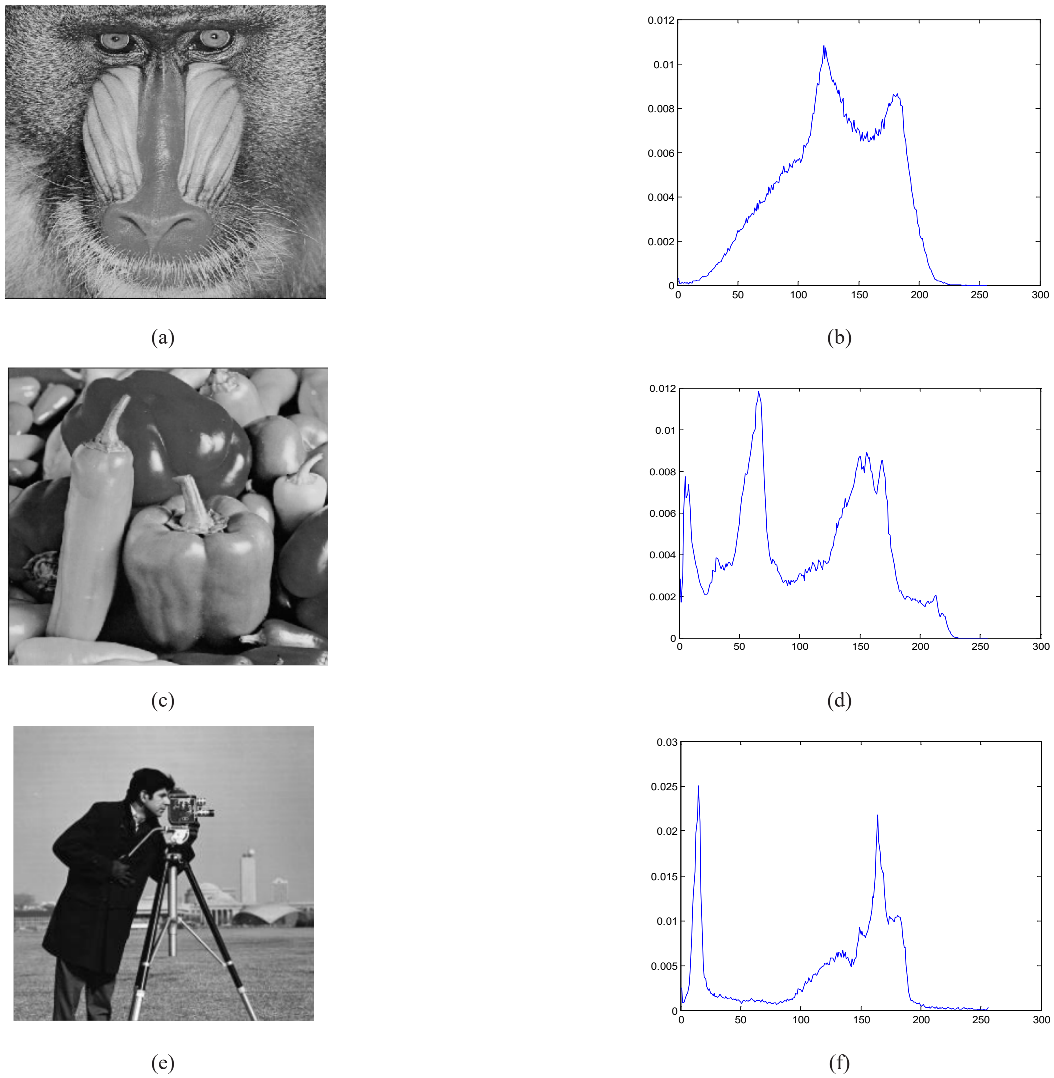


Fig. 2. Benchmark test images: (a) Baboon, (c) pepper, (e) Cameraman, (b), (d), (f) histogram of these images, respectively.

Fig. 2 indicates these test images and their histogram distributions which refer to the number of pixels in the images at each different intensity value found in these images.

The performance of proposed EMO with levy algorithm based multilevel thresholding is compared with some well-known optimization algorithms such as; CS [31, 32], MFO [18], WOA [18], SCA [33, 34]. All algorithms have been tested with the same condition of stop criterion;100 iterations and 25 population. At the end of each test, the Peak Signal-to-Noise Ratio (PSNR) is computed as:

$$PSNR = 20\log_{10}\left(\frac{255}{RMSE}\right) \tag{30}$$

The PSNR is an important value which measures the accuracy of a segment image comparing with the original image.

where, RMSE is the root mean square error, which can be calculated as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{ro}\sum_{j=1}^{co}\left(I_0^r(i,j)-I_{th}^r(i,j)\right)}{ro\times co}} \tag{31}$$

where, ro is the total number of rows of an image, co is the total number of column of the image, r is depending on the type of image (RGB image r=1,2,3),I_{th}^r refers to the segmented image, I_o^r refers to the original image.

In each experiment the stop criteria is set to 100 iterations. In order to verify the stability at the end of each test the standard deviation (STD) is obtained (Eq. (32)). If the STD value increases the algorithms becomes more instable.

$$STD = \sqrt{\sum_{i=1}^{iter_{max}}\frac{(s_i-m)}{Ru}} \tag{32}$$

Table I indicates the parameters of EMO with Levy function (EMO.levy). The maximum number of iterations is taken as 100, this parameter represents the stop criterion of the optimization process. However, which the stop criterion is taken as the number of times in which the best fitness values remains without change. $Iter_{local}$=100 is the number of times that the algorithm do local search with 25 population at every external iteration.

TABLE I. Parameters of Emo With Levy

Itermax	Iterlocal	d	M(population)
100	10	0.025	25

TABLE II. Results After Applying the Emo with Levy Using Otsu´S Function

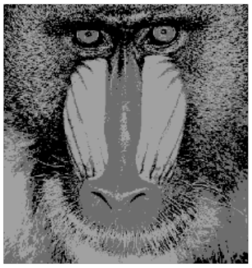
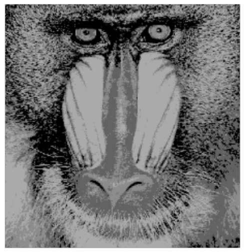
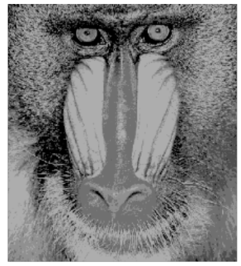
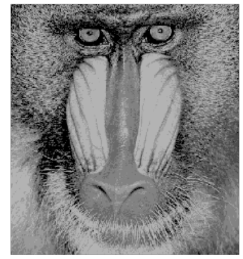
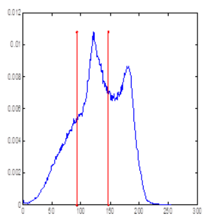
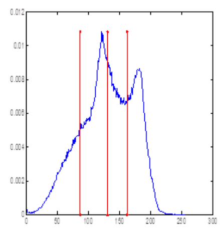
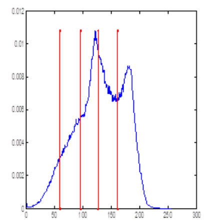
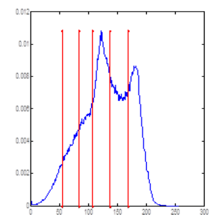
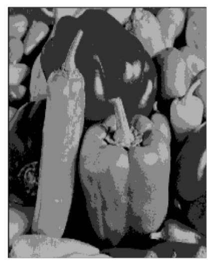
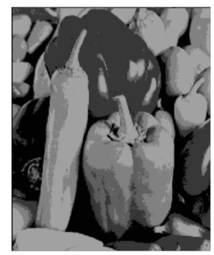
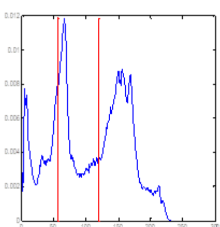
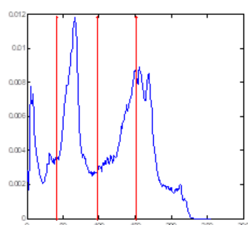
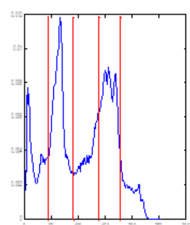
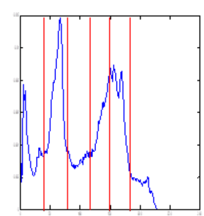
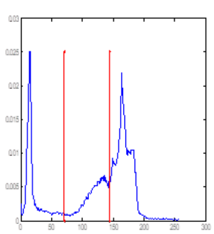
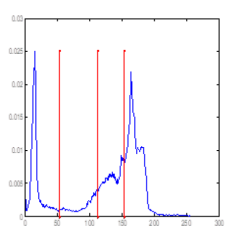
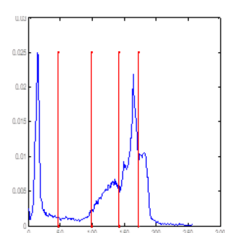
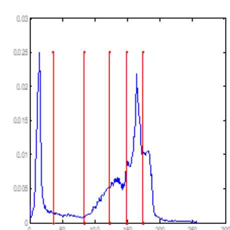
TH=2	TH=3	TH=4	TH=5
			
			
			
			
			
			

TABLE III .The Best Fitness Value Obtained from all Algorithms for Test Image

Image	K	EMO.levy	EMO	CS	SCA	MFO	WOA
Baboon	2	1552.5	1550.2	1552.5	1455.2	1538.81	1545.92
	3	1644.7	1645	1645	1527.8	1592.98	1635.7
	4	1699.4	1691	1699.5	1663.4	1647.93	1669.83
	5	1725.6	1720	1725.8	1628.2	1705.93	1682.48
pepper	2	2858.5	2857.5	2861.1	2324.5	2435.5	2433.36
	3	3059.9	3059.8	3059.9	2550.4	2574.7	2493.18
	4	3145.7	3144.8	3145.7	2537.9	2647.67	2632.9
	5	3189.8	3189.4	3189.7	2603.8	2669.47	2682.01
Cameraman	2	3646.6	3646.5	3643.6	3515.8	3643.31	3649.36
	3	3721.8	3721	3718.7	3680.5	3720.25	3690.11
	4	3778.2	3777.4	3774.5	3749.18	3748.75	3760.5
	5	3809.4	3809.4	3805.7	3718.18	3758.72	3795.86

TABLE IV. The Best Threshold Values Obtained from All Algorithms for the Test Image

Image	k	EMO with levy	EMO	CS	MFO	WOA	SCA
Baboon	2	97,149	93,147	97,149	111,133	83,151	103,150
	3	86,126,162	85,124,160	84,123,160	74,137,168	27,83,144	81,112,145
	4	71,105,136,167	60,96,127,161	71,106,137,168	44,84,133,173	96,109,134,152	61,82,128,164
	5	64,95,121,146,173	54, 83,107,137,168	68,100,125,149,174	46,94,158,189,219	32,77,101,148,207	75,99,140,159,188
pepper	2	82,143	56,120	49,116	100,182	56,152	66,143
	3	43,99,153	43,100,154	43,99,153	92,124,161	72,119,165	59,127,163
	4	41,89,135,175	43,91,138,178	41,89,135,175	41,139,157,187	47,74,132,155	68,94,136,167
	5	39,81,119,151,183	39,79,116,149,183	39,80,118,150,183	59,81,121,165,207	29,85,97,133,157	38,81,125,181,183
Camera man	2	70,144	71,144	70,144	96,167	94,146	86,145
	3	57,117,155	53,112,153	59,119,156	65,155,197	59,143,208	73,151,196
	4	41,94,140,170	47,99,142,172	41,94,140,170	59,117,133,203	77,135,200,222	76,126,166,188
	5	36,83,122,149,173	36,83,122,149,173	37,84,123,150,174	47,85,110,173,210	3,53,111,166,184	80,126,171,202,207

TABLE V. The Average of the Psnr Measure of All Algorithms

Image	K	EMO	EMO .levy	CS	MFO	WOA	SCA
Baboon	2	15.4129	15.6662	15.4129	15.8636	16.0311	15.2642
	3	17.6283	17.7558	17.8713	18.5521	18.3643	16.6058
	4	20.2143	20.9297	20.3029	20.0673	20.4056	17.5951
	5	21.9684	22.5824	21.4839	22.6505	20.2745	18.0885
Pepper	2	15.5245	16.4296	16.2354	16.6278	16.2716	14.2092
	3	18.8675	18.8469	18.8675	18.1623	18.0953	16.466
	4	20.4503	20.4749	20.4503	20.63	19.7982	16.9392
	5	21.8329	21.8597	21.855	21.4056	21.608	19.5709
camera man	2	17.253	17.3213	17.253	17.3744	18.2862	15.8362
	3	20.2116	20.2304	20.1796	20.0161	19.9084	18.0557
	4	21.4177	21.625	21.4177	21.9564	21.1897	20.219
	5	23.2749	23.2749	23.2735	21.6302	22.1592	18.6505

TABLE VI. CPU Time for All Algorithms (Per Second)

image	k	EMO.levy	EMO	CS	WOA	MFO	SCA
Baboon	2	9.34	10.39	11.67	12.55	12.88	12.97
	3	12.19	13.96	14.67	15.40	16.06	16.98
	4	14.53	18.82	19.31	20.01	21.55	22.89
	5	16.01	19.66	20.39	21.83	22.09	22.77
Pepper	2	9.67	10.31	12.31	13.95	14.33	14.87
	3	12.16	14.47	15.48	15.78	16.09	16.59
	4	14.78	16.26	17.45	18.66	19.23	20.01
	5	16.38	18.43	19.85	20.28	21.91	22.01
Camera man	2	5.34	7.53	10.32	11.03	12.22	12.83
	3	8.90	10.95	12.36	13.02	13.89	14.66
	4	12.29	14.39	16.23	17.15	17.87	18.32
	5	16.26	18.29	19.43	20.33	20.85	21.22

TABLE VII. Comparison Between Otsu And Kapur Thresholding Methods

		Otsu method			Kapur method		
Image	**K**	**Threshold**	**STD**	**PSNR**	**Threshold**	**STD**	**PSNR**
Baboon	**2**	97,149	6.92 E-13	15.6662	79, 144	1.08 E-14	15.016
	3	86,126,162	7.66 E-01	17.7558	79, 143, 232	3.60 E-15	16.018
	4	71,105,136,167	2.65 E-02	20.9297	44, 98, 152, 231	2.10 E-03	18.485
	5	64,95,121,146,173	4.86 E-02	22.5824	33, 74, 115, 159, 231	1.08 E-14	20.507
Pepper	**2**	82,143	1.38 E-12	16.4296	67, 143	7.21 E-15	16.265
	3	43,99,153	4.61 E-13	18.8469	62, 112, 162	2.80 E-03	18.367
	4	41,89,135,175	4.61 E-13	20.4749	62, 112, 162, 227	1.28 E-01	18.754
	5	39,81,119,151,183	2.33 E-02	21.8597	48, 86, 127, 171, 227	1.37 E-01	20.643
Camera man	**2**	70,144	1.40 E-12	17.3213	128, 196	3.60 E-15	13.626
	3	57,117,155	3.07 E-01	20.2304	97, 146, 196	4.91 E-02	18.803
	4	41,94,140,170	8.40 E-03	21.625	44, 96, 146, 196	1.08 E-14	20.586
	5	36,83,122,149,173	2.12 E+00	23.2749	24, 60, 98, 146, 196	6.35 E-02	20.661

TABLE VIII. Proposed Algorithm For Low Contrast Images

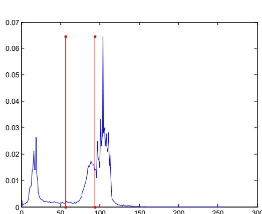
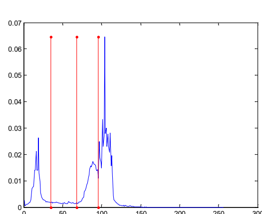
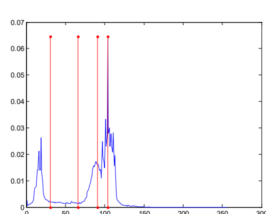
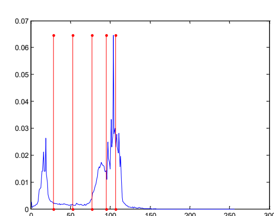
Th=2	Th=3	Th=4	Th=5
			
			

TABLE IX. Proposed Algorithm for High Contrast Images

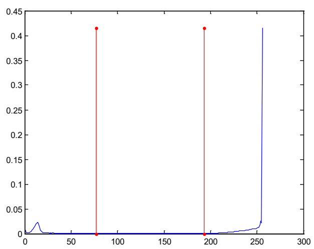
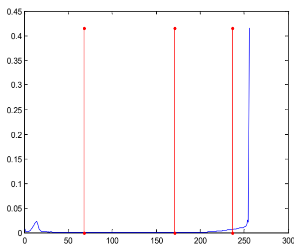
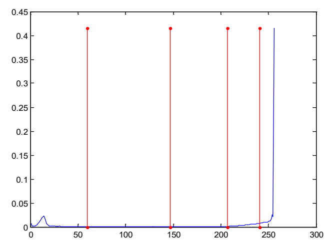
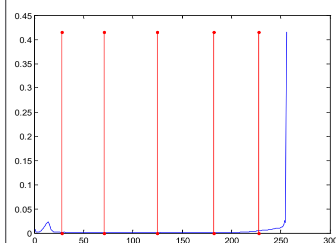
Th=2	Th=3	Th=4	Th=5
			
			

Table II shows the results of segmented images after applied four optimal thresholds TH= {2,3,4,5} based on Otsu's objective function.

Table III and Table IV show the result of the best fitness function and threshold values of three images obtained by all algorithms using Otsu's function.

From Table III, it can be observed that the proposed algorithm (EMO.levy) has a better fitness value than other algorithms when TH={2,3,4,5} .

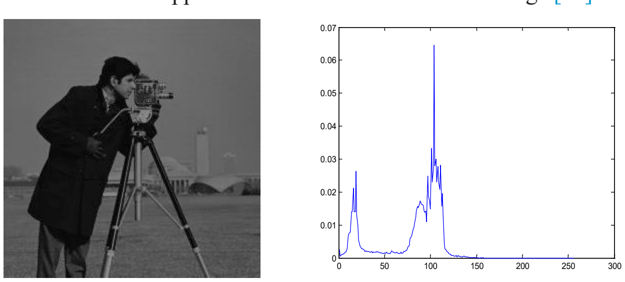
Table V shows that the proposed algorithm has mostly the highest values for PSNR compared to other multilevel algorithms (EMO, CS, MFO, WOA, SCA) also note whenever the number of threshold values increases, PSNR value also is increased for all algorithms.

Table VI presents the CPU time for all algorithms when applied on multilevel thresholding problem Experiments were conducted on MATLAB R2014 a running on an Intel®CoreTM i5 PC with 2.50 GHz CPU and 8 GB RAM. From this table, it can be observed that the developed algorithm gives low CPU time compared with the traditional EMO and CS algorithm.

Two multilevel thresholding-based segmentation methods; Otsu and Kapur are compared in Table VII. From this table, it can be observed that the Otsu thresholding method generates segmented images with high accuracy and higher values of PSNR and STD compared with Kapur method.

B. Low and High Contrast Test Image

In order to validate the effectiveness of the proposed algorithm, high and low contrast are applied to benchmark camera man image [36].

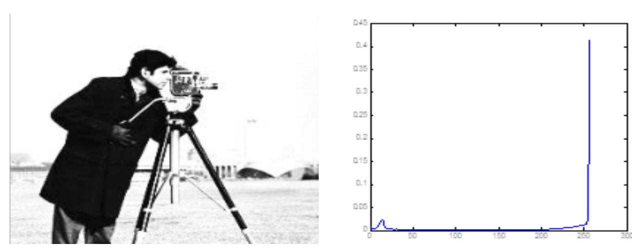


(a) (b)

Fig. 3. Low contrast camera man image and its histogram.

TABLE X. Result of Low Contrast Images

Comparison of Fitness values between algorithms				
Image	K	EMO.levy	EMO	CS
Cameraman	2	1.1517e+03	1.1516e+03	1.1516e+03
	3	1.1806e+03	1.1802e+03	1.1804e+03
	4	1.1922e+03	1.1923e+03	1.1922e+03
	5	1.1989e+03	1.1966e+03	1.1984e+03
Comparison of thresholds values between algorithms				
Image	k	EMO.levy	EMO	CS
Cameraman	2	52 ,95	52,94	52,93
	3	35,68,96	34,69, 96	34,67,96
	4	31,66,91,104	34,68,93,106	34,68,92 106
	5	29,53,77,95,107	33,67,92,104,118	32,65,87,98,109
Comparison of PSNR values between algorithms				
Image	K	EMO.levy	EMO	CS
Cameraman	2	21.4769	21.3387	21.4768
	3	24.5621	24.4175	24.5621
	4	25.6215	25.5492	25.6210
	5	27.3095	25.7825	26.8439



(a) (b)

Fig. 4. High contrast camera man image and its histogram.

TABLE XI. Result Of High Contrast Images

Comparison of Fitness values between algorithms				
Image	K	EMO.levy	EMO	CS
Camera man	2	9.2645e+03	9.2646e+03	8.6620e+03
	3	9.3622e+03	9.3622e+03	8.7206e+03
	4	9.4934e+03	9.4028e+03	8.7585e+03
	5	9.4170e+03	9.4012e+03	8.7738e+03
Comparison of PSNR values between algorithms				
Image	k	EMO.levy	EMO	CS
Camera man	2	18.9340	13.9395	12.8589
	3	19.6216	19.5743	19.2122
	4	22.3761	21.3739	20.8283
	5	22.7940	22.0383	22.4284
Comparison of thresholds values between algorithms				
Image	K	EMO.levy	EMO	CS
Camera man	2	77,193	78,193	73,184
	3	68,171,237	68,169,236	66,163,228
	4	60,147,207,241	42,108,185,238	41,106,180,232
	5	28,71,125,182,228	39,100,169,218,244	37,93,154,204,236

Tables VIII, shows the proposed algorithm for low contrast images, when (th=2,3,4,5). Table IX, shows the proposed algorithm for high contrast images, (th=2,3,4,5). Table X, shows a comparative study between the proposed algorithm and other algorithms in case of low contrast images. Table XI, shows a comparative study between the proposed algorithm and other algorithms in case of high contrast images. Fig. 3 depicts the low contrast camera man image and its histogram. Fig. 4 depicts the high contrast camera man image and its histogram. From Fig. 3, 4 and Table VIII, IX, X, XI, it can be observed that the proposed algorithm still keeps with better performance compared to the other algorithms in terms of threshold, PSNR and fitness.

From the obtained results, the proposed algorithm has a powerful computational, outperforms, high converges speed, consuming less time and saves energy to perform the computational tasks. The drawback of the proposed algorithm is that the threshold and fitness function values are not usually the best for the studied cases.

V. Outstanding Features of Proposed Algorithm

In this section, the main features of the proposed optimization algorithm (EMO.levy) are summarized as follows:

- Based on the obtained results, the proposed algorithm gives the highest values for PSNR in the most studied cases compared to other algorithms (EMO, CS, MFO, WOA, and SCA) as presented in Table V.
- The proposed algorithm (EMO.levy) gives the best fitness values in the most cases compared with the other optimization algorithms when TH={2, 3, 4, 5} as presented in Table III.
- The computation time of the proposed algorithm is lower than those obtained by the other optimization techniques as presented in Table VI.
- The reliability of the proposed algorithm has been proved using different high and low contrast images.

However, the proposed optimization algorithm has a distinguished performance in the most studied cases compared with other well-known optimization algorithms. In a few cases, the PSNR and fitness values of the proposed algorithm are little bit worse than those obtained by the other optimization techniques.

VI. Conclusion

In this paper, we have proposed an improved version for electromagnetism optimization based on levy function for multilevel thresholding of image segmentation. The aim of this algorithm is to determine the best threshold values that segment the tested image accurately, producing better results compared with the original EMO version. The peak signal-to-noise ratio (PSNR) value has been used to measure the segmentation quality and similarity between original image and segmented image. The PSNR measure proved the efficiency of the algorithm compared with the other algorithms. From the experimental results of the proposed algorithm comparing with original EMO, CS, SCA, MFO and WOA algorithms, the developed algorithm has good performance regarding to the fitness function and PSNR in all images.

The future work will focus on applying the proposed optimization algorithm for higher numbers of thresholds, solving dynamic multilevel thresholding and multi-objective problems. In addition, recent optimization techniques will be applied to in order to attain better segmentation results.

References

[1] Kumar MJ, Kumar DG, Reddy RV. Review on image segmentation techniques. International Journal of Scientific Research Engineering & Technology (IJSRET), ISSN. 2014 Sep:2278-0882.

[2] Muthukrishnan R, Radha M. Edge detection techniques for image segmentation. International Journal of Computer Science & Information Technology. 2011 Dec 1;3(6):259.

[3] Wang Z, Ma Y, Cheng F, Yang L. Review of pulse-coupled neural networks. Image and Vision Computing. 2010 Jan 1;28(1):5-13.

[4] Grau V, Mewes AU, Alcaniz M, Kikinis R, Warfield SK. Improved watershed transform for medical image segmentation using prior information. IEEE transactions on medical imaging. 2004 Apr;23(4):447-58.

[5] Chuang KS, Tzeng HL, Chen S, Wu J, Chen TJ. Fuzzy c-means clustering with spatial information for image segmentation. computerized medical imaging and graphics. 2006 Jan 1;30(1):9-15.

[6] Sezgin M, Sankur B. Survey over image thresholding techniques and quantitative performance evaluation. Journal of Electronic imaging. 2004 Jan;13(1):146-66.

[7] Kasturi R, Goldgof D, Soundararajan P, Manohar V, Garofolo J, Bowers R, Boonstra M, Korzhova V, Zhang J. Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2009 Feb;31(2):319-36.

[8] Sauvola J, Pietikäinen M. Adaptive document image binarization. Pattern recognition. 2000 Feb 29;33(2):225-36.

[9] Du S, Ibrahim M, Shehata M, Badawy W. Automatic license plate recognition (ALPR): A state-of-the-art review. IEEE Transactions on circuits and systems for video technology. 2013 Feb;23(2):311-25.

[10] Bazi Y, Bruzzone L, Melgani F. Image thresholding based on the EM algorithm and the generalized Gaussian distribution. Pattern Recognition. 2007 Feb 1;40(2):619-34.

[11] Hammouche K, Diaf M, Siarry P. A comparative study of various meta-heuristic techniques applied to the multilevel thresholding problem. Engineering Applications of Artificial Intelligence. 2010 Aug 1;23(5):676-88.

[12] Cheng HD, Chen YH, Jiang XH. Thresholding using two-dimensional histogram and fuzzy entropy principle. IEEE Transactions on Image Processing. 2000 Apr;9(4):732-5.

[13] Zahara E, Fan SK, Tsai DM. Optimal multi-thresholding using a hybrid optimization approach. Pattern Recognition Letters. 2005 Jun 30;26(8):1082-95.

[14] Zhang J, Hu J. Image segmentation based on 2D Otsu method with histogram analysis. In Computer Science and Software Engineering, 2008 International Conference on 2008 Dec 12 (Vol. 6, pp. 105-108). IEEE.

[15] Manic KS, Priya RK, Rajinikanth V. Image multithresholding based on Kapur/Tsallis entropy and firefly algorithm. Indian Journal of Science and Technology. 2016 Apr 1;9.

[16] Bhandari AK, Kumar A, Singh GK. Modified artificial bee colony based computationally efficient multilevel thresholding for satellite image segmentation using Kapur's, Otsu and Tsallis functions. Expert Systems with Applications. 2015 Feb 15;42(3):1573-601.

[17] Kennedy J. Particle swarm optimization. InEncyclopedia of machine learning 2011 (pp. 760-766). Springer US.

[18] El Aziz MA, Ewees AA, Hassanien AE. Whale Optimization Algorithm and Moth-Flame Optimization for multilevel thresholding image segmentation. Expert Systems with Applications. 2017 Oct 15; 83:242-56.

[19] Elsayed SM, Sarker RA, Essam DL. A new genetic algorithm for solving optimization problems. Engineering Applications of Artificial Intelligence. 2014 Jan 1; 27:57-69.

[20] Tao W, Jin H, Liu L. Object segmentation using ant colony optimization algorithm and fuzzy entropy. Pattern Recognition Letters. 2007 May 1;28(7):788-96.

[21] Birbil Şİ, Fang SC. An electromagnetism-like mechanism for global optimization. Journal of global optimization. 2003 Mar 1;25(3):263-82.

[22] Guan X, Dai X, Li J. Revised electromagnetism-like mechanism for flow path design of unidirectional AGV systems. International Journal of Production Research. 2011 Jan 15;49(2):401-29.

[23] Jhang JY, Lee KC. Array pattern optimization using electromagnetism-like algorithm. AEU-International Journal of Electronics and Communications. 2009 Jun 1;63(6):491-6.

[24] Lee CH, Chang FK. Fractional-order PID controller optimization via improved electromagnetism-like algorithm. Expert Systems with Applications. 2010 Dec 1;37(12):8871-8.

[25] Yurtkuran A, Emel E. A new hybrid electromagnetism-like algorithm for capacitated vehicle routing problems. Expert Systems with Applications. 2010 Apr 1;37(4):3427-33.

[26] Hung HL, Huang YF. Peak to average power ratio reduction of multicarrier transmission systems using electromagnetism-like method. Int J Innov Comput Inf Control. 2011 May 1;7(5A):2037-50.

[27] Naderi B, Tavakkoli-Moghaddam R, Khalili M. Electromagnetism-like mechanism and simulated annealing algorithms for flowshop scheduling problems minimizing the total weighted tardiness and makespan. Knowledge-Based Systems. 2010 Mar 1;23(2):77-85.

[28] Cuevas E, Oliva D, Zaldivar D, Pérez-Cisneros M, Sossa H. Circle detection using electro-magnetism optimization. Information Sciences. 2012 Jan 1;182(1):40-55.

[29] Birbil Şİ, Fang SC, Sheu RL. On the convergence of a population-based global optimization algorithm. Journal of global optimization. 2004 Nov 1;30(2-3):301-18.

[30] Oliva D, Cuevas E, Pajares G, Zaldivar D, Osuna V. A multilevel thresholding algorithm using electromagnetism optimization. Neurocomputing. 2014 Sep 2; 139:357-81.

[31] Suresh S, Lal S. An efficient cuckoo search algorithm based multilevel thresholding for segmentation of satellite images using different objective functions. Expert Systems with Applications. 2016 Oct 1; 58:184-209.

[32] Suresh, S., Lal, S., Chen, C., & Celik, T. Multispectral Satellite Image Denoising via Adaptive Cuckoo Search-Based Wiener Filter. IEEE Transactions on Geoscience and Remote Sensing. 2018 99:1-12.

[33] Mirjalili S. SCA: a sine cosine algorithm for solving optimization problems. Knowledge-Based Systems. 2016 Mar 15;96:120-33.

[34] Yildiz S., Yildiz, R. comparison of grey wolf, whale, water cycle, ant lion and sine-cosine algorithms for the optimization of a vehicle engine connecting rod. Materials Testing, 2018, 60.3: 311-315.

[35] Pavlyukevich I. Lévy flights, non-local search and simulated annealing. Journal of Computational Physics. 2007 Oct 1;226(2):1830-44.

[36] Martin D, Fowlkes C, Tal D, Malik J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. InComputer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on 2001 (Vol. 2, pp. 416-423). IEEE. 2004 Jan;13(1):146-66.

Ashraf M. Hemeida

Ashraf Mohamed Hemeida, was born in Elmenia, Egypt. He obtained his B.Sc., and M.Sc. in electrical engineering from Faculty of Engineering, Elminia University, Elmenia, Egypt in 1992 and 1996 respectively. In Nov. 1992, he engaged Higher Institute of Energy, Aswan as Teaching Assistant. From 1998 to 1999 he was a full time Ph.D. student with Electrical and Computer Engineering Department, University of New Brunswick, Canada. In Sept., 2000 he obtained his Ph.D. From Electrical Engineering Department, Faculty of Engineering, Assiut University, Assiut Egypt. From Oct. 2000 till August 2005 he was assistant professor at Electrical Engineering Department, Higher Institute of Energy, South Valley University, Aswan, Egypt. In Nov. 2005 he obtained associate professor rank. Dr. Hemeida is a full professor since July, 2011. His research interest artificial intelligence applications, optimization techniques, fuzzy systems, image processing, and advanced control techniques in electrical systems. Professor Hemeida is a member of editorial board of ARPN Journal of Engineering and Applied Sciences.

Radwa Mansour

Radwa, Mansour received the B.Sc. degree in computer science from the Faculty of Science, Aswan University, Aswan, Egypt, in 2011. Her research interests include artificial intelligence applications, optimization techniques, fuzzy systems, and image processing.

Mohamed Eid Hussein

Mohamed Eid Hussein was born in Aswan, Egypt, in 1967. He received the B.Sc. (with honors) and M.Sc. degrees from faculty of science, Assuit University, Egypt, in 1988 and 1995 respectively. In 2005, he received the Ph.D. degree in computer science from college of electrical engineering & Information Technology, Dortmund University, Germany.

He has been a Lecturer in the computer science branch, Mathematics Department, faculty of science, Aswan University, Egypt, since 2005. He is currently the Director of the Management Information System (MIS) Unit, Aswan University, Egypt. From 2008 to 2016, he was an Assistant Professor at College of Applied Sciences & Community Service , King Faisal University, KSA. His current research interests include Task Scheduling Systems, Genetic Algorithms, and Image Processing.

QoS based Web Service Selection and Multi-Criteria Decision Making Methods

Pallavi Bagga[1], Aarchit Joshi[2], Rahul Hans [2]*

[1] Department of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab (India)
[2] Department of Computer Science and Engineering, DAV University, Jalandhar, Punjab (India)

Received 8 June 2017 | Accepted 26 October 2017 | Published 15 December 2017

Abstract

With the continuing proliferation of web services offering similar efficacies, around the globe, it has become a challenge for a user to select the best web service. In literature, this challenge is exhibited as a 0-1 knapsack problem of multiple dimensions and multiple choices, known as an NP-hard problem. Multi-Criteria Decision Making (MCDM) method is one of the ways which suits this problem and helps the users to select the best service based on his/her preferences. In this regard, this paper assists the researchers in two conducts: Firstly, to witness the performance of different MCDM methods for large number of alternatives and attributes. Secondly, to perceive the possible deviation in the ranking obtained from these methods. For carrying out the experimental evaluation, in this paper, five different well-known MCDM methods have been implemented and compared over two different scenarios of 50 as well as 100 web services, where their ranking is defined on an account of several Quality of Service (QoS) parameters. Additionally, a Spearman's Rank Correlation Coefficient has been calculated for different pairs of MCDM methods in order to provide a clear depiction of MCDM methods showing the least deviation in their ranking. The experimental results comfort web service users in conquering an appropriate decision on the selection of suitable service.

Keywords

AHP, COPRAS, SAW, Spearman's Rank Correlation Coefficient, TOPSIS, VIKOR.

DOI: 10.9781/ijimai.2017.12.001

I. Introduction and Background

A web service is a self-describing software application which can be advertised, located and used over the Internet [5]. To initiate the provisioning of web service, a user must first identify the service that is desired. "How will I effectively select the web service that will meet my performance requirements?" [1]; the answer to this question, however, still remains a challenge faced by the user because of numerous services sharing similar functionalities in the web environment. The rise in the number of Web Services has been caused due to growing demands of increasing the flexibility of IT infrastructure in order to support rapidly evolving business needs [2]. The selection of web services is not only limited to meeting the users' needs, but also, non-functional information, including, reliability, response time, etc. [3].

A. Motivation

A Web Service can also be viewed as one of the encouraging technologies that could help business units to systematize their web operations on a large scale by automatic discovery and consumption of services [5]. In the Web Service Architecture [6], the Service Requester (client or user) may receive a pool of web services from the Service Provider (server) as per the initial query in "Service Discovery" stage. Subsequently, in "Service Selection" stage, the "best" web service, which satisfies all the constraints set by the original requester, is selected from the pool. This process of service filtration is carried out based on the degree of satisfaction to the users' non-functional requirements known as Quality of Service (QoS) parameters. For example, while booking flights or downloading music, there exists a number of available services sharing identical functionalities, however, they exhibit different QoS. A web service with remarkable QoS can deliver big competitive influence to service providers while bringing the social prosperity to service consumers. It has been acknowledged from the literature that due to intensive global competition, the experts recognize the decision on web service selection an important activity.

* Corresponding author.

E-mail address: rahulhans@gmail.com

B. Quality of Service (QoS) Parameters for Web Service Selection

The QoS based Web Service selection has gained the attention of many researchers in recent years, since maintaining the quality of their web services has become the topmost priority of each web service provider. In this paper, therefore, the activity of web service selection is carried out based on both functional as well as non-functional QoS parameters [4, 11]. QoS attributes are measured on a scale of 0 to 9. To facilitate the description, the set of QoS attributes is divided into two subsets: Benefit (Positive) attributes and Cost (Negative) attributes, as shown in Table I. The values of positive attributes need to be maximized, whereas the values of negative attributes need to be minimized. For the sake of simplicity, the values closer to 9 for benefit criteria and closer to 1 for cost criteria are considered good.

C. Multi-Criteria Decision Making (MCDM) Methods

In past many years, Multi-Criteria Decision Making (MCDM) methods have proven their effectiveness in addressing different complex real-world decision making problems. MCDM methods use knowledge from many fields, including economics, mathematics, behavioral decision theory, software engineering, computer technology and information systems [9]. In the context of this paper, the goal of MCDM methods is to find one web service from a pool of several web services such that the QoS is optimized and users' end-to-end QoS requirements are satisfied. In MCDM methods, each problem is stated in matrix design as:

	C_1	…	C_n
A_1	x_{11}	…	x_{1n}
…	…	…	…
A_m	x_{m1}	…	x_{mn}

where,

A_1, A_2, … A_m are possible 'm' alternatives or choices among which decision makers have to choose, C_1, C_2, ..C_n are possible 'n' criteria or attributes the basis of which the alternatives are ordered or selected,

x_{ij} is the rating of alternative A_i with respect to criterion C_j.

The general flowchart of MCDM method is shown in Fig 1. The sensitivity analysis is done to address the problems of uncertainty, imprecision, and inaccurate determination [10]. Various MCDM methods used for web service selection in the literature are briefly summarized in Table II, however Tables III and IV conveys overview and the advantages and disadvantages of various MCDMs used in this research concluded from literature by the authors. For a particular case, two or more MCDM methods often generate different rankings of web services.

D. Objectives of Present Research

The objectives of present study can be stated as follows:

- To address the problem of web service selection based on QoS.
- To study how service consumers are benefited by selecting the appropriate Web Service based on QoS using MCDM in literature.
- To implement and compare five different MCDM methods (AHP, TOPSIS, SAW, VIKOR and COPRAS) with two different large sets of web services (50 and 100) against 9 QoS attributes.

To evaluate the deviation in the rankings of different MCDM methods using Spearman's Rank Correlation Coefficient and to conclude which MCDM methods produce similar ranking.

E. Organization of Paper

The remainder of this paper is organized as follows: Section 2 throws light upon the dataset and several methods used for present study. Section 3 discusses the results from experimental evaluation. Finally, the conclusion and highlights on possible continuations of this work are addressed in Section 4.

II. Material and Methods

A. Dataset Used

The Quality of web service (QWS) dataset version 2.0 [12, 43] is chosen for the present study. This dataset includes a set of 2,507 web services and their QWS measurement which were conducted in 2007, using Web Service Broker (WSB) framework. Each row in the dataset represents a web service and its corresponding QWS measurements on nine different QoS parameters. In this paper, two different sets of 50 and 100 web services have been constructed by random sampling for the experimental study of various MCDM methods.

B. Various MCDM Methods

MCDM methods help in selecting the optimal one from a set of alternatives with respect to the predefined set of attributes. MCDM methods are continuously growing in the application areas of Business, Mathematics, Decision Sciences, Management and Accounting, Social Sciences, Medicine, Environmental Science, Economics, Econometrics and Finance, etc. [19]. In this paper, for QoS based web service selection, five different existing MCDM methods are taken into account: Simple Additive Weighting (SAW) [15], Analytic Hierarchy Process (AHP) [14], Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) [9], VIšekriterijumsko KOmpromisno Rangiranje (VIKOR) [17] and COmplex PRoportional ASsessment (COPRAS) [18]. These methods are summarized in Tables III and IV. The general formal definition of MCDM is represented in (1):

$$mcdm_k\{f_k\,(A_{ij})\} \quad (1)$$

where,

i=1,2,…,I; k=1,2,..,K ; j=1,2,…,J

k is the set of different MCDM methods, i.e. k ={AHP=1, TOPSIS=2, COPRAS=3, VIKOR=4, SAW=5}

I is the number of alternatives

J is the number of criterion

f is the Aggregating function/method

A_{ij} is the decision matrix A obtained by taking criteria 'i' and alternative 'j' as row and column respectively

C. AHP (Analytical Hierarchy Process)

Suppose J represents the number of criterion and a_{jk} represents the importance of j^{th} criterion relative to k^{th} criterion in PC matrix. The AHP consists of following steps [14]:

i) Breaking down of a problem into a hierarchy of decision criteria and alternatives.

ii) Relative importance of each criterion is measured with respect to other criterion, which is known as Pairwise Comparison PC process. The consistency checks are also performed while the evaluations are made by the decision makers.

iii) Normalization of pairwise comparison matrix by using (2.1).

$$A_{jk} = \frac{a_{jk}}{\sum_{i=1}^{J} a_{ik}}$$

where, j=1,2….J; k=1,2,….J (2.1)

iv) The Criteria Weight Vector is built with the help of (2.2).

$$W_j = \frac{\sum_{i=1}^{J} a_{ik}}{J}$$

where j=1,2,…J (2.2)

v) Obtaining the matrix S of scores of alternatives as shown in (2.3).

$$S = [s^1, s^2, \ldots s^J] \quad (2.3)$$

vi) Using (2.4) to calculate the matrix of global scores V.

$$V = S.\,W \quad (2.4)$$

vii) Ranking of alternatives is done as per the decreasing order of global scores.

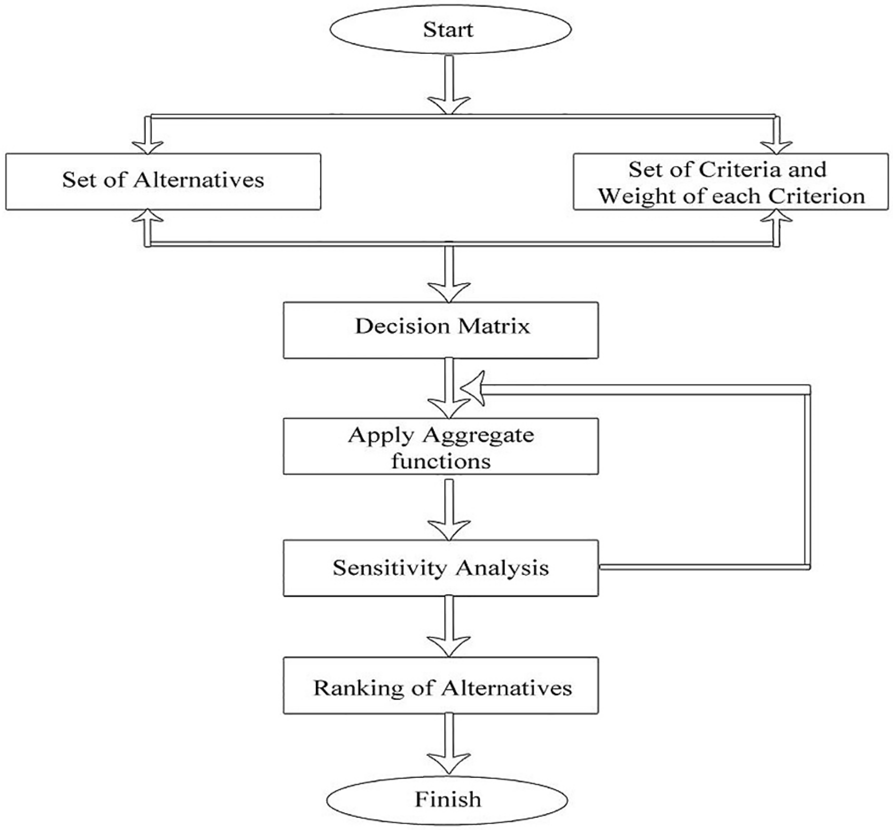


Fig. 1. General Flowchart of MCDM Method.

TABLE I. Various QoS Attributes Used for Web Service Selection [7-8, 27]

Attribute	Description	Expected Value	Units
Availability	Probability that the service is available. It is computed from historical data as: Amount of time the service is available Total time monitored Or Number of Successful Invocations Total Invocations	Maximum	%
Throughput	Total number of invocations for a given period of time	Maximum	ips
Successability	It can be defined as: Number of Responses Number of Service Requests	Maximum	%
Reliability	Probability that a request is correctly handled within the expected time. It is computed from historical data as: Number of Requests successfully responded Total number of requests	Maximum	%
Compliance	The extent to which a WSDL (Web Services Definition Language) document is followed	Maximum	%
Best Practices	The extent to which a Web Service Interoperability industry consortium (WS-I) is followed	Maximum	%
Documentation	Measure of documentation (i.e. description tags) in WSDL	Maximum	%
Latency*	Time taken for the server to process a given request.	Minimum	ms
Response Time*	Time interval between when a user requests a service and when the user receives a response.	Minimum	ms

Note: * denotes Cost Attributes; % means Percentage for a particular attribute

TABLE II. Summary of MCDM Based Existing Related Researches

Reference No.	Authors	Year	Publisher	No. of Citations	No. of Criteria	MCDM	Target area
[34]	Shaikh et al.	2004	IEEE	10	3	AHP	E-business processes
[33]	Chemane et al.	2005	IEEE	3	4	AHP	Internet Access Technology
[30]	Colace et al.	2006	IEEE	40	5	AHP	E-learning
[26]	Zhuang et al.	2007	IEEE	8	NS	Fuzzy Multi-Attributive Group Decision-Making	Web Service Selection
[38]	Xiong et al.	2007	IEEE	53	6	Fuzzy MCDM	Web service selection
[29]	Godse et al.	2008	IEEE	23	5	AHP	Prioritizing Web Service Features
[16]	Tseng	2009	Elsevier	210	21	Grey theory, Fuzzy theory, DEMATEL	Customer Service Quality Expectation
[25]	Sun et al.	2009	Elsevier	140	12	Fuzzy TOPSIS	Shopping Websites
[40]	Pervaiz	2010	IEEE	35	4	AHP	Access Network Selection in WLAN
[41]	Yang et al.	2010	IEEE	0	4	CRML, TOPSIS	Cross-organizational service selection
[37]	Karim et al.	2011	IEEE	21	9	Enhanced PROMETHEE	Web Services
[39]	Garg et al.	2011	IEEE	193	6	AHP	Cloud Services Ranking
[36]	Luo et al.	2012	Elsevier	20	6	SAW	Wireless Network
[31]	Park et al.	2013	Springer	13	6	Pairwise comparison	Enterprise Resource Planning
[32]	Sun et al.	2013	IEEE	17	5	AHP	Consumer-centered Cloud services
[35]	Fakhfakh et al.	2013	Springer	4	3	MACBETH integrated with 2-additive Choquet integral	Degree of Service Orchestrations Measurement
[28]	Dragović et al.	2014	Taylor & Francis	14	5	AHP, Fuzzy Logic	Web services
[27]	Almulla et al.	2015	Elsevier	5	9	FDCRT, FSIRT	Real world web services
[21]	Lin et al.	2016	Elsevier	0	5	Hybrid of DEMATEL, PCA, ANP, VIKOR	Digital Music Services
[20]	Huang et al.	2016	Springer	0	6	DEMATEL, DANP	Social Networking sites
[22]	Sánchez-Lozano et al.	2016	Elsevier	5	10	Fuzzy AHP, Fuzzy TOPSIS	Geographical Information System
[23]	Gupta et al.	2016	Springer	0	6	Ordered Weighted Operators	Handling Outliers in Web Data
[24]	Rhimi et al.	2016	IEEE	0	NS	Fuzzy logic, TOPSIS	Skyline computation
[42]	Taibi et al.	2017	IMAI Software	0	4	Fuzzy AHP	Industrial Site Selection

Note: #Citations are taken up to November 2016, NS means Not Specified

DEMATEL: Decision MAking Trial and Evaluation Laboratory, ANP: Analytical Network Procedure, PCA: Principal Component Analysis, DANP: Decision MAking Trial and Evaluation Laboratory based Network Process, FSIRT: Fuzzy Interval-based Ranking Technique, FDCRT: Fuzzy Distance Correlation Ranking Technique, MACBETH: Measuring Attractiveness by a Categorical Based Evaluation TecHnique, SAW: Simple Additive Weighting, AHP: Analytic Hierarchy Process, TOPSIS: Technique for Order Preference by Similarity to Ideal solution, VIKOR: VIšekriterijumsko KOmpromisno Rangiranje

D. TOPSIS (Technique for Order Preference by Similarity to Ideal Solution)

Suppose I represents the number of alternatives and J represents the number of criterion; x_{ij} is the value assigned to i^{th} alternative with respect to j^{th} criterion. The TOPSIS consists of following steps [9]:

i) Normalization of decision matrix using (3.1).

$$r_{ij} = \frac{x_{ij}}{[\,{}^{I}\sum_{k=1} (x_{kj})^2\,]^{1/2}}$$

where i=1,2…I, j=1,2,….J (3.1)

ii) Calculation of the weighted normalized decision matrix using (3.2).

$$v_{ij} = w_j \cdot r_{ij} \quad (3.2)$$

where W={w_1,w_2,…w_J}=relative weight about the criterion

iii) Determination of the positive ideal and negative ideal solutions which maximizes and minimizes the benefit criteria and cost criteria respectively using (3.3) and (3.4).

$$A^+ = \{v_1^+, v_2^+, \ldots\ldots v_J^+\}$$
$$= \{\max_j v_{ij} \mid j\in \text{set of Benefit Criteria},$$
$$\min_j, v_{ij} \mid j \in \text{set of Cost Criteria}\} \quad (3.3)$$

$$A^- = \{v_1^-, v_2^-, \ldots\ldots v_J^-\}$$
$$= \{\min_j v_{ij} \mid j \in \text{set of Benefit Criteria},$$
$$\max_j v_{ij} \mid j \in \text{set of Cost Criteria}\} \quad (3.4)$$

iv) Calculation of each alternative from positive ideal solution and negative ideal solution with the help of (3.5) and (3.6) respectively (Euclidean distance).

$${}^{+}D_i = [\,{}^{J}\sum_{j=1} (v_{ij} - v_j^+)^2\,]^{1/2} , i=1,2\ldots I \quad (3.5)$$

$${}^{-}D_i = [\,{}^{J}\sum_{j=1} (v_{ij} - v_j^-)^2\,]^{1/2} , i=1,2...I \quad (3.6)$$

v) Calculation of the relative closeness of each alternative to the ideal solutions using (3.7).

$$RC_i = \frac{{}^{-}D_i}{{}^{+}D_i + {}^{-}D_i}$$

where i= 1,2,…I (3.7)

vi) Ranking of alternatives is done as per the increasing order of relative closeness. Higher the relative closeness to ideal solution is, better is the alternative.

E. COPRAS (COmplex PRoportional ASsessment)

Suppose I represents the number of alternatives and J represents the number of criterion; x_{ij} is the value assigned to i^{th} alternative with respect to j^{th} criterion; q_j represents the significance of j^{th} criterion. The steps of COPRAS are summarized as follows [18]:

i) Normalization of decision matrix using (4.1).

$$r_{ij} = \frac{x_{ij}}{{}^{I}\sum_{i=1} x_{ij}}$$

where i=1,2…I and j=1,2,….J (4.1)

ii) Calculation of Maximizing Criteria S_+ and Minimizing Criteria S_- using (4.2) and (4.3) respectively.

$$S_{+i} = {}^{J'}\sum_{j=1} r_{ij} \cdot q_j \quad (4.2)$$

where i=1,2,…I; j' is the number of benefit criterion

$$S\text{-}_i = {}^{J''}\sum_{j=1} r_{ij} \cdot q_j \quad (4.3)$$

where i=1,2,…I; j" is the number of cost criterion

iii) Relative weight of each alternative is obtained using 4.4.

$$Q_i = S_{+i} \frac{{}^{J}\sum_{i=1} S_{-i}}{S_{-i} \cdot {}^{J}\sum_{i=1} 1/S_{-i}}$$

where i=1,2,…I (4.4)

iv) Ranking of alternatives is done according to the ascending order of relative weight. Higher the relative weight of alternative is, higher is the priority of alternative.

F. VIKOR (VIšekriterijumsko KOmpromisno Rangiranje)

Suppose I represents the number of alternatives and J represents the number of criterion; f_{ji} is the value of j^{th} criterion for the i^{th} alternative The VIKOR involves the aforementioned steps [17]:

i) Determination of the best and the worst values of criterion using (5.1) and (5.2) respectively.

${}^{+}f_j = \max_i f_{ji}$ for benefit criterion

${}^{+}f_j = \min_i f_{ji}$ for cost criterion

where j =1,2,….J (5.1)

${}^{-}f_j = \min_i f_{ji}$ for benefit criterion

${}^{-}f_j = \max_i f_{ji}$ for cost criterion

where j =1,2,….J (5.2)

ii) Computation of the values S_j and R_j using (5.3) and (5.4) respectively.

$$S_i = {}^{I}\sum_{i=1} \frac{w_j ({}^{+}f_j - f_{ji})}{({}^{+}f_j - {}^{-}f_j)}$$

where j =1,2,….J (5.3)

$$R_i = \max_j [w_j ({}^{+}f_j - f_{ji}) / ({}^{+}f_j - {}^{-}f_j)]$$

where j =1,2,….J (5.4)

iii) Computation of Q_j using (5.5).

$$Q_i = v(S_i - S^+) / (S^- - S^+) + (1\text{-}v)(R_i - R^+) / (R^- - R^+) \quad (5.5)$$

where i =1,2,….I; and

$S^- = \max_i\{S_i\}$; $S^+ = \min_i\{S_i\}$; $R^- = \max_i\{R_i\}$; $R^+ = \min_i\{R_i\}$

iv) Three different Ranking of alternatives is done according to S, R and Q. Lower the value of Q is, better is the alternative.

v) The alternative A is considered as the compromised solution if the following two conditions are satisfied:

- *C1-Acceptable Advantage*: $Q(a^2)$- $Q(a^1) \geq DQ$, where a^2 and a^1 are the alternatives with ranking second and first respectively;

DQ=1/(Number of alternatives -1).

- *C2-Acceptable stability in decision making*: the alternative a^1 must also be best ranked by S and R.

vi) If any of the conditions is not true, then the set of compromised solutions are proposed consisting of

- If only C2 is not satisfied, then alternatives a^1 and a^2 are proposed.
- If only C1 is not satisfied, then alternatives a^1, a^2, …a^m are proposed; a^m is determined by the relation $Q(a^m)-Q(a^1)<DQ$ for maximum 'm' (the positions of these alternatives are "in closeness").

G. SAW (Simple Additive Weighting)

Suppose I represents the number of alternatives and J represents the number of criterion; x_{ij} is the value assigned to i^{th} alternative. The SAW method consists of the following steps [15]:

i) Normalization of decision matrix using (6.1) and (6.2) respectively for cost and benefit criterion.

$$V_{ij} = \frac{x_{ij}}{{}^{max}x_j} \tag{6.1}$$

$$V_{ij} = \frac{{}^{min}x_j}{x_{ij}} \tag{6.2}$$

ii) Calculation of weighted normalized values is done using (6.3).

$$v_{ij} = Vij\ .\ wj \tag{6.3}$$

where $W=\{w_1,w_2,\ldots w_J\}$=relative weight about the criterion; j=1,2…J

iii) The sum S is calculated correspond to each alternative using (6.4).

$$S_i = {}^{J}\sum_{i=1} v_{ij}$$

where i=1,2….I (6.4)

iv) Ranking of alternatives is done according to the increasing order of S.

H. Spearman's Rank Correlation Coefficient

The Spearman's rank correlation coefficient method assists in finding the similarity between two sets of ranking obtained from two different 'k^{th}' and 'i^{th}' MCDM methods, using (7):

$$\rho_{ki} = 1 - \frac{6\ {}^{n}\sum_{i=1} d_i^2}{n^3-n} \tag{7}$$

where, n is the number of web services, and d_i is the difference between the ranks of two MCDM methods. A larger absolute value indicates a good agreement between one MCDM method and other MCDM method [13].

III. Results and Discussions

For carrying out the experiments, two sets of web services are constructed out of present dataset: one set of 50 web services (Scenario 1) and other set of 100 web services (Scenario 2). Five different MCDM methods such as AHP, COPRAS, SAW, TOPSIS, and VIKOR

TABLE III. Brief Overview of Various MCDM Methods Used in Present Study

Method Name	Introduced By	Originating Year	Normalization Method	Distance Method	Aggregating Method
SAW	MacCrimmon	1968	Linear Normalization	NA	Additive Weights
AHP	Thomas Satty	1977	Linear Normalization	NA	Priority Vector
TOPSIS	Hwang and Yoon	1981	Vector Normalization	Euclidean	Closeness Coefficient
VIKOR	Serafim Opricovic	1990	Linear Normalization	Manhattan and Chebyshev	Distance from ideal solution
COPRAS	Zavadskas and Kaklauskas	1996	Linear Normalization	NA	Relative weight

Note: NA means Not Applicable

TABLE IV. Comparison of Various MCDM Methods

Method Name	Advantages	Disadvantages
AHP	• Hierarchical structure of criteria. • Pairwise comparison gives better comparisons of criteria. • Gives option to evaluate quantitative and qualitative criteria and alternatives.	• If Hierarchical structure of criteria is not made properly user may get worst ranking. • In special cases (currencies exchange), it may not work. • Absolute zero doesn't exist.
TOPSIS	• Scalar value that accounts for both the best and worst alternatives simultaneously. • Sound logic that represents the thesaurus of human choice.	• Possess rank reversal problem. • Its use of Euclidian distance does not consider the correlation of attributes, difficult to weight and keep consistency of judgment.
COPRAS	• Degree of utility is the bases of ranking. • It doesn't require transformation of cost and benefit criteria.	It has complex aggregation procedure.
SAW	Simple to understand and implement.	Result not always real to situation.
VIKOR	It gives Ideal and compromised solution.	Complex to understand and implement.

are implemented in which the weights for different QoS attributes are calculated using AHP method in each scenario. The outline of experiments can be illustrated from Fig 2. The top five ranking of each method for each scenario is shown in Fig 3.

In order to check the consistency of user inputs to QoS attributes, the matrix analysis is usually done for AHP [14, 30], however, this effective technique of checking the evaluations made by the decision maker is used for each method while constructing the pairwise comparison matrices. The formula used for obtaining the Consistency Index is shown in (8).

$$CI = ((\lambda_{max}-n) / (n-1)) \quad (8)$$

where, n is the matrix size and λ_{max} is the eigen value.

Generally, the value CI = 0 is obtained by a perfectly consistent decision-maker, however, the smaller inconsistency may be tolerated. In particular, the inconsistencies are tolerable and reliable results have been expected from each method, if (9) holds true.

$$CI / RI < 0.1 \quad (9)$$

where, RI is Random Index, means the Consistency Index when matrix has entries which are completely random [30], and the ratio CI/RI is known as Consistency Ratio.

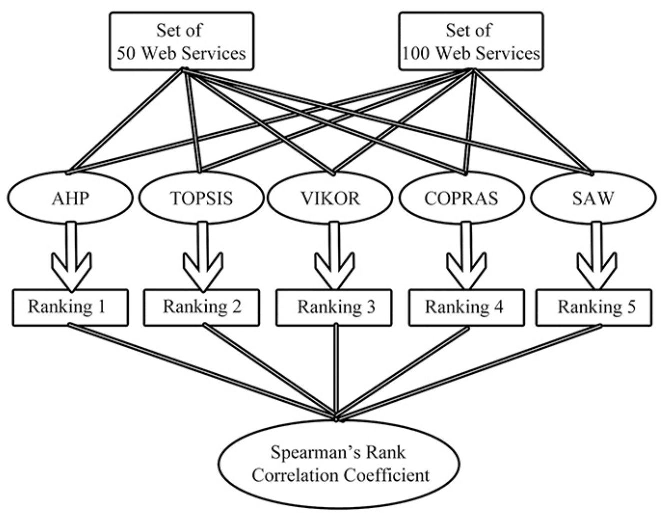


Fig. 2. Outline of Experimental Evaluation Process.

Scenario 1: 50 Web services and 5 MCDM methods (Table V)

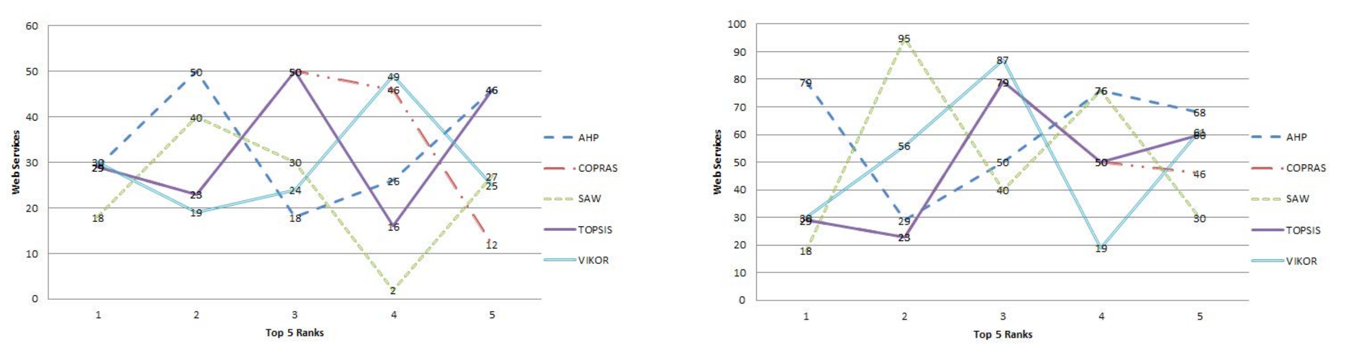
AHP method shows 29th, 50th and 18th numbered are the best three web services to be selected, whereas 21st, 6th and 38th numbered are the worst three to be selected. Moreover, COPRAS method produces 29th, 23rd and 50th as the best three web services and 38th, 7th and 27th as worst three web services. On the other hand, using SAW method, 18th, 40th and 30th numbered three best web services are obtained, whereas 21st, 38th and 6th are worst. Furthermore, TOPSIS method gives best three ranks to 29th, 23rd and 50th numbered web services, while worst three ranks to 7th, 6th and 38th. Also, 30th, 19th and 24th numbered are best three web services, while 32nd, 36th and 22nd are worst three web services chosen by VIKOR method.

Scenario 2: 100 Web Services and 5 MCDM methods (Table VI)

AHP method shows 79th, 29th and 50th numbered are the best three web services to be selected, whereas 38th, 96th and 50th numbered are the worst three web services to be selected. While, COPRAS method produces 29th, 23rd and 79th as best three web services and 64th, 77th and 27th as worst three web services. Similarly, using SAW method, 18th, 95th and 40th numbered three best web services are obtained, whereas 6th, 96th and 53rd are worst. On one hand, TOPSIS method gives best three ranks to 29th, 23rd and 79th numbered web services, on the flip side, worst three ranks to 96th, 38th and 53rd. Also, 30th, 56th and 87th numbered are best three web services, while 78th, 22nd and 73rd are worst three web services chosen by VIKOR method.

Clearly, in both the scenarios (1 and 2), rankings obtained from five different MCDM methods in the QoS based Web Service Selection problem yields divergent results. This difference in ranking can be seen due to either the use of different normalization techniques on decision matrix or the use of different aggregating methods in each MCDM method. Further, in order to evaluate the closeness of ranking, for both the scenarios, Spearman's rank correlation coefficients are calculated using (7).

The closeness of the correlation coefficient value (example, 0.9526), ranging between -1 to 1, in approximation to unity indicates complete dependency and reliability of either of the methods used. The dependency reduces with each unit reduction in the coefficient value. The negative sign indicates reverse trend existing between the two methods i.e. the rank value increasing in one method shall be declining under the second method in comparison. The inter-relationship between the MCDM methods is analyzed through correlation matrix using (7) as shown in Tables VII and VIII. It is found that AHP and TOPSIS show maximum Correlation value i.e. 0.9535 in Scenario 1 (Table VII) and 0.9526 in Scenario 2 (Table VIII), indicating the strongest correlation, as the values are generously high. Thus, it can be concluded that in this web service selection problem, AHP and TOPSIS can be used effectively for making similar types of decisions. All other combinations show positive correlation except COPRAS and VIKOR. These findings hold true for both scenarios of 50 as well as 100 web services.



(a)

(b)

Fig. 3. Graphical Comparison of Top 5 Ranks obtained by each MCDM method for both scenarios: (a) 50 Web Services (b) 100 Web Services.

TABLE VII
Spearman's Rank Correlation Coefficients with Respect to Scenario 1

Methods	AHP	COPRAS	SAW	TOPSIS	VIKOR
AHP	-	0.7880	0.6379	**0.9535**	0.2949
COPRAS	0.7880	-	0.3290	0.8226	-0.0131
SAW	0.6379	0.3290	-	0.4982	0.3886
TOPSIS	**0.9535**	0.8226	0.4982	-	0.1602
VIKOR	0.2949	-0.0131	0.3886	0.1602	-

TABLE VIII
Spearman's Rank Correlation Coefficients with Respect to Scenario 2

Methods	AHP	COPRAS	SAW	TOPSIS	VIKOR
AHP	-	0.6872	0.7343	0.9526	0.3769
COPRAS	0.6872	-	0.3306	0.7391	-0.0432
SAW	0.7343	0.3306	-	0.6151	0.4748
TOPSIS	0.9526	0.7391	0.6151	-	0.7109
VIKOR	0.3769	-0.0432	0.4748	0.7109	-

IV. Conclusions and Future Scope

Nowadays, there is a need to distinguish increasing number of web services with similar functionalities, being made accessible across the Internet, using a set of QoS parameters. The QoS level displays abundant influence on degree of the web service usability as well as effectiveness, both of which further influences the service popularity. In this regard, the problem of web service selection based on QoS using MCDM method is addressed in this paper. Firstly, the rankings of web services are calculated using five different MCDM methods, including AHP, TOPSIS, VIKOR, SAW, and COPRAS. Secondly, since these rankings show divergent results, Spearman's rank correlation coefficient is used to compute the degree of similarity in the rankings of one MCDM method with other MCDM methods. The whole process is done for two larger sets of web services: 50 (Scenario 1) and 100 (Scenario 2). Maximum co-efficient correlation value is deduced for the combination of AHP and TOPSIS in both the scenarios. The experimental outcomes on different sets of web services using different MCDM methods reveal that AHP and TOPSIS methods show good agreement with each other. In future, the work can be extended using different correlation methods such as Pearson, Kendall or any other correlation method. The more focus can also be put on Rank Reversal problem of MCDM methods.

V. Acknowledgements

The authors express their gratitude to E. Al-Masri and Q. H. Mahmoud for providing the QWS dataset which has supported this research to evaluate MCDM methods on well-validated web services based on different Quality of Service parameters.

References

[1] G. Chang. (2012). QoS-Based Web Service Selection Approach. In: Wu Y. (eds.) Software Engineering and Knowledge Engineering: Theory and Practice, Advances in Intelligent and Soft Computing, Vol. 115, pp. 887-892, Springer, Berlin, Heidelberg.

[2] I. Egambaram, G. Vadivelou and S. P. Sivasubramanian. (2010). QoS Based Web Service Selection. In: The International Conference on Computing, Communications and Information Technology Applications (CCITA-2010).

[3] H. Wang, P. Tong and P. Thompson. (2007). QoS-Based Web Services Selection. In: IEEE International Conference on e-Business Engineering (ICEBE 2007), pp. 631-637.

[4] S. Ran. (2003). A Model for Web Services Discovery with QoS. ACM SIGecom Exchanges, Vol. 4, No. 1, pp. 1-10.

[5] Y. Liu, A. H. Ngu and L. Z. Zeng. (2004). QoS Computation and Policing in Dynamic Web Service Selection. In: Proceedings of the 13th International World Wide Web Conference on Alternate Track Papers & Posters, New York, USA, pp. 66-73.

[6] S. Haddad, L. Mokdad and S. Youcef. (2010). Selection of the best composite Web service based on quality of service. In: Proceedings of the 2nd International Symposium on Services Science and 3rd International Conference on Business Process and Services Computing (ISSS/BPSC,'10), pp. 255-266.

[7] S. Susila and S. Vadivel. (2011). Web Service Selection based on QoS Attributes using Entropy Discretization Method. International Journal of Computer Applications, Vol. 30, No. 2, pp. 47-53.

[8] T. Yu and K. Lin. (2005). Service selection algorithms for Web services with end-to-end QoS constraints. Information Systems and e-Business Management, Vol. 3, No. 2, pp. 103-126.

[9] M. Behzadian, S. K. Otaghsara, M. Yazdani and J. Ignatius. (2012). A state-of-the-art survey of TOPSIS applications, Vol. 39, pp. 13051-13069.

[10] P. Jankowski. (1995). Integrating geographical information systems and multiple criteria decision-making methods. International Journal of Geographical Information Systems, Vol. 9, No. 3, pp. 251-273.

[11] P. Wang. (2009). QoS-aware web services selection with intuitionistic fuzzy set under consumer's vague perception. Expert Systems with Applications, Vol. 36, No. 3 (Part 1), pp. 4460-4466.

[12] E. Al-Masri and Q. H. Mahmoud (2007). QoS-based Discovery and Ranking of Web Services. In: 16th IEEE International Conference on Computer Communications and Networks (ICCCN), pp. 529-534.

[13] G. Kou, Y. Lu, Y. Peng and Y. Shi. (2012). Evaluation of classification algorithms using MCDM and rank correlation. International Journal of Information Technology & Decision Making, Vol. 11, No. 1, pp. 197–225.

[14] T. L. Saaty. (2008). Decision making with the analytic hierarchy process. International Journal of Services Sciences, Vol. 1, No. 1, pp. 83-98.

[15] K. Madani, M. Sheikhmohammady. S. Mokhtari, M. Moradi and P. Xanthopoulos. (2014). Social Planner's Solution for the Caspian Sea Conflict. Group Decision and Negotiation, Vol. 23, No. 3, pp. 579-596.

[16] M.L. Tseng. (2009). A causal and effect decision making model of service quality expectation using grey-fuzzy DEMATEL approach. Expert Systems with Applications, Vol. 36, No. 4, pp. 7738-7748.

[17] H. Liao and Z. Xu. (2013). A VIKOR-based method for hesitant fuzzy multi-criteria decision making. Fuzzy Optimization and Decision Making, Vol. 12, No. 4, pp. 373-392.

[18] M. M. Fouladgar, A. Yazdani-Chamzini, A. Lashgari, E. K. Zavadskas and Z. Turskis. (2012). Maintenance strategy selection using AHP and COPRAS under fuzzy environment. International Journal of Strategic Property Management, Vol. 16, No. 1, pp. 85-104.

[19] R. Rekik, I. Kallel, J. Casillas and A. M. Alimi. (2016). Using Multiple Criteria Decision Making Approaches to Assess the Quality of Web Sites. International Journal of Computer Science and Information Security, Vol. 14, No. 7, pp. 747-761.

[20] C. Y. Huang, H. H. Wu and H. H. Lu. (2016). A MCDM Methods Based TAM for Deriving Influences of Privacy Paradox on User's Trust on Social Networks. In: International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, pp. 356-363.

[21] C. L. Lin, Y. H. Shih, G. H. Tzeng and H. C. Yu. (2016). A service selection model for digital music service platforms using a hybrid MCDM approach. Applied Soft Computing, Vol. 48, pp. 385-403.

[22] J. M. Sánchez-Lozano, M. S. García-Cascales and M. T. Lamata. (2016). GIS-based onshore wind farm site selection using Fuzzy Multi-Criteria Decision Making methods. Evaluating the case of Southeastern Spain. Applied Energy, Vol. 171, pp. 86-102.

[23] A. Gupta and S. Kohli. (2015). An MCDM approach towards handling outliers in web data: a case study using OWA operators. Artificial Intelligence Review, Vol. 46, No. 1, pp. 59-82.

[24] F. Rhimi, S. B. Yahia and S. B. Ahmed. (2016). Refining the Skyline with fuzzy similarity measures and Topsis method for the optimization of web services composition. In: 2016 IEEE International Conference on Fuzzy Systems, IEEE, pp. 2091-2097.

[25] C.C. Sun and G. T. Lin. (2009). Using fuzzy TOPSIS method for evaluating the competitive advantages of shopping websites. Expert Systems with Applications, Vol. 36, No. 9, pp. 11764-11771.

[26] L. Zhuang, G. HeQing, L. Dong, H. Tao and Z. JuanJuan. (2007). Solving Multi-Objective and Fuzzy Multi-Attributive Integrated Technique for QoS-Aware Web Service Selection. In: International Conference on Wireless Communications, Networking and Mobile Computing (WiCom), pp. 735-739.
[27] M. Almulla, H. Yahyaoui and K. Al-Matori. (2015). A new fuzzy hybrid technique for ranking real world Web services. Knowledge-Based Systems, Vol. 77, pp. 1-15.
[28] I. Dragović, N. Turajlić, D. Radojević and B. Petrović. (2013). Combining boolean consistent fuzzy logic and AHP illustrated on the web service selection problem. International Journal of Computational Intelligence Systems, Vol. 7, pp. 84-93.
[29] M. Godse, R. Sonar and S. Mulik. (2008). The Analytical Hierarchy Process Approach for Prioritizing Features in the Selection of Web Service. In: IEEE Sixth European Conference on Web Services (ECOWS'08), pp. 41-50.
[30] F. Colace, M. D. Santo and A. Pietrosanto. (2006). Evaluation Models for E-Learning Platform: an AHP approach. In: IEEE 36th Annual Frontiers in Education Conference, pp. 1-6.
[31] J. H. Park and H. Y. Jeong. (2013). The QoS-based MCDM system for SaaS ERP applications with Social Network. The Journal of Supercomputing, Vol. 66, No. 2, pp. 614-632.
[32] M. Sun, T. Zang, x. Xu and R. Wang. (2013). Consumer-Centered Cloud Services Selection Using AHP. In: International Conference on Service Sciences (ICSS), pp. 1-6.
[33] L. Chemane, L. Ekenberg, O. Popov and T. Cossa. (2005). MCDM Model for Selecting Internet Access Technologies - A Case Study in Mozambique. In: The 2005 International Conference on Computer as a Tool, Vol. 2, pp. 1738-1741.
[34] S. E. Shaikh and N. Mehandjiev. (2004). Multi-attribute negotiation in e-business process composition. In: 13th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, pp. 141-146.
[35] N. FakhFakh, H. Verjus, Frédéric Pourraz and P. Moreaux. (2013). QoS aggregation for service orchestrations based on workflow pattern rules and MCDM method: evaluation at design time and runtime. Service Oriented Computing and Applications, Vol. 7, No. 1, pp. 15-31.
[36] Y. S. Luo, K. Yang, Q. Tang, J. Zhang and B. Xiong. (2012). A multi-criteria network-aware service composition algorithm in wireless environments. Computer Communications, Vol. 35, No. 15, pp. 1882-1892.
[37] R. Karim, C. Ding and C. H. Chi. (2011). An Enhanced PROMETHEE Model for QoS-Based Web Service Selection. In: IEEE 2011 International Conference on Services Computing (SCC), pp. 536-543.
[38] P. Xiong and Y. Fan. (2007). QoS-aware Web Service Selection by a Synthetic Weight. In: Fourth International conference on Fuzzy Systems and Knowledge Discovery (FSKD), Vol. 3, pp. 632-637.
[39] S. K. Garg, S. Versteeg and R. Buyya. (2011). SMICloud: A Framework for Comparing and Ranking Cloud Services. In: Fourth IEEE International Conference on Utility and Cloud Computing (UCC), pp. 210-218.
[40] H. Pervaiz. (2010). A Multi-Criteria Decision Making (MCDM) Network Selection Model providing enhanced QoS differentiation to customers. In: 2010 International Conference on Multimedia Computing and Information Technology (MCIT), pp. 49-52.
[41] R. Yang, L. Qi, W. Lin, W. Dou and J. Chen. (2010). A Comprehensive Evaluation Method for Cross-Organizational Service Selection. In: IEEE 13th International Conference on Computational Science and Engineering (CSE), pp. 357-362.
[42] A. Taibi and B. Atmani (2017). Combining Fuzzy AHP with GIS and Decision Rules for Industrial Site Selection. International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 4, No. 6, pp. 60-69.
[43] E. Al-Masri and Q. H. Mahmoud (2007). Discovering the best web service. (Poster) In: 16th IEEE International Conference on World Wide Web (WWW), pp. 1257-1258.

Pallavi Bagga

Pallavi Bagga is presently working as an Assistant Professor in the Department of Computer Science and Engineering at Lovely Professional University, Phagwara, Punjab, India. She has accomplished her Master of Technology in Computer Science & Engineering from DAV University, Jalandhar, Punjab, India in 2016. Earlier, she has completed her Bachelor of Technology in Computer Science & Engineering from Punjab Technical University, Punjab, India in 2014. Her predominant areas of research interests include Distributed Computing and Machine Learning. Nowadays, she is actively pursuing her research career in the sphere of Mobile Agents System Security using Machine Learning Algorithms.

Aarchit Joshi

Aarchit Joshi has completed his M.Tech in Computer Science and Engineering from DAV University, Jalandhar, Punjab, India. Earlier, he received his B.E. degree from Chitkara University, Barotiwala, Solan, H.P., India. His research interest topics include Multi Criteria Decision Making techniques and Optimization techniques.

Rahul Hans

Rahul Hans is currently working as an Assistant Professor in the Department of Computer Science and Engineering at DAV University, Jalandhar, Punjab, India. He has received his B.Tech degree in Computer Science and Engineering from Punjab Technical University, Jalandhar, Punjab, India and subsequently his M.Tech degree in Computer Science and Engineering from Guru Nanak Dev University, Amritsar, Punjab, India. His area of research includes Machine learning, Mobile Agent Systems and cloud computing. He has published research papers in various reputed International Journals and Conferences.

Multimodal Generic Framework for Multimedia Documents Adaptation

Hajar Khallouki*, Mohamed Bahaj

Department of Mathematics & Computer Science, Faculty of Science and Technologies, Hassan 1st University, Settat (Morocco)

Received 10 July 2017 | Accepted 8 February 2018 | Published 16 February 2018

LA UNIVERSIDAD EN INTERNET

Abstract

Today, people are increasingly capable of creating and sharing documents (which generally are multimedia oriented) via the internet. These multimedia documents can be accessed at anytime and anywhere (city, home, etc.) on a wide variety of devices, such as laptops, tablets and smartphones. The heterogeneity of devices and user preferences has raised a serious issue for multimedia contents adaptation. Our research focuses on multimedia documents adaptation with a strong focus on interaction with users and exploration of multimodality. We propose a multimodal framework for adapting multimedia documents based on a distributed implementation of W3C's Multimodal Architecture and Interfaces applied to ubiquitous computing. The core of our proposed architecture is the presence of a smart interaction manager that accepts context related information from sensors in the environment as well as from other sources, including information available on the web and multimodal user inputs. The interaction manager integrates and reasons over this information to predict the user's situation and service use. A key to realizing this framework is the use of an ontology that undergirds the communication and representation, and the use of the cloud to insure the service continuity on heterogeneous mobile devices. Smart city is assumed as the reference scenario.

Keywords

Multimedia Documents, Heterogeneity Of Devices, Sensors, Multimodal Architecture, Ubiquitous Computing, Cloud Computing, Smart City, Internet Of Things.

DOI: 10.9781/ijimai.2018.02.009

I. Introduction

A smart city is an extension of a city which allows the intelligent use of resources to improve the quality of life in urban areas. Various aspects are considered to be the focus of the transformation towards smart cities, e.g. smart home, smart mobility, smart parking, smart lighting, and many more. Computing is moving toward Smart, ubiquitous environments in which heterogeneous devices (Smartphone, tablet, laptop, smart TV, etc.), applications and services are all expected to integrate and cooperate in support of human objectives; anticipating needs, negotiating for the service, acting on our behalf, and delivering services in anywhere (home, city, etc.) and anytime.

Currently, people are capable of creating and sharing documents (which generally are multimedia oriented) via the internet according to their context. Context has been defined in [1] as "the knowledge that can be used to characterize the situation of any entity that is relevant for the (pervasive) system under consideration". Many context information models are limited from the view of interoperability. Thus, the need for ontology based context modeling is given for many context-aware applications [2].

Multimodal interactions and multimodality refer to the process in which different devices and people are able to interact aurally, visually, by touch or by gesture. One of the main purposes of multimodality is the improvement of user interactions with devices, such as smartphones, laptops, tablet, etc [3] [4]. User interfaces should allow users to interact with the content or a service through the most appropriate mode or through multiple modes, taking into account user preferences and context [5].

In this paper, we introduce a multimodal generic framework to support context-aware computing in smart fields (Home, City, Tourism, etc.). Our proposed approach is developed using a semantic multimodal user context to create a more natural communication system, adaptation decision making and a cloud as a way to ensure the service continuity since a wide variety of devices have to coexist in a very heterogeneous environments.

The rest of the paper is organized as follows. In Section 2, we survey the related work in literature. In Section 3, we introduce and discuss our contribution. Possible application scenarios are discussed in section 4. The result and discussion are provided in section 5. Finally, we conclude our paper and present future work in Section 6.

II. Related Work

Making cities smarter is spreading all over the world during the latest five years. The term "smart city" was defined by [6] and [7] as a strategy to surround the modern urban production factors in a common platform and to accentuate the importance of information technologies and communications during the last twenty years to increase the competing profile of a city.

In [8] García et al. described the difference between Smart Objects and Not-Smart Objects and their relation with the Internet of Things (IoT). In their review, they realized that the combination between Smart Objects and the IoT can offer many advantages and improve the people's life because it can interconnect and communicate the different

* Corresponding author.

E-mail address: hajar.khallouki@gmail.com

objects to create more complex applications.

Solanki et al. introduced in [9] an approach for irrigation and highway lamps using IoT in order to preserve energy and resource such as water and electricity. The proposed approach is based on an advanced irrigation system for parks and road side plantation which includes grouping together various peripherals using IoT, and, an advanced highway and high mast lighting system which provides automatic control of the lights of the Highway and High Mast Light.

Smart cities applications must be able to adapt and react dynamically according to the heterogeneous and dynamic environment. They must also be able to adapt dynamically to the utilization context; which represents the relation between the user and the application, such as, its preferences, its physical capabilities, its intentions, etc. And, to the execution environment; which concerns all the information related to the system, such as, user interaction, sensors, networks, etc. [10]

Since the last decade, a lot of approaches have been proposed in order to model devices characteristics and users contexts that are further exploited by multimedia document adaptation processes. We have noticed that some of these approaches provide exclusively a descriptive view of context information (e.g., CC/PP (Composite Capability / Preference Profiles) [11], UAProf (User Agent Profile) [12]), while others propose enhancements with some constraints expressions (e.g., CSCP (Comprehensive Structured Context Profiles) [13], Context-ADDICT [14], SGP (Semantic Generic Profile) [15]). Furthermore, these approaches lack of a dynamic context modeling that is dedicated to describe situation dependent user information and preferences and enables the multimedia documents adaptation.

WURFL (User Agent Profile) [16] is a XML file of description of resources of mobile terminals. This language contains information on the capacities and features of the mobile devices. This project is intended for the adaptation of Web pages on mobile terminals [17]. In [18] a generic model of profile specified in UML was proposed which makes it possible to describe the structure and the semantic of any type of information or user profile. This contribution defines semantic links between elements and integrates the weighting of the elements. This semantic graph is described with a logical directed approach of description [19] via formalisms RDF/RDFS/OWL. However, these models do not make it possible to express actions (e.g. to disable sound).

An adaptive context-aware application for assisting tourists in a smart environment was proposed in [20]. This solution is able to collect not-structured data, belonging to heterogeneous sources and develop recommendations for the user, in order to support a tourist inside a town. In [21] an active learning support system for context-aware ubiquitous learning environments is developed, using contextual information including the location, the current capacity of the learning object, the time available, etc. However, these approaches don't insure the service continuity on mobile devices. In a moving environment, the services can themselves be lost or not be able to function correctly, for example, because of the disappearance of certain resources.

Multidimensional context-aware social network architecture was proposed in [22] to develop a mobile ecosystem to enable context awareness in the development and utilization of mobile crowdsensing applications. Maarala et al. [23] proposed system architecture for providing semantic data and reasoning process with different Semantic Web technologies and methods on context aware IoT environment. Hence, they do not allow a user to express his/her requirements using different interaction modes (e.g. speech).

In [24] Khari et al. proposed a model for secure transmission of data in smart cities. They implemented their model using digital signatures for authentication and triple DES for data transmission. Khari et al. proposed another approach in [25] which focused on Security Classification based on Cloud Layered Framework.

The adaptation, in particular, real time adaptation, raises complex scientific issues as well as new challenges for the execution and the development of applications. For example, the possibility of collecting context information in a dynamic and heterogeneous environment, exploiting heterogeneous infrastructures while benefiting from the opportunities given by these environments, etc, creating a more natural system which takes into account the cultural issues, working environment and physical capabilities.

Our work differs from all of the above in many aspects. Our main objectives are to design an innovative architecture, enabling: (1) a semantic representation and manipulation of multimodal input information, sensed data and services (2) dynamic relevant adaptation services (3) cloud services as a way to insure the service continuity on mobile devices.

III. Multimodal Generic Context-Based Framework for Multimedia Documents Adaptation

Intuitively, people navigate information which generally contains multimedia objects (text, image, audio and video) and interact with multiple devices every day, anytime and anywhere. These corporal interactions fundamentally change how we communicate with devices, because they influence how we process information and thus how we obtain knowledge.

Multimodal interfaces are the scope of our work; they allow users to process and display multimedia documents better. In order to adapt these documents, we propose, in this section, a system which benefits from user multimodal input along with the user context to make it easier to the user to display multimedia content according to their preferences.

A. System Architecture

We first describe the architecture based on a distributed implementation of W3C's Multimodal Architecture and Interfaces [26] applied to ubiquitous computing. The proposed architecture is shown in Fig.1.

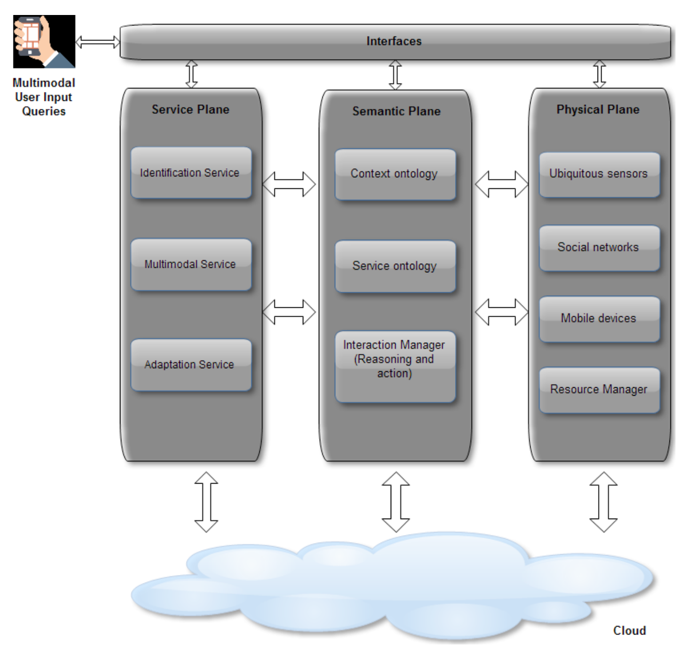


Fig. 1. Architecture description.

1) Physical Plane

The first step of enabling smart services is to collect contextual information about environment, location, device characteristics, user input, etc. For example, sensors can be used to continuously monitor human's physiological activities and actions such as health status and motion patterns. The Resource Manager able to interact with different components coordinates their activities and collect data for the semantic layer. It provides a uniform management of heterogeneous sensors.

2) Cloud Plane

The emergence of mobile devices has made millions of people turn to cloud-based services. The reason behind this shift is the need for seamless syncing of contacts, emails, calendar events, and all kinds of data within a heterogeneous environment comprising different devices, operating systems and applications.

3) Service Plane

The service plane is composed of three components:

- The identification Service is able to identify the user and give him the full access to the platform.
- The multimodal services allow the recognition of the input modalities and present their semantic [27], [28]. There are many multimodal services integrated in devices such as: the HandWriting Recognition (HWR), Speech Recognition (SR) and EmotionML for user emotions recognition.
- The adaptation service allows the transformation of a multimedia object into another multimedia object satisfying a given profile.

4) Semantic Plane

The semantic plane is composed of three main components:

- The context ontology for modeling the user context, such as, their personal characteristics (language, physical capabilities, etc), their preferences, the capacity of their terminal (screen size, battery, etc.), the characteristics of the network (type, bandwidth, etc.), etc.
- The service ontology for modeling the services used in our system, such as, the identification service, multimodal service, etc.
- The Interaction Manager (IM) is the core of the architecture presented in this paper. Its role is the management of the user interactions between the multimodal interface and the user. To access to the platform functionalities, the user must identify himself, for that, the IM sends a request to the identification service. If the user is already identified, the Interaction Manager receives the multimodal input fragments from the user's device and processes them to obtain a meaningful input using the multi-modal services. Thus, the IM starts to analyze the user request and get data from the context ontology. From all the information and data gathered in the analyze step, the IM generates a set of actions, then gets the adapted document from the adaptation services and displays it to the user. If there is no action, the process will start again and wait for another request. When it comes to a low battery level, the data flow could be stored in the cloud as a way to insure the service continuity on mobile devices.

The algorithm (see Fig.2) begins by testing if the user is already logged into the system using *identified()* method. If the user is not identified yet, an identification service will be called using the method *identificationService()*.Once the user is logged, the system allows the user to interact more naturally and gets their requests through the method *multimodalService.getRequest()*. The analysing process is done by *analyse()* which receives as parameters the user request, the multimedia document and the context ontology (described in the next section). Thus, depending on the analysis result, an adaptation service will be called through the method *adaptationService()*.

```
Algorithm : interactionManager()
Inputs   m: Modality //speech, touch, handwriting, typing, etc
         imd : InputMultimediaDocument //image, audio, video, text
output   omd : OutputMultimediaDocument //image, audio, video, text

query= NULL;
owl=NULL; //refers to the ontology used in our system
r=NULL;

if !identified() then
// to check if the user is already identified or not
identificationService();

else
query=multimodalService.getRequest(m);
// get the user request
r=analyse(query, owl, imd); //the result of analysing process
omd=adaptationService(r);
end if

return omd; //return the result of the adaptation
```

Fig. 2. Proposed algorithm.

B. The Proposed Context-Aware Ontology for Multimedia Documents Adaptation

In order to realize the multimodal generic architecture for multimedia documents adaptation, we need to develop an ontology for enabling knowledge sharing and reasoning.

1) The Context Ontology Representation

The first step of building the context ontology is the data collection from different sources, such as the ubiquitous sensors, social networks and mobile devices. The raw resource data is summarized in Table I.

We develop an ontology (Fig. 3) for a semantic modeling of context information using OWL language. OWL is one of the emerging Semantic Web technologies that are endorsed by the W3C for building ontologies [29], [30]. Table II shows some classes of the context ontology and their contextual data.

TABLE I. Raw Resource

Device data	• Hardware data (GPS, Wifi, bluetooth, battery, CPU, etc.) • Software (Microsoft, agenda, etc.)
Sensor data	• Environment data (temperature, humidity, pollution) • Health data (heart rate, blood pressure, stress level, physical capability)
Social data	• Facebook • Twitter • Etc.
Multimodal data	• Haptic • Visual • Auditory
Multimedia data	• Text • Audio • Video • Image

Fig. 3. A snapshot of the context ontology.

TABLE II. CONTEXTUAL DATA

Physical Context	• Location (Home, hospital, street, office, etc.) • Time (Morning, afternoon, etc.) • Environment (temperature, humidity, pollution)
Environment	• Environment data (temperature, humidity, pollution) • Health data (heart rate, blood pressure, stress level, physical capability)
User Context	• Social profile information (language, connection, etc.) • Online/offline social behaviors • Activity (working, meeting, driving, etc.)
User Interaction	• Haptic • Visual • Auditory
Multimedia data	• Text/ Audio/Video/Image
Resource Context	• Hardware (Battery level, device screen size, etc.) • Software

The contextual data is updated automatically depending to environment and situation changes.

The User class contains information about the user. It is divided into two sub classes (see Fig.4):

- The Non_Security class which represents general information such as, the user name, age, gender and physical capability.
- The Security class related to user identification.

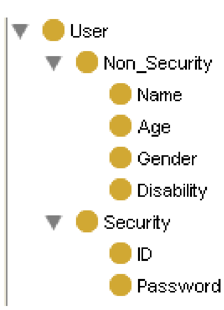


Fig. 4. Subclasses of User class.

In the next section, we introduce the service ontology which is composed of different types of services.

2) The Service Ontology Representation

Fig. 5 presents the services composition in a simple OWL ontology.

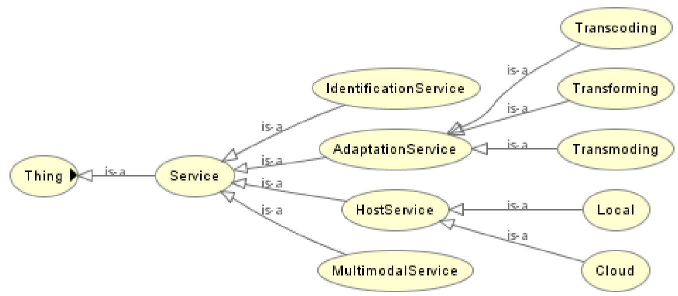


Fig. 5. A snapshot of the service ontology.

- The host service is responsible of migrating contextual data and services to the cloud for further processing in case of a limited computing power of mobile devices. The Local class contains information about fixed devices or mobile devices. The Cloud class contains information about the cloud server that can be used for hosting contextual data and services [31].
- The adaptation service generates the transformation process of a multimedia object into another multimedia object satisfying a given decision.
  - Transcoding: conversion of format, e.g., JPEG to PNG.
  - Transmoding: conversion of types, e.g., text to sound.
  - Transforming: it allows the content change without changing the media type and format, e.g., text summarization, language translation, etc.
- The identification service allows the user to be identified and give them the full access to the system.
- The multimodal service enables the recognition of user multimodal inputs.

3) Case Study on Context Reasoning

In this section, we define rules that can for example be represented using the generic rule languages in Jena reasoner which we intend to use in our prototype.

Example 1: to receive text content if User1 is in a meeting

[rule1: (? Schedule mgo:is-a ?meeting)
(? Location mgo: is-a ?office)
(? MultimediaDocument mgo: is-a ?audio)
➔ (?AdaptationService mgo:is-a ?transmoding)

Example 2: to translate text to English

[rule2: (?UserLanguage mgo:is-a ?Spanish)
(? MultimediaDocument mgo:is-a ?Text)
➔ (?Text "SetLanguage" ?English)]

Example 3: to migrate to the cloud if the battery level is 10%

[rule3: (? location mgo:is-a ?car)
(? batteryLevel mgo:is-a ?10)
➔ (?ContextHost mgo:is-a ?Cloud)]

IV. APPLICATION SCENARIOS: SMART CITIES

The introduced systems may serve people in different application scenarios. We list some of them in the following to illustrate the practicality and benefits of using the developed mobile application.

- Tourism

Traveling somewhere new where you have few to no language skills is a real obstacle, especially when you want to navigate native websites (restaurants, malls, museums, etc.) without any translation. Tourists may benefit from the language translation functionalities without being distracted or even blocked by language barriers in foreign countries.

- International Students Commuting

Many students prefer to study abroad. One of the problems that are internationally faced, language barrier becomes the biggest problem for international students. The students can get information, adapt with the city system and probably display any multimedia document without counting for language barriers. For instance, any student may display any courses videos, these videos can be subtitled automatically to their own language.

- Health

People navigate and display content using different methods depending on their preferences, skills, and abilities. Thus, they can display multimedia contents depending to their physical capabilities. For example, if a blind person receives a text, he would rather receive an audio instead, which means that we need to transform the text to audio.

V. Result and Discussion

We have partially implemented the proposed architecture on Androïd platforms using the AndroJENA framework. This framework allows to maintain OWL descriptions on tablets and smartphones, and to query them with the support of SPARQL queries.

Fig.6 shows screenshots of the speech input queries.

Our work will go beyond previous work in building intelligent multimodal context aware framework for multimedia documents adaptation. Some mobile computing research projects have explored context awareness as a means to improve user interfaces of mobile devices. In our framework, the notion of context awareness goes beyond the basic sensing of how the devices are being used. In our work, context ontology is explicitly represented using ontology language (i.e. OWL) that undergirds the communication and representation and supports preferences sharing, prediction of user mobility and prediction of service use.

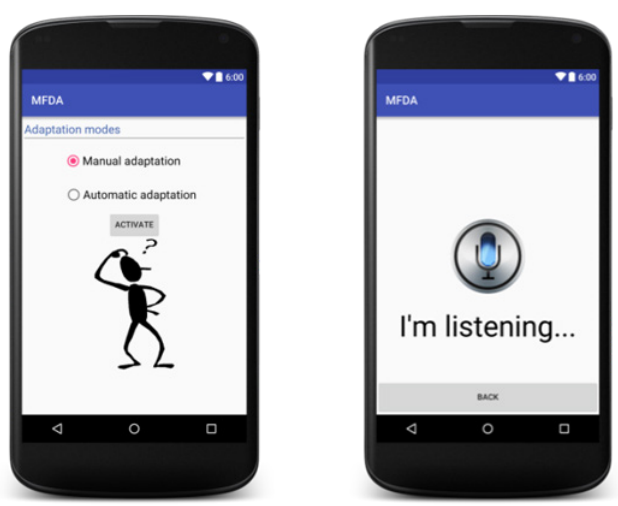


Fig. 6. (a) Screenshot of the adaptation modes interface. (b) Screenshot of the speech recognition.

VI. Conclusion

The heterogeneity of devices and user preferences has raised the problem of multimedia documents adaptation according to user context and condition. This paper applied a semantic multimodal approach to multimedia documents and their adaptation through a multimodal architecture that allows the interaction of user and his devices to satisfy his preferences and according to his context.

As discussed above, there remains more work to be carried out for covering all the aspects of adapting multimedia documents and for improving our semantic and multimodal architecture by integrating a trust model for data sharing in smart cities. Another work aims to build a context-based agent architecture in which the various components are implemented as autonomous agents.

References

[1] Colace F., Moscato V., Quintarelli E., Rabosio E., & Tanca L. Context awareness in pervasive information management. In Data Management in Pervasive Systems (pp. 235-256). Springer International Publishing. (2015).

[2] Khallouki H., Bahaj M., Roose P., Laborie S. SMPMA: Semantic multimodal Profile for Multimedia documents adaptation, In Proceedings of the 5th IEEE International Workshop on Codes, Cryptography and Communication Systems (IWCCCS'14), 27-28 Nov 2014, 142-147, El Jadida, Morocco.

[3] Abascal J., Boanil B., Gardeazabal L., Lafuente A., and Salvador Z., Managing Intelligent Services for People with Disabilities and Elderly People, Lecture Notes in Computer Science, vol. 5615/2009, pp. 623-630, July 2009.

[4] Raman T. V., "Open, Interoperable Means for Integrated Multimodal Interaction", in W3C Workshop on Multimodal Interaction, 2004.

[5] Pous M. & Ceccaroni L. Multimodal interaction in distributed and ubiquitous computing. In Internet and Web Applications and Services (ICIW), 2010 Fifth International Conference on (pp. 457-462). IEEE.

[6] Caragliu A., Del Bo C., & Nijkamp P. Smart cities in Europe. Journal of Urban Technology, vol. 18, no. 2, pp. 65–82. (2011).

[7] Dameri R. P. Smart City Definition, Goals and Performance. In Smart City Implementation (pp. 1-22). Springer International Publishing. (2017).

[8] García C. G., Meana-Llorián D. & Lovelle J. M. C. A review about Smart Objects, Sensors, and Actuators. International Journal of Interactive Multimedia and Artificial Intelligence, vol. 4, no. 3, pp. 7-10. (2017).

[9] Solanki V. K., Venkatesan M. & Katiyar S. Conceptual Model for Smart Cities: Irrigation and Highway Lamps using IoT. International Journal of Interactive Multimedia and Artificial Intelligence, vol. 4, no. 3, pp. 28-33. (2017).

[10] Peyman Oreizy, Michael M Gorlick, Richard N Taylor, Dennis Heimbigner, Gre-gory Johnson, Nenad Medvidovic, Alex Quilici, David S Rosenblum, and Alexander L Wolf. An architecture-based approach to self-adaptive software. IEEE Intelligent systems, vol. 14, no. 3, pp. 54–62, 1999.

[11] Klyne G., Reynolds F., Woodrow C., Ohto H., Hjelm J., Butler M. H., and Tran L., Composite Capability/Preference Profiles (CC/PP): Structure and Vocabularies 1.0, W3C, Recommendation, January 2004.

[12] WAP Forum, User agent profile, Open Mobile Alliance, Specifications, October 2001.

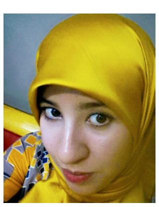
[13] Buchholz S., Hamann T., and Hübsch G., Comprehensive structured context profiles (cscp): Design and experiences,in Proceedings of the Second IEEE Annual Conference on Pervasive Computing and Communications Workshops, ser. PERCOMW'04. Washington, DC, USA: IEEE Computer Society, 2004, pp. 43–48.

[14] Bolchini C., Orsi G., Quintarelli E., Schreiber F. A., and Tanca L., Context Modeling and Context Awareness: steps forward in the Context- ADDICT project, Bulletin of the IEEE Technical Committee on Data Engineering, vol. 34, no. 2, pp. 47–54, 2011.

[15] Dromzée C., Laborie S., and Roose P., A Semantic Generic Profile for Multi-media Documents Adaptation, in Intelligent Multimedia Technologies for Net-working Applications: Techniques and Tools. IGI Global, 2012.

[16] Passani L., WURFL (Wireless Universal Resource File), 2007. http://

wurfl.sourceforge.net.
[17] Veaceslav C., Applying Next Generation Web Technologies in the Configuration of Customer Designed Products, Royal Institute of Technology, Stockholm, 2011.
[18] Chevalier M., Soulé-Dupuy C., Tchienehom P., Semantics-based Profiles Modeling and Matching for Resources Access, Journal des Sciences pour l'Ingénieur, African Journals Online (AJOL), vol. 1, no. 7, pp. 54-63, 2006.
[19] Jouanot F., Cullot N., and Kokou Y., Context Comparison for Object Fusion, Lecture Notes in Computer Science, vol. 2681, p. 1031, 2003.
[20] Colace F., Lemma S., Lombardi M., De Santo M. & Casillo M. A Context-Aware Mobile Solution for Assisting Tourists in a Smart Environment. In Proceedings of the 50th Hawaii International Conference on System Sciences. (2017, January).
[21] Hsu T. Y., Chiou C. K., Tseng J. C., & Hwang G. J. Development and evaluation of an active learning support system for context-aware ubiquitous learning. IEEE Transactions on Learning Technologies, vol. 9, no. 1, pp. 37-45. (2016).
[22] Hu X., Li X., Ngai E. C. H., Leung V. C. & Kruchten P. Multidimensional context-aware social network architecture for mobile crowdsensing. IEEE Communications Magazine, vol. 52, no. 6, pp. 78-87. (2014).
[23] Maarala A. I., Su X. & Riekki J. Semantic Reasoning for Context-aware Internet of Things Applications. arXiv preprint arXiv:1604.08340. (2016).
[24] Khari M., Kumar M., Vij S. & Pandey P. Smart Cities: A Secure Data Transmission Model. In Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies (p. 10). ACM. (2016, March).
[25] Khari M., Gupta S. & Kumar, M. Security outlook for cloud computing: A proposed architectural-based security classification for cloud computing. In Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on (pp. 2153-2158). IEEE. (2016, March).
[26] António Teixeira, Nuno Almeida, Carlos Pereira, Miguel Oliveira e Silva, W3C MMI Architecture as a Basis for Enhanced Interaction for Ambient Assisted Living, W3C Workshop on Rich Multimodal Application Development.
[27] Maes S. H. A VoiceXML framework for reusable dialog components. In Applications and the Internet, 2002. (SAINT 2002). Proceedings. 2002 Symposium on (pp. 28-30). IEEE.
[28] McGlashan S., Burnett DC., Carter J., Danielsen P., Ferrans J., Hunt A., Lucas B., Porter B., Rehor K., Tryphonas S. Voice Extensible Markup Language (VoiceXML 2.0). W3C. (2004).
[29] Smith M., Welty C., and McGinness D. Web Ontology Language (OWL) Guide Version 1. 2003.
[30] Harmelen F., Hendler J., Horrocks I., McGinness D., Patel-Schneider P., and Stein L. Web Ontology Language (OWL) Reference Version 1.0. 2003.
[31] Alti A., Achouri M., Derdour M. & Roose P. Smart semantic-based approach for mobile applications in pervasive environments. In 2016 International Conference on Information Technology for Organizations Development (IT4OD) (pp. 1-6). IEEE.



H. Khalloukii

H. Khallouki born in 1989, she completed her Master's degree in computer science from Faculty of Sciences, Hassan II University, Casablanca, Morocco. In pursuit of a PH.D, she joined the Department of Mathematics and Computer Sciences, Faculty of Sciences and Technology, Hassan I University Settat, Morocco, in 2013. Her actual main research interests concern multimedia documents adaptation, context awareness and semantic web.

M. Bahaj

M. Bahaj is a full professor in the Department of Mathematics and Computer Sciences from the University Hassan 1st Faculty of Sciences & Technology Settat Morocco. He has published over 80 peer-reviewed papers. His research interests are intelligent systems, ontologies engineering, partial and differential equations, numerical analysis and scientific computing. He is Editor-in-Chief for Australasian Journal of Computer Science and associate editor of Journal of Artificial Intelligence and Journal Software Engineering.

Two Hand Gesture Based 3D Navigation in Virtual Environments

I. Rehman*, S. Ullah, M. Raees

University of Malakand, Department of CS & IT, Chakdara (Pakistan)

Received 30 April 2018 | Accepted 22 June 2018 | Published 6 July 2018

Abstract

Natural interaction is gaining popularity due to its simple, attractive, and realistic nature, which realizes direct Human Computer Interaction (HCI). In this paper, we presented a novel two hand gesture based interaction technique for 3 dimensional (3D) navigation in Virtual Environments (VEs). The system used computer vision techniques for the detection of hand gestures (colored thumbs) from real scene and performed different navigation (forward, backward, up, down, left, and right) tasks in the VE. The proposed technique also allow users to efficiently control speed during navigation. The proposed technique is implemented via a VE for experimental purposes. Forty (40) participants performed the experimental study. Experiments revealed that the proposed technique is feasible, easy to learn and use, having less cognitive load on users. Finally gesture recognition engines were used to assess the accuracy and performance of the proposed gestures. kNN achieved high accuracy rates (95.7%) as compared to SVM (95.3%). kNN also has high performance rates in terms of training time (3.16 secs) and prediction speed (6600 obs/sec) as compared to SVM with 6.40 secs and 2900 obs/sec.

Keywords

Human Computer Interaction, Virtual Reality, Augmented Reality, Gesture Recognition, 3D Navigation.

DOI: 10.9781/ijimai.2018.07.001

I. Introduction

The effective and realistic interaction is a prerequisite to attain/explore physical/behavioral properties of a Virtual Environment (VE) and its constituent objects. The interaction in VEs ranges from keyboard and mouse where button clicks (key/mouse) carry no information [1], to touch based, which uses 2D gestures with finger/hand actions via touchscreens/interactive panels, and recently to touchless (natural) interfaces which involves 3D gestures in midair [1]. Hand gesture recognition plays a vital role in various applications such as sign language recognition [2-5], virtual and augmented reality [6-10], robotics [11-13], physical (health) sciences [14-16], natural sciences [17], computer engineering [18, 19], and industrial areas [20, 21] etc.

Gesture based interaction is the most attractive form of natural interaction due to the exclusion of physical contact with hardware where input/interaction devices need to be permanently connected (via some physical means) with computer. These physical media (cables) are used for delivery of input to the computer. These media lead to extra burden, need space, extra cost and complexity. Gesture based interaction offers more natural and intuitive HCI with various multimodal forms [22-26].

Different sensors have been used in computer vision and image processing for recognition purposes, such as Koller et al. [27], who used monocular camera for tracking in augmented reality applications. Jalal et al., presented different systems for Human Activity Recognition (HAR) based on depth video [28, 29] and depth imaging [30, 31]. Some authors proposed, Depth Images-based Human Detection and Activity Recognition [32], and Human Pose Estimation and Recognition from RGB-D Video [33].

Feature selection has a vital role in any recognition system. For face recognition different methods have been used. In Holistic Matching method [34, 35], the complete face area is taken as input to the recognition system. In Feature based method, the position and statics of nose, eyes and mouth are considered as input to the system. The Hybrid method uses the combination of both Holistic and feature based methods [36].

For action recognition, Lowe [37] introduced Scale Invariant Feature Transform (SIFT). Dense Sampling Scale Invariant Feature Transform (DSIFT) was used by different authors for action recognition [38-48]. Histogram of oriented gradient (HOG) was used by [41, 47-55]. Shape context (SC) was proposed by Belongie and Malik [56] for feature extraction and also was used by Wang et al. [57], Gupta et al. [51], and Yao and Fei-Fei [58]. For recognition of action from still images, GIST was proposed by Oliva and Torralba [59] and used by Gupta et al. [51], Prest et al. [60], and Li and Ma [48]. The Speeded Up Robust Features (SURF), is proposed by Bay et al. [61] and used by Ikizler et al. [62] to represent the human silhouettes for action recognition.

Different techniques are proposed by various authors for gait recognition. Ahmed et al. [63], used horizontal and vertical distances of selected joint pairs. Andersson et al. [64], calculated mean and standard deviation in the signals of lower joint angles. Ball et al. [65] used mean, standard deviation and maximum of the signals of lower joint angles. Dikovski et al. [66] proposed a set of seven different features, such as joint angles and inter-joint distances aggregated within a gait cycle, body parameters, along with various statistics. Kwolek et al. [67], used static body parameter, bone rotations, and the person's height. Preis et al. [68], presented 13 pose attributes. Sinha et al. [8], used multiple gait features, upper and lower body area, inter-joint distances and other

* Corresponding author.

E-mail address: inam.btk@gmail.com

features [65, 68]. Skin Joints Features were also proposed by [69, 70] for gait recognition.

Different approaches are proposed for hand recognition such as mount based sensors [71], multi-touch screen sensors[72, 73] and vision based sensors [74-76]. The depth based hand gesture recognition have three types i.e. static hand gesture recognition [77], hand trajectory gesture recognition [78, 79] and continuous hand gesture recognition [80, 81]. Most of the authors used computer vision and image processing methods [82-85], along with some newly introduced input devices such as Leap Motion [86-88] and Kinect [89]. For a natural interaction with AR environments, fiducial markers are used with fingers [90, 91]. Different computer vision techniques are used for detection of hand and fingertips for AR interaction [92-94]. These systems are commercially limited due to problems such as skin color and precise depth sense [94]. Different glove based techniques have been used for accurate interaction [95-97] but limited due to its cumbersome nature.

For gesture based navigation, different systems have been proposed so far, but have limited commercial application due to cumbersome or inaccurate nature, cost, or dependency/need for special devices and their limited range. Recent research mostly stresses to deal with these problems, but simple and intuitive interaction is still the major area than needs to be improved. The previous gesture based navigation techniques, mostly rely on the coarse/unrealistic alteration/shape of hands and fingers layout for transition among different gestures which results in an increased physical and mental load on the user.

In this paper, we propose a novel two hand gesture based 3D navigation technique for VEs with the objective of providing intuitive and easy navigation. Navigation includes 3D movement, i.e. forward, backward, up, down, left, and right along with an effective speed control mechanism. Computer vision techniques are used for detection of gestures (colored thumb, fingers) from the real scene while OpenGl is used as a front end for navigation in the VE. Machine learning tools such as SVM and kNN are used to assess the accuracy and performance of the proposed gestures.

The rest of the paper is organized as follows: section 2 presents related work, section 3 describes the proposed system, section 4 consists of experiments and evaluation and finally section 5 is related to conclusion and future work.

II. Related Work

In daily life communication, hand gestures cover the gap of merely verbal information, and so is a necessary part of effective and meaningful communication with the receiver. In HCI, hand gesture based interaction is more valuable due its natural and attractive nature. Hand gestures refer to meaningful movement of the hand and fingers[86], which entails most valuable information[98].

In the past different navigation techniques in VEs have been proposed [99, 100]. Different types of sensing techniques have been used for recognition of patterns in gestures [101]. Different sensing systems have been proposed so far, such as glove based, vision based, along with some newly introduced devices such Leap Motion [102] and Kinect [89].

Glove based devices use movement based approach with high performance in some applications such as recognition of sign language [103]. CyberGlove, a type of data gloves is used for tracking hand gestures [104]. Cooper et al. [105] used color coded glove for tracking of hand movement, but the system needs wearable gloves which decrease user's experience in the environment. Kim et al. [106], used a Cyber Glove, a wearable device for recognition of hand gestures and performed different navigation tasks as shown in Fig. 1.

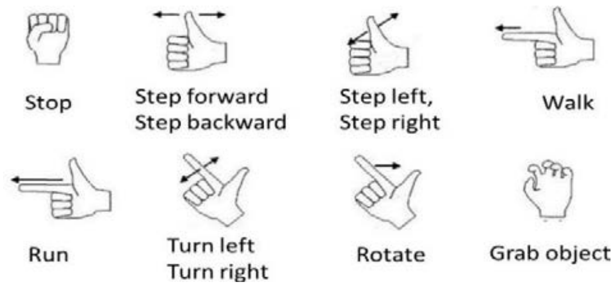


Fig. 1. Gestures with Cyber Glove.

Although wearable devices were used mostly for gesture based interaction in the past, the cumbersome and costly nature of gloves, limit its widespread use in HCI [101].

Chen et al. [107], used computer vision for the detection of hand gestures from the video taken by a webcam. Two types of hand gestures were proposed i.e. appearance based which used bare hand and marker based (with colored markers on a black glove) 3D hand model. For moving a virtual car, different gestures were proposed as shown in Fig. 2. The system is unable to provide complete 3D movement and speed control. The system has less degree of correspondence with real world navigation.

a. Two fingers b. Palmc c. Little finger

Fig. 2. Gestures for virtual car navigation [107].

Krum et al. [108], presented a navigation interface (earth 3D visualization) using verbal and hand gestures. The system used image processing techniques for detection of hand gestures taken from a Gesture Pendant video camera. Multiple infrared emitting LEDs were used for illuminating the hand gestures in front of camera. Different types of hand gestures were used for navigation.

Shao Lin [109], used Leap Motion controller for detection of hand gestures. The proposed technique used both hands for three different types of gestures as shown in Fig. 3. Similarly clockwise and anti-clockwise rotation of hands produced the same rotation in the VE but the technique has no mechanism for up, down movements and speed control.

Kerefeyn et al. [110], used different gestures for controlling and manipulation of virtual objects in a VE using Leap Motion controller. The system proposed five different gestures for interaction with virtual objects using right hand (see Fig. 4). The system does not provide complete 3D navigation i.e. forward, backward, up, down, left, and right as well as there is no mechanism for speed control.

Khundam et al. [111], proposed single hand gestures via Leap Motion controller for navigation in VE as shown in Fig. 5. Forward movement is done by raising the hand with straight palm. While backward movement is achieved via turning palm direction or reversing hand facing. Pushing the palm to left causes step movement to left side while its movement to right results in step right. Grasping or moving out of display area causes hold position. Forward speed depends on advance movement of palm while pulling of palm to body controls backward speed. More or less movement towards left or right side causes more or less speed. The proposed gestures are hard to learn.

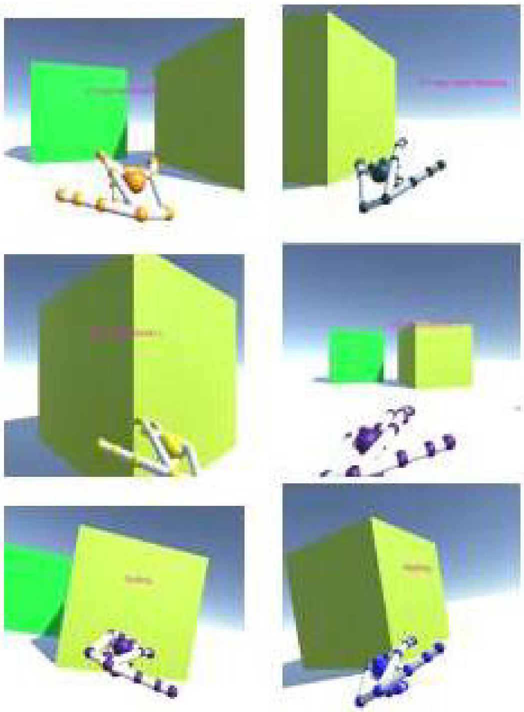


Fig. 3. Gesture based navigation by Shao Lin [108].

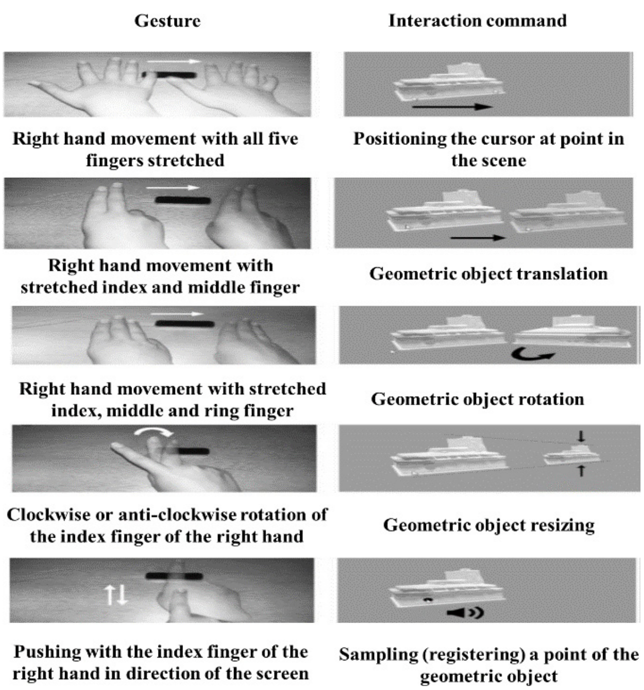


Fig. 4. Gesture based interaction by Kerefeyn et al. [110].

Fig. 5. Hand gestures proposed by Khundam et al. [111].

Batista et al. [112], used Leap Motion controller and proposed different hand gestures as shown in Fig. 6, for controlling an avatar in a virtual tour but there is no mechanism for up/down movement and turning left/right.

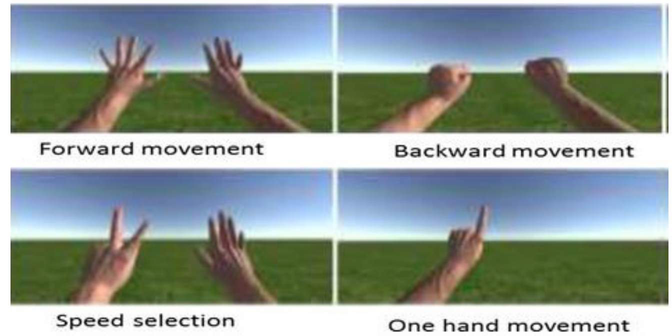


Fig. 6. Gestures proposed by Batista et al. [112].

Liang et al. [113] presented a system using Leap Motion controller for the detection of hand gestures. There were different types of modules for navigation such as single hand gestures, designed for children which causes avatar in VE to fly left, right, up, and down as shown in Fig. 7.

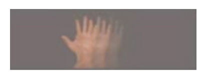
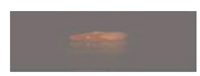
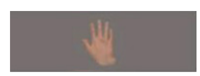
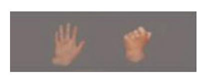
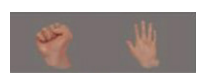
Hand gesture	Movement	Target action
	Move right	Fly to the right
	Move left	Fly to the left
	Move down	Fly down
	Move up	Fly up
	Stretch	Hover
	Stretch to grip	Grasp stick
	Grip to stretch	Drop stick

Fig. 7. Hand gestures proposed by Liang et al. [113].

Leap Motion controller is unable to detect all fingers of hand, specially middle and pinky finger, moreover, it has a limited working space [109]. Other problems include misdetection during overlapping of hands [114], crossing of field boundaries, and varying lighting conditions [112]. Leap Motion gives inaccurate results for hands beyond 250mm upside the controller as well as it gives unstable sampling frequency [115]. A comparative study conducted by [114] states that Kinect provides inaccurate but comprehensive data while Leap Motion gives comparatively accurate results. Nabiyouni et al. [114] stated that Leap Motion fails to recognize cross over fingers or if they are next to each other. Rotation of palm or fingers more than 80 degrees causes failure in tracking. Moreover it produces significant fatigue.

Kinect is used for detection of full human body gestures to interact with virtual objects [117]. Kumerburg et al. [116] used Microsoft Kinect as input device for navigation in a VE. Different gestures for navigation were proposed such as raising both arms upside for fly and forward movement as shown in Fig. 8. The proposed navigation gestures were hard to learn and use.

Gesture	Command	Characteristic
	Tilt up	Bend backward with hands down
	Turn left	Raise the left hand above shoulder sideways with the right hand down
	Pan left	Move left hand across the body and right hand down
	Fly, move foward	Raise both arms up
	Fly pan right	Raise both arms and move right hand across head

Fig. 8. Gestures proposed by Kumerburg et al. [116] using Microsoft Kinect.

Vulture et al. [118], also used Microsoft Kinect and proposed gesture for navigation using both arms as shown in Fig. 9.

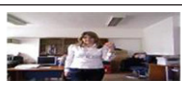
Gesture	Characteristic
	Foward navigation
	Backward navigation
	Navigation to the right
	Navigation to the left
	Up to a higher level
	Down to a lower level

Fig. 9. Gestures proposed by Vulture et al. [118].

The previous techniques for gesture based navigation mostly depend on variant shapes/layouts of hands, fingers, and arms which lead to extra mental and physical load on user in learning and usage. They are limited in use due to less realistic nature.

III. Proposed System

We propose a new, two hand gesture based navigation technique for VEs with a close resemblance to car steering driving. The relative position of both thumbs determines various gestures. These gestures are used for 3D navigation in the VE. Navigation includes Forward, Backward, Up, Down, and Left and Right side movement. The green and yellow color caps made of paper/rubber are used with thumbs. The VE consists of different 3D objects as shown in Fig. 10. Identification of different gestures leads the virtual object/camera to navigate accordingly.

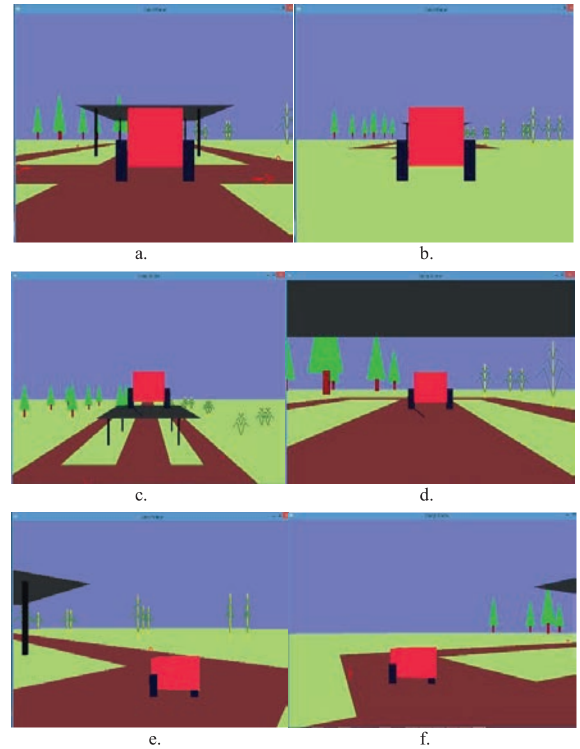


Fig. 10. Scenario of VE (from different perspectives). (a). forward movement, (b). backward movement, (c). Upward movement, (d). Downward movement, (e). Right turn, (f). Left turn.

A. System Architecture

The proposed system uses OpenCV as backend and OpenGl as frontend tool. OpenCV is used for image processing, consists of different phases such as image acquisition, conversion of image to HSV (Hue, Saturation, Value), thresholding for green and yellow colors, and finally the calculation of 2D position (x, y) of the specified colors as shown in Fig. 11. OpenGL is used for designing and interaction of the VE. OpenGL, based on position of colors (Green and Yellow), identifies various gestures that leads to different navigation tasks in the VE.

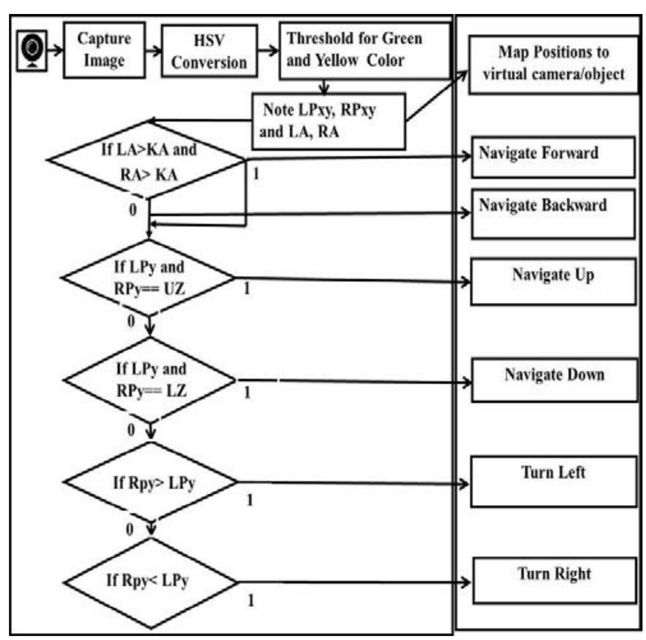


Fig. 11. (a) Backend (Image processing), (b) Frontend (VE).

First of all OpenCV performs image acquisition via a camera. The image is then converted to HSV for realistic performance which is then thresholded for Green and Yellow colors. In the last stage, position of both colors is calculated dynamically from the image. OpenGL receives the positions of both thumbs colors. On the basis of these positions, different gestures are identified which lead to 3D navigation in the VE.

OpenCV performs image acquisition via ordinary webcam. Finger caps of green and yellow colors are used for left and right thumbs of both hands. Skin color is omitted to achieve best results as it varies from person to person. A rang of Hue, Saturation and Values are selected for green and yellow colors to detect thumbs in stable lighting conditions. First the region of interest of the image (RI_img) is extracted from the Frame Image (F_img) to avoid false detection of the background green and yellow colors. The IR_img is then segmented from the F_img based on the skin area which is the most probable area to get the thumb fingers. YCbCr model has the capability to distinguish between skin and non-skin colors [119].

$$F - Img - YCbCr \begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.1 & 128 & 24.8 \\ -37.3 & -74 & 110 \\ 110 & -93.2 & -18.2 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

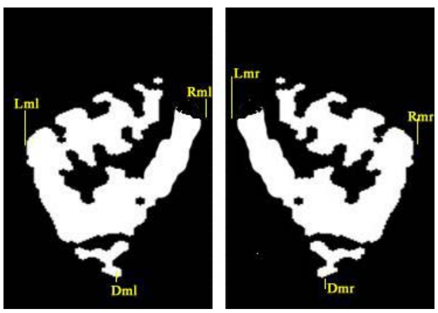


Fig. 12. Binary images of both hands after conversion to YCbCr.

After getting the binary image of F-Img (see Fig. 12), *RI-Img* with rows *'m'* and column *'n'* is extracted from *F-Img* using algorithm [120] as,

$$RI\_Img(m,n) = \left(\bigcup_{r=Dml}^{F\_Img.Row(0)} (F\_Img) , \bigcup_{c=Lml}^{F\_Img.Colum(Rml)} (F\_Img) \right) \quad (2)$$

Where *Lml*, *Rml* and *Dml* represent Left-most, Righ-most and Down-most skin pixels of the left hand.

$$RI\_Img(m,n) = \left(\bigcup_{r=Dmr}^{F\_Img.Row(0)} (F\_Img) , \bigcup_{c=Lmr}^{F\_Img.Colum(Rmr)} (F\_Img) \right) \quad (3)$$

Where *Lmr*, *Rmr* and *Dmr* represent *Left-most*, *Righ-most* and *Down-most* skin pixels of the right hand.

The segmented image RI-Img is then thresholded for green and yellow color simultaneously, using HSV color space (see Fig. 13).

$$RI\_Img(x,y) = \begin{cases} 1 \text{ if } RI\_Img.H(x,y) \geq 63 \wedge RI\_Img.H(x,y) \leq 98 \\ 1 \text{ if } RI\_Img.S(x,y) \geq 98 \wedge RI\_Img.S(x,y) \leq 255 \\ 1 \text{ if } RI\_Img.V(x,y) \geq 37 \wedge RI\_Img.V(x,y) \leq 255 \\ 0, \text{ Otherwise} \end{cases} \quad (4)$$

$$RI\_Img(x,y) = \begin{cases} 1 \text{ if } RI\_Img.H(x,y) \geq 65 \wedge RI\_Img.H(x,y) \leq 80 \\ 1 \text{ if } RI\_Img.S(x,y) \geq 110 \wedge RI\_Img.S(x,y) \leq 255 \\ 1 \text{ if } RI\_Img.V(x,y) \geq 145 \wedge RI\_Img.V(x,y) \leq 255 \\ 0, \text{ Otherwise} \end{cases} \quad (5)$$

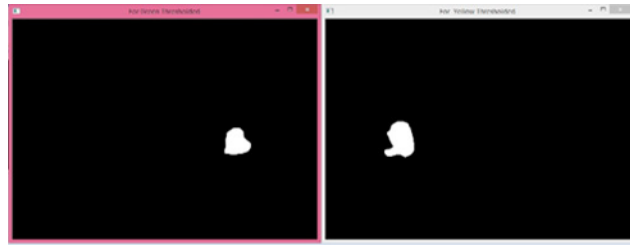


Fig. 13. Detected thumbs of both hands.

B. Navigation

Navigation is the movement towards the desired position in a VE. OpenCV computes positions of both thumbs in 2D. The z-axis movement is deduced from the area variation of the detected thumbs. As the area of the thumbs increases with inward and decreases with outward movement in z-axis (towards the camera eye) as shown in Fig. 14 (a, b). So increase from a predefined (threshold) area KA results in forward navigation while decrease results in backward navigation. The value of KA is half of the fully detected thumb area (near to camera eye) which divides the navigation space (z-axis) into two zones i.e. Forward and Backward zone, as shown in Fig. 15. For accurate navigation in the VE, it is necessary that both thumbs should be visible to camera. Movement of both thumbs forward (where detected thumb area >KA) towards camera eye results in forward navigation while moving thumbs backward (where detected thumb area <KA) produces backward navigation.

LPx and LPy represent position of left hand thumb (with green cap) in x and y-axis, RPx and RPy are position of right hand thumb (with yellow color) in x and y-axis, UZ is the upper zone, LZ is the lower zone, and LA and RA represent the detected thumb areas.

The algorithm for forward and backward movement is given below:

If(LA>KA AND RP>KA)
Forward navigation
If(LA<KA AND RA<KA)
Backward navigation

Fig. 14. (a) Detected (Left and Right hand) areas in forward navigation.

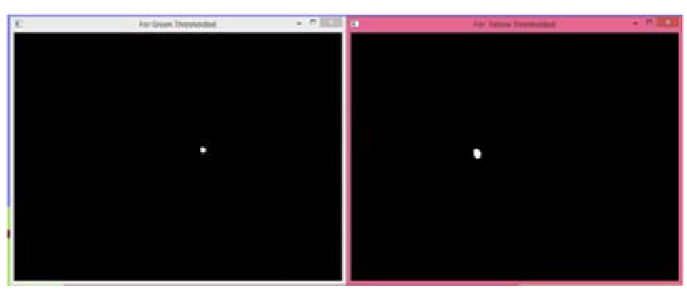


Fig. 14. (b) Detected (Left and Right hand) areas in backward movement.

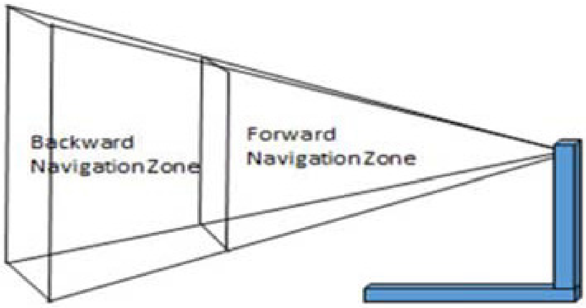


Fig. 15. Forward and backward navigation zones.

For navigation in y-axis, the thumbs of both hands use UZ or LZ. If both thumbs simultaneously move to upper zone (UZ), upward navigation (along y-axis) is produced, while at lower zone (LZ), it leads to downward navigation as shown in Fig. 16.

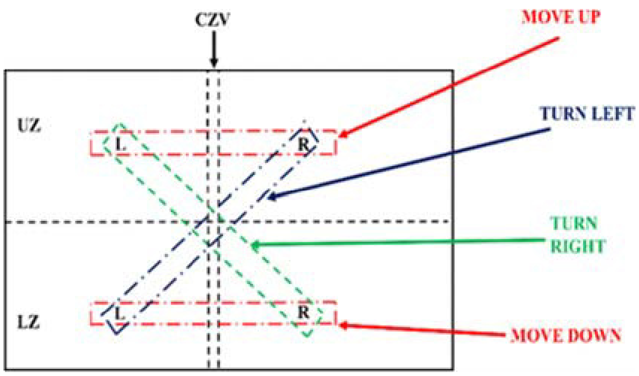


Fig. 16. Navigation space for gestures.

RPx,y is the right hand position (x,y), LPx,y left hand position (x,y), UZ upper zone, UL lower zone, and CZV is central zone vertical.

The algorithm for moving in vertical (y-axis) direction is given below:

```
If(RPy== UZ AND LPy== UZ)
Upward movement (along +ve y-axis) (see Fig.  17.a)
If(RPy== LZ AND LPy== LZ)
Downward movement (along -ve y-axis) (see Fig.  17.b)
```

a. b.

Fig. 17. Gestures for movement (a) Up (b) Down.

C. Speed Control

Speed control (Sp) is an important requirement in any navigation technique. In the proposed technique, speed can be controlled via changing the horizontal (x-axis) distance between both thumbs in navigation and turning. The speed remains normal for maximum distance while it increases with decrease in relative distance between both thumbs i.e. the speed is inversely proportional to the relative distance between both thumbs.

Mathematically it can be written as:

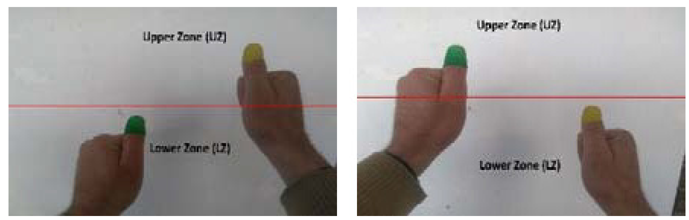
$$Sp \propto 1/abs(RPx - LPx) \tag{6}$$

D. Turning

Left or right turn is taken by simply comparing the y-position of both thumbs. For left turn, the right hand thumb will be in upper zone (UZ) while left hand thumb will be in lower zone (LZ). For right turn, the right hand thumb will be in lower zone (LZ) while left hand thumb will be in upper zone (UZ) as shown in Fig. 16.

The algorithm for turning right and left is given below:

```
If(RPy> LPy)
Turn left (see Fig. 18.a)
If(RPy< LPy)
Turn right (see Fig. 18.b)
```



a. b.

Fig. 18. Gestures for Turning (a) Left (b) Right.

E. System Implementation

The proposed system was implemented using Visual Studio 2013 with corei3 laptop having 1.7 GHz processor, 4GB RAM, and 640x480 resolution low cost built in camera. The OpenCV, as a backend tool, performs different image processing tasks such as acquisition of image from camera, identification of thumbs colors (green and yellow), and dynamic area and pose calculation of these colors. OpenGL, as a frontend tool is responsible for creation and interaction with VE as shown in Fig. 19. The system allows users to interact (navigate) with VE using his/her both hands thumbs having colored (green, yellow) caps. The left hand thumb uses green color while right hand thumb uses yellow color cap. The combination and relative position of both colored thumb makes different gestures which are used to perform 3D navigation in the VE.

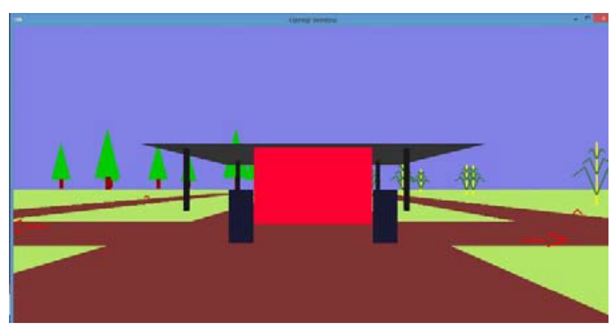


Fig. 19. Screen shot of the experimental scenario.

IV. Experiments and Evaluation

We performed objective and subjective evaluation to assess the accuracy and effectiveness of the proposed navigation technique in VEs. We also used machine learning models i.e. SVM and kNN to assess the accuracy and performance of the proposed gestures.

A. Protocol and Task

Forty (40) volunteer male students participated in the experimental study. Their ages were in the range of 22 to 35 years. All of them had gaming experience with keyboard, mouse, and touch screen but had no experience with gesture based VEs. In the training phase all the students were demonstrated about the use of proposed system and gestures. After that they performed different navigation tasks (Forward, Backward, Up, Down, and turning Left, and Right) using the specified gestures in the VE. Each student performed five pre-trails of each navigation task. After the training session, each participant performed four trails of three different navigation tasks in three different lighting conditions making a total of 1440 trails.

1) Interaction Routes

The experimental environment consisted of three different routes from start to stop position as shown in Fig. 20.

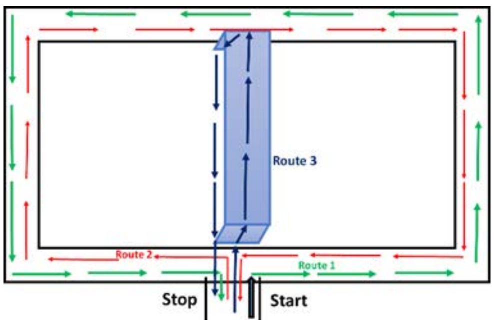


Fig. 20. Complete 2D route model in VE.

2) Interaction task

The students need to perform four (4) trials of each of the following three tasks:

- Task 1. The first task was to follow the Route 1 which covers forward navigation (five times), one right turn and four left turns as shown in Fig. 20.
- Task 2. The second task was to follow the Route 2, which covers five forward movements, one left and four right turns.
- Task 3. The third task was to follow the Route 3 which covers one upward, two forward, one downward, and one backward movement.

3) Lighting conditions

The proposed system uses colored thumbs for interaction with the VE. The detection thumb is highly dependent on the surrounding light. So we performed all the tasks in three different lighting intensities.

- Low light intensity (3-7 Lux) Dark limit of civil twilight under a clear sky [128].
- Medium light intensity (50-70 Lux) Family living room lights (Australia, 1998) [129] to office building hallway/toilet lighting [130-131].
- High light intensity (200-500 Lux) Office lighting [129] [132-133], sunrise or sunset on a clear day.

After performing the tasks, task completion time (TCT) and errors for each task were recorded for objective analysis. Misdetection and deviation from the specified route were considered as errors. Finally, each participant filled a questionnaire for subjective analysis.

B. Result Analysis

In this section, we performed objective analysis (task completion time and errors) and subjective analysis (questionnaire) to assess the accuracy and effectiveness of the proposed navigation technique in VEs.

Objective Analysis

1) Task Completion Time and Errors

The Mean and SD of time and errors for task 1 is shown in Fig. 21 and 22. The Mean time and SD is (89.01, 20.02), (82.30, 19.19), and (61.98, 15.16) for low, medium, and high lighting intensity respectively. The Mean and SD of errors is (3.31, 1.49), (2.19, 1.37), (1.71, 0.97) for low, medium, and high light intensities. The Mean for both time and errors are minimum for high lighting intensity as compared to other. It means that the technique has good performance in higher lighting intensity.

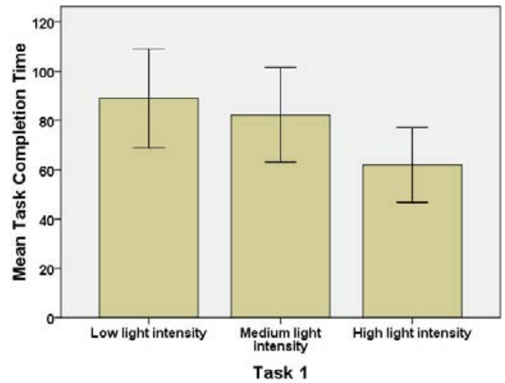


Fig. 21. Mean and SD of time for Task 1.

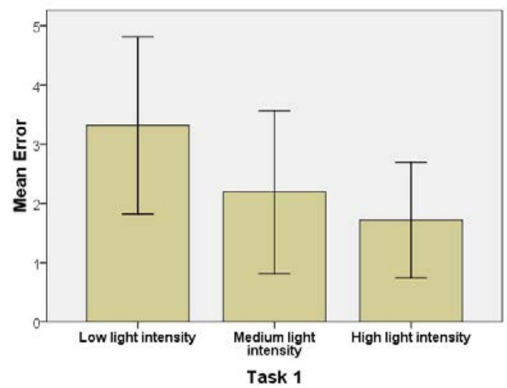


Fig. 22. Mean and SD of errors for Task 1.

The Mean and SD of time and errors for task 2 is shown in Fig. 23 and 24. The Mean time and SD is (68.24, 17.80), (62.53, 15.00), and (51.48, 14.06) for low, medium, and high light intensity respectively. The Mean and SD for errors is (2.31, 1.33), (1.96, 1.14), (1.67, 1.02) for low, medium, and high light intensities. The Mean and SD for both time and errors decreases with increase in lighting.

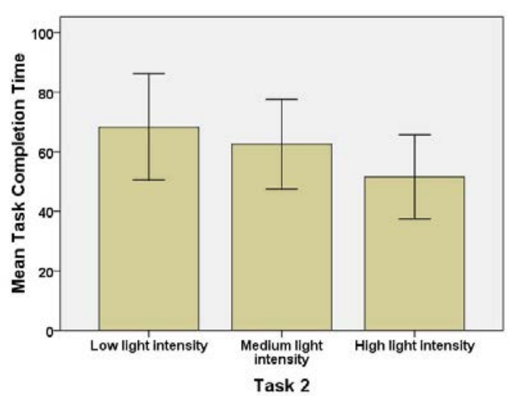


Fig. 23. Mean time and SD for Task 2.

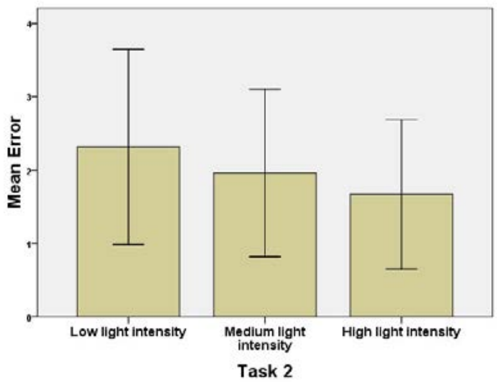


Fig. 24. Mean error and SD for Task 2.

The Mean and SD of time and errors for task 3 is shown in Fig. 25 and 26. The Mean and SD of time is (62.01, 14.11), (56.28, 13.58), and (50.24, 13.89) respectively for low (3-7 Lux), medium (50-70 Lux), and high (200-500 Lux) light intensity.

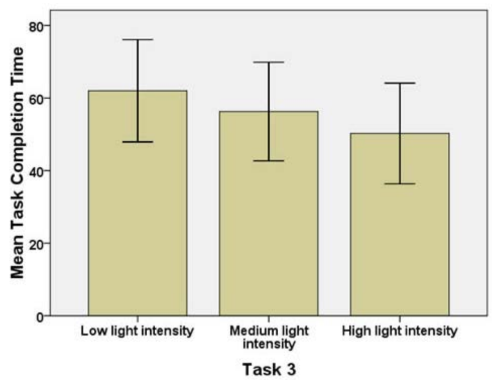


Fig. 25. Mean time and SD for Task 3.

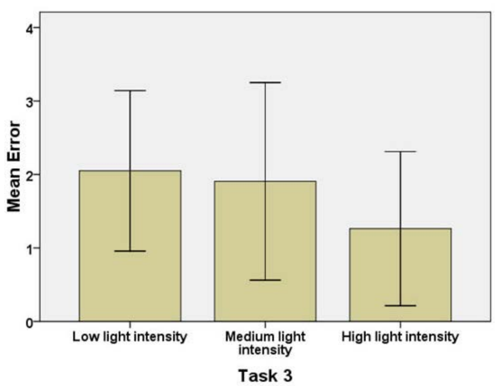


Fig. 26. Mean error and SD for Task 3.

2) Task Learning

The mean task completion time of each task is shown in Fig. 27, 28, and 29. The results show that the mean time decreases with each successive trial which leads to improved task learning. So it means that task learning improves with experience. The system intrinsically works better in high lighting conditions, while the repetition of task improves learning of task performance. The enhanced learning effect is obvious due to simple and realistic nature of gestures.

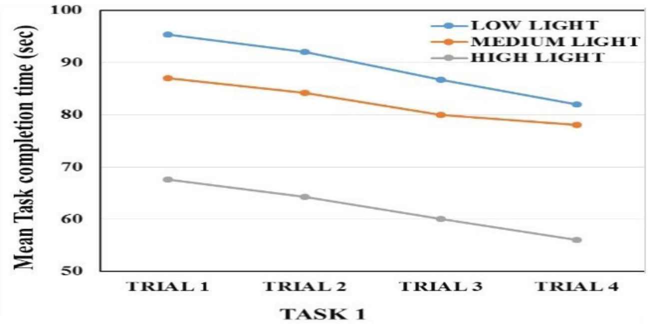


Fig. 27. Mean time of each trial for Task 1.

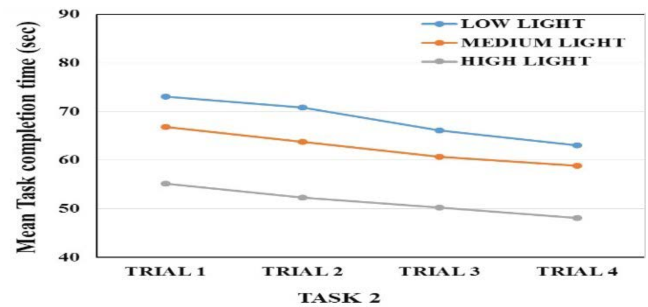


Fig. 28. Mean time of each trial for Task 2.

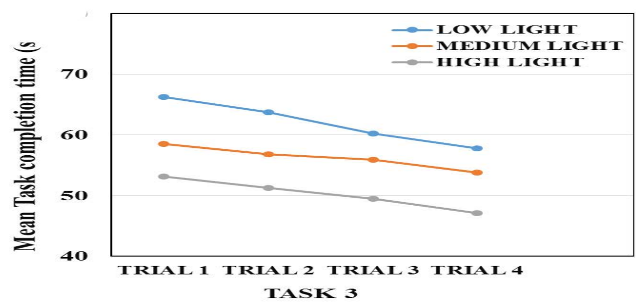


Fig. 29. Mean time of each trial for Task 3.

Subjective Evaluation

For the subjective evaluation, each student filled a questionnaire. The questionnaire was consisted of different questions related to the following topics:

1. Cognitive load during interaction
2. Technique learning
3. Fatigue during interaction

Likert scale is used for scaling purposes where 1 is for lowest level and 5 for highest level.

1) Cognitive Load During Interaction

The main goal of the proposed technique is to use simple and realistic alignment of fingers or hands in order to get an easy to learn and use interaction. The responses for the question concerned to cognitive load on users during interaction is shown in Fig. 30. The results show that most of the students opted for lowest (52.5%) and low level (35%) of mental load created during interaction.

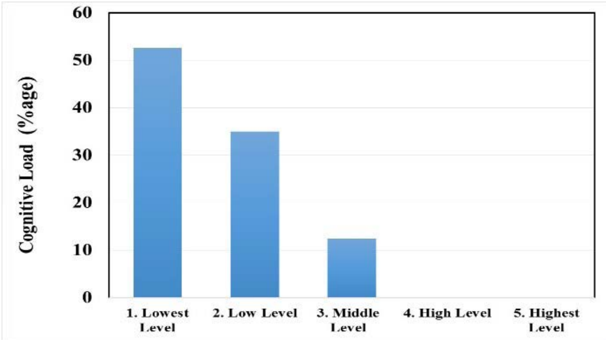


Fig. 30. Cognitive load on user during interaction in VE.

2) Technique Learning

The navigation technique is developed with the aim of ease in learning. Most of the students selected the proposed technique as easy to learn as shown in Fig. 31. This figure shows that 35% and 40% students opted the highest and high level learning. We used real word phenomena/experience for the proposed navigation technique.

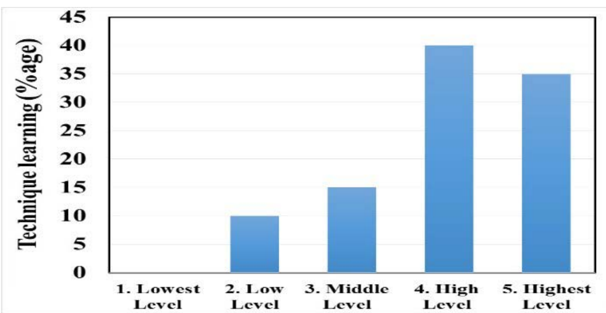


Fig. 31. Technique learning.

3) Fatigue

The students responses concerning the created fatigue during interaction is shown in Fig. 32. The results show that there is a considerable fatigue (40% students opted for high level and 10% for highest) developed during interaction. So we can conclude that fatigue (physical) increases gradually with time in restless operations while the main reason for creation of fatigue is due to baseless/midair operation of hands gestures for long times.

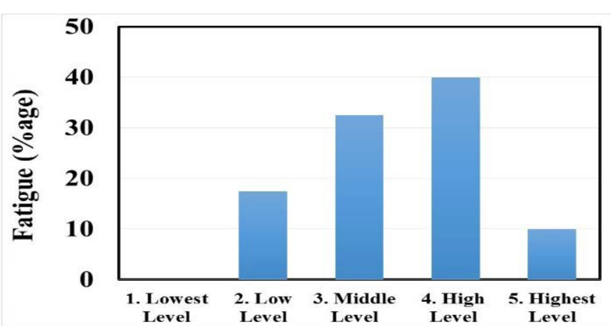


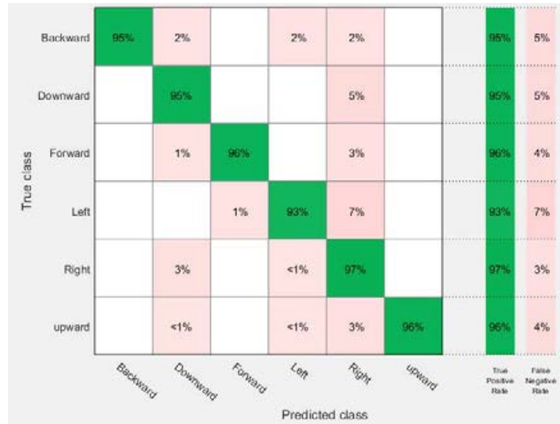
Fig. 32. Fatigue in interaction.

Performance Evaluation of Gesture Recognition

Various authors used different gesture recognition engines for recognition such as HMM [121-124], Advanced HMM [125], Combined HMM and SVM [126], and kNN [127]. We used SVM and kNN models to assess the accuracy of the proposed gestures. To evaluate the performance of the proposed system, 20 participants were selected to perform 10 trials of each of the six (6) different gestures i.e. Forward, backward, upward, downward, left and right. It makes a total of 1200 gestures.

The confusion matrix for SVM (see Table I) shows 95.3% mean accuracy of the proposed system with 97% (maximum) accuracy for right gesture and 93% (minimum) accuracy for left gesture.

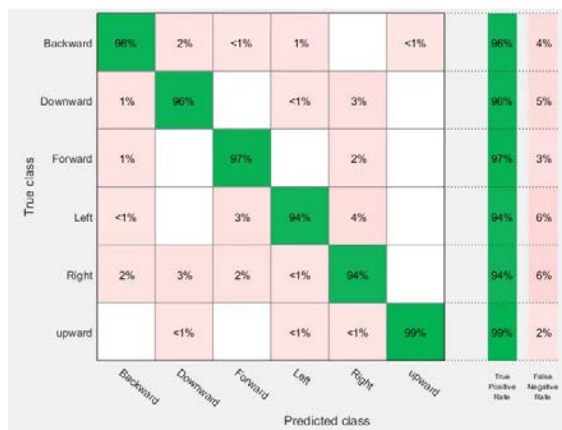
TABLE I. Confusion Matrix For Svm



True class \ Predicted class	Backward	Downward	Forward	Left	Right	upward	True Positive Rate	False Negative Rate
Backward	95%	2%		2%	2%		95%	5%
Downward		95%			5%		95%	5%
Forward		1%	96%		3%		96%	4%
Left			1%	93%	7%		93%	7%
Right		3%		<1%	97%		97%	3%
upward		<1%		<1%	3%	96%	96%	4%

The confusion matrix for the proposed system with kNN model shows (see Table II) mean accuracy of 95.7% with 99.0% (maximum) accuracy for upward gestures and 94% (minimum) accuracy for left and right gesture.

TABLE II. Confusion Matrix for KNN



True class \ Predicted class	Backward	Downward	Forward	Left	Right	upward	True Positive Rate	False Negative Rate
Backward	96%	2%	<1%	1%		<1%	96%	4%
Downward	1%	96%		<1%	3%		96%	5%
Forward	1%		97%		2%		97%	3%
Left	<1%		3%	94%	4%		94%	6%
Right	2%	3%	2%	<1%	94%		94%	6%
upward		<1%		<1%	<1%	99%	99%	2%

1) Comparison of SVM and kNN

Comparison of both SVM and kNN in terms of accuracy and performance is shown in Table III. The results show that kNN has high recognition accuracy (95.7 %) as compared to SVM (95.3%). The kNN also has high performance rates in terms of training time (3.16 secs) and prediction speed (6600 obs/sec) as compared to SVM with 6.40 secs and (2900 obs/sec).

TABLE III. Comparison of SVM and KNN Based on Recognition Accuracy, Training Time, and Prediction Speed

Method	Recognition accuracy	Training Time	Prediction Speed
SVM	95.3%	6.40 secs	2900 obs/sec
kNN	95.7%	3.16 secs	6600 obs/sec

V. Conclusion and Future Work

In this paper, we proposed a novel two hand gesture based interaction technique for 3 dimensional (3D) navigation in virtual environments (VEs). The system uses computer vision techniques for the detection of hand gestures (colored thumbs) from real scene and performs different navigation (forward, backward, up, down, left, and right) tasks in the VE. The proposed technique also allows users to efficiently control speed during navigation. This is implemented via a VE for experimental purposes. Forty participants tested the proposed technique in different lighting scenarios. Experiments revealed that the technique is feasible in normal lighting conditions, easy to learn and use, having less cognitive load on users. Its performance is evaluated using gesture recognition engines i.e. SVM and kNN. kNN achieves high accuracy rates 95.7% as compared to SVM (95.3). kNN also has high performance rates in terms of training time (3.16 secs) and prediction speed (6600 obs/sec) as compared to SVM with 6.40 secs and (2900 obs/sec).

The proposed system is sensitive to lighting conditions.

In future we will compare our proposed camera based system with Motion leap in terms of accuracy, interaction space, update rate, and positional distortion.

References

[1] Saffer, D., *Designing gestural interfaces: Touchscreens and interactive devices*. 2008: " O'Reilly Media, Inc.".

[2] Yang, R., S. Sarkar, and B. Loeding, *Handling movement epenthesis and hand segmentation ambiguities in continuous sign language recognition using nested dynamic programming.* IEEE transactions on pattern analysis and machine intelligence, 2010. **32**(3): p. 462-477.

[3] Zafrulla, Z., et al. *American sign language recognition with the kinect*. in *Proceedings of the 13th international conference on multimodal interfaces*. 2011. ACM.

[4] Uebersax, D., et al. *Real-time sign language letter and word recognition from depth data*. in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. 2011. IEEE.

[5] Pugeault, N. and R. Bowden. *Spelling it out: Real-time ASL fingerspelling recognition*. in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. 2011. IEEE.

[6] Wickeroth, D., P. Benolken, and U. Lang. *Markerless gesture based interaction for design review scenarios*. in *Applications of Digital Information and Web Technologies, 2009. ICADIWT'09. Second International Conference on the*. 2009. IEEE.

[7] Frati, V. and D. Prattichizzo. *Using Kinect for hand tracking and rendering in wearable haptics*. in *World Haptics Conference (WHC), 2011 IEEE*. 2011. IEEE.

[8] Choi, J., H. Park, and J.-I. Park. *Hand shape recognition using distance transform and shape decomposition*. in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. 2011. IEEE.

[9] Guo, Z.-m. *Research of hand positioning and gesture recognition based on binocular vision*. in *VR Innovation (ISVRI), 2011 IEEE International Symposium on*. 2011. IEEE.

[10] Raees, M., S. Ullah, and S.U. Rahman, *VEN-3DVE: vision based egocentric navigation for 3D virtual environments.* International Journal on Interactive Design and Manufacturing (IJIDeM), 2018: p. 1-11.

[11] Wu, X.-H., M.-C. Su, and P.-C. Wang. *A hand-gesture-based control interface for a car-robot*. in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. 2010. IEEE.

[12] Jalal, A., et al., *Human activity recognition via recognized body parts of human depth silhouettes for residents monitoring services at smart home.* Indoor and Built Environment, 2013. **22**(1): p. 271-279.

[13] Zhan, Y. and T. Kuroda, *Wearable sensor-based human activity recognition from environmental background sounds.* Journal of Ambient Intelligence and Humanized Computing, 2014. **5**(1): p. 77-89.

[14] Jalal, A. and Y.A. Rasheed. *Collaboration achievement along with performance maintenance in video streaming*. in *Interactive Computer Aided Learning, IEEE Conference on*. 2007.

[15] Jalal, A. and A. Shahzad. *Multiple facial feature detection using vertex-modeling structure*. in *Proceedings of the IEEE Conference on Interactive Computer Aided Learning, Villach, Austria*. 2007.

[16] Jalal, A., S. Kim, and B. Yun. *Assembled algorithm in the real-time H. 263 codec for advanced performance*. in *Enterprise networking and Computing in Healthcare Industry, 2005. HEALTHCOM 2005. Proceedings of 7th International Workshop on*. 2005. IEEE.

[17] Jalal, A. and S. Kim, *Advanced performance achievement using multi-algorithmic approach of video transcoder for low bit rate wireless communication.* ICGST International Journal on Graphics, Vision and Image Processing, 2005. **5**(9): p. 27-32.

[18] Jalal, A. *Security architecture for third generation (3g) using gmhs cellular network*. in *Emerging Technologies, 2007. ICET 2007. International Conference on*. 2007. IEEE.

[19] Jalal, A. and M.A. Zeb, *Security enhancement for e-learning portal.* International Journal of Computer Science and Network Security, 2008. **8**(3): p. 41-45.

[20] Jalal, A. and S. Kim. *The mechanism of edge detection using the block matching criteria for the motion estimation*. in *Proceedings of the Conference on Human Computer Interaction, Daegu, Korea*. 2005.

[21] Jalal, A. and S. Kim. *Algorithmic implementation and efficiency maintenance of real-time environment using low-bitrate wireless communication*. in *Software Technologies for Future Embedded and Ubiquitous Systems, 2006 and the 2006 Second International Workshop on Collaborative Computing, Integration, and Assurance. SEUS 2006/ WCCIA 2006. The Fourth IEEE Workshop on*. 2006. IEEE.

[22] Cassell, J., *A framework for gesture generation and interpretation.* Computer vision in human-machine interaction, 1998: p. 191-215.

[23] Freeman, W.T. and C.D. Weissman. *Television control by hand gestures*. in *Proc. of Intl. Workshop on Automatic Face and Gesture Recognition*. 1995.

[24] Hong, P., M. Turk, and T.S. Huang. *Gesture modeling and recognition using finite state machines*. in *Automatic face and gesture recognition, 2000. proceedings. fourth ieee international conference on*. 2000. IEEE.

[25] Moeslund, T.B., M. Störring, and E. Granum. *A natural interface to a virtual environment through computer vision-estimated pointing gestures*. in *International Gesture Workshop*. 2001. Springer.

[26] Streitz, N.A., et al., *Roomware: Towards the next generation of human-computer interaction based on an integrated design of real and virtual worlds.* Human-Computer Interaction in the New Millenium, Addison Wesley, 2001: p. 551-576.

[27] Koller, D., et al. *Real-time vision-based camera tracking for augmented reality applications*. in *Proceedings of the ACM symposium on Virtual reality software and technology*. 1997. ACM.

[28] Jalal, A., M.Z. Uddin, and T.-S. Kim, *Depth video-based human activity recognition system using translation and scaling invariant features for life logging at smart home.* IEEE Transactions on Consumer Electronics, 2012. **58**(3).

[29] Jalal, A., et al. *Human activity recognition via the features of labeled depth body parts*. in *International Conference on Smart Homes and Health Telematics*. 2012. Springer.

[30] Jalal, A., J.T. Kim, and T.-S. Kim. *Development of a life logging system via depth imaging-based human activity recognition for smart homes*. in *Proceedings of the International Symposium on Sustainable Healthy Buildings, Seoul, Korea*. 2012. Citeseer.

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

[32] Kamal, S., A. Jalal, and D. Kim, *Depth images-based human detection, tracking and activity recognition using spatiotemporal features and modified HMM.* J. Electr. Eng. Technol, 2016. **11**(3): p. 1921-1926.

[33] Jalal, A., Y. Kim, and D. Kim. *Ridge body parts features for human pose estimation and recognition from RGB-D video data*. in *Computing, Communication and Networking Technologies (ICCCNT), 2014 International Conference on*. 2014. IEEE.

[34] Turk, M.A. and A.P. Pentland. *Face recognition using eigenfaces*. in *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*. 1991. IEEE.

[35] Jalal, A. and S. Kim, *Global security using human face understanding under vision ubiquitous architecture system.* World academy of science, engineering, and technology, 2006. **13**: p. 7-11.

[36] Parmar, D.N. and B.B. Mehta, *Face recognition methods & applications.* arXiv preprint arXiv:1403.0485, 2014.

[37] Lowe, D.G. *Object recognition from local scale-invariant features*. in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*. 1999. Ieee.

[38] Li, L.-J. and L. Fei-Fei. *What, where and who? classifying events by scene and object recognition*. in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. 2007. IEEE.

[39] Yao, B. and L. Fei-Fei. *Grouplet: A structured image representation for recognizing human and object interactions*. in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. 2010. IEEE.

[40] Delaitre, V., I. Laptev, and J. Sivic. *Recognizing human actions in still images: a study of bag-of-features and part-based representations*. in *BMVC 2010-21st British Machine Vision Conference*. 2010.

[41] Shapovalova, N., et al. *On importance of interactions and context in human action recognition*. in *Iberian conference on pattern recognition and image analysis*. 2011. Springer.

[42] Yao, B., A. Khosla, and L. Fei-Fei. *Combining randomization and discrimination for fine-grained image categorization*. in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. 2011. IEEE.

[43] Koniusz, P. and K. Mikolajczyk. *Soft assignment of visual words as linear coordinate coding and optimisation of its reconstruction error*. in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. 2011. IEEE.

[44] Koniusz, P. and K. Mikolajczyk. *Spatial coordinate coding to reduce histogram representations, dominant angle and colour pyramid match*. in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. 2011. IEEE.

[45] Li, P., J. Ma, and S. Gao. *Actions in still web images: Visualization, detection and retrieval*. in *International Conference on Web-Age Information Management*. 2011. Springer.

[46] Yao, B., et al. *Human action recognition by learning bases of action attributes and parts*. in *Computer Vision (ICCV), 2011 IEEE International Conference on*. 2011. IEEE.

[47] Delaitre, V., J. Sivic, and I. Laptev. *Learning person-object interactions for action recognition in still images*. in *Advances in neural information processing systems*. 2011.

[48] Li, P. and J. Ma. *What is happening in a still picture?* in *Pattern Recognition (ACPR), 2011 First Asian Conference on*. 2011. IEEE.

[49] Thurau, C. and V. Hlavác. *Pose primitive based human action recognition in videos or still images*. in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. 2008. IEEE.

[50] Dalal, N. and B. Triggs. *Histograms of oriented gradients for human detection*. in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. 2005. IEEE.

[51] Gupta, A., A. Kembhavi, and L.S. Davis, *Observing human-object interactions: Using spatial and functional compatibility for recognition.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009. **31**(10): p. 1775-1789.

[52] Desai, C. and D. Ramanan. *Detecting actions, poses, and objects with relational phraselets*. in *European Conference on Computer Vision*. 2012. Springer.

[53] Raja, K., et al. *Joint pose estimation and action recognition in image graphs*. in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. 2011. IEEE.

[54] Yao, B. and L. Fei-Fei, *Recognizing human-object interactions in still images by modeling the mutual context of objects and human poses.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012. **34**(9): p. 1691-1703.

[55] Sener, F., C. Bas, and N. Ikizler-Cinbis. *On recognizing actions in still images via multiple features*. in *European Conference on Computer Vision*. 2012. Springer.

[56] Belongie, S., G. Mori, and J. Malik, *Matching with shape contexts*, in *Statistics and Analysis of Shapes*. 2006, Springer. p. 81-105.

[57] Wang, Y., et al. *Unsupervised discovery of action classes*. in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*. 2006. IEEE.

[58] Yao, B. and L. Fei-Fei. *Modeling mutual context of object and human pose in human-object interaction activities*. in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. 2010. IEEE.

[59] Oliva, A. and A. Torralba, *Modeling the shape of the scene: A holistic representation of the spatial envelope.* International journal of computer vision, 2001. **42**(3): p. 145-175.

[60] Prest, A., C. Schmid, and V. Ferrari, *Weakly supervised learning of interactions between humans and objects.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012. **34**(3): p. 601-614.

[61] Bay, H., et al., *Speeded-up robust features (SURF).* Computer vision and image understanding, 2008. **110**(3): p. 346-359.

[62] Ikizler, N., et al. *Recognizing actions from still images*. in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. 2008. IEEE.

[63] Ahmed, M., N. Al-Jawad, and A.T. Sabir. *Gait recognition based on Kinect sensor*. in *Real-Time Image and Video Processing 2014*. 2014. International Society for Optics and Photonics.

[64] Andersson, V.O. and R.M. de Araújo. *Person Identification Using Anthropometric and Gait Data from Kinect Sensor*. in *AAAI*. 2015.

[65] Ball, A., et al. *Unsupervised clustering of people from'skeleton'data*. in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. 2012. ACM.

[66] Dikovski, B., G. Madjarov, and D. Gjorgjevikj. *Evaluation of different feature sets for gait recognition using skeletal data from Kinect*. in *Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th International Convention on*. 2014. IEEE.

[67] Kwolek, B., et al. *3D gait recognition using spatio-temporal motion descriptors*. in *Asian Conference on Intelligent Information and Database Systems*. 2014. Springer.

[68] Preis, J., et al. *Gait recognition with kinect*. in *1st international workshop on kinect in pervasive computing*. 2012. New Castle, UK.

[69] Farooq, A., A. Jalal, and S. Kamal, *Dense RGB-D Map-Based Human Tracking and Activity Recognition using Skin Joints Features and Self-Organizing Map.* KSII Transactions on Internet & Information Systems, 2015. **9**(5).

[70] Yoshimoto, H., N. Date, and S. Yonemoto. *Vision-based real-time motion capture system using multiple cameras*. in *Multisensor Fusion and Integration for Intelligent Systems, MFI2003. Proceedings of IEEE International Conference on*. 2003. IEEE.

[71] Zhang, X., et al., *A framework for hand gesture recognition based on accelerometer and EMG sensors.* IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 2011. **41**(6): p. 1064-1076.

[72] Kammer, D., et al. *Towards a formalization of multi-touch gestures*. in *ACM International Conference on Interactive Tabletops and Surfaces*. 2010. ACM.

[73] Hoggan, E., et al. *Multi-touch rotation gestures: Performance and ergonomics*. in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2013. ACM.

[74] Zabulis, X., H. Baltzakis, and A.A. Argyros, *Vision-Based Hand Gesture Recognition for Human-Computer Interaction.* The universal access handbook, 2009. **34**: p. 30.

[75] Buehler, P., et al., *Upper body detection and tracking in extended signing sequences.* International journal of computer vision, 2011. **95**(2): p. 180.

[76] Rautaray, S.S. and A. Agrawal, *Vision based hand gesture recognition for human computer interaction: a survey.* Artificial Intelligence Review, 2015. **43**(1): p. 1-54.

[77] Ren, Z., J. Yuan, and Z. Zhang. *Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera*. in *Proceedings of the 19th ACM international conference on Multimedia*. 2011. ACM.

[78] Liou, W.-G., C.-Y. Hsieh, and W.-Y. Lin. *Trajectory-based sign language recognition using Discriminant Analysis in higher-dimensional feature space*. in *Multimedia and Expo (ICME), 2011 IEEE International Conference on*. 2011. IEEE.

[79] Tran, C. and M.M. Trivedi, *3-D posture and gesture recognition for interactivity in smart spaces.* IEEE Transactions on Industrial Informatics, 2012. **8**(1): p. 178-187.

[80] Alon, J., et al., *A unified framework for gesture recognition and spatiotemporal gesture segmentation.* IEEE transactions on pattern

analysis and machine intelligence, 2009. **31**(9): p. 1685-1699.

[81] Sarkar, S., et al. *Segmentation-robust representations, matching, and modeling for sign language*. in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*. 2011. IEEE.

[82] Lee, B. and J. Chun. *Interactive manipulation of augmented objects in marker-less ar using vision-based hand interaction*. in *Information Technology: New Generations (ITNG), 2010 Seventh International Conference on*. 2010. IEEE.

[83] Störring, M., et al. *Computer vision-based gesture recognition for an augmented reality interface*. in *4th IASTED international conference on visualization, imaging, and image processing*. 2004.

[84] Yang, M.-T., W.-C. Liao, and Y.-C. Shih. *VECAR: Virtual English classroom with markerless augmented reality and intuitive gesture interaction*. in *Advanced Learning Technologies (ICALT), 2013 IEEE 13th International Conference on*. 2013. IEEE.

[85] Lee, M., R. Green, and M. Billinghurst. *3D natural hand interaction for AR applications*. in *Image and Vision Computing New Zealand, 2008. IVCNZ 2008. 23rd International Conference*. 2008. IEEE.

[86] Hassanpour, R., S. Wong, and A. Shahbahrami. *Visionbased hand gesture recognition for human computer interaction: A review*. in *IADIS international conference interfaces and human computer interaction*. 2008. Published.

[87] Leap, C., *API--https://developer. leapmotion. com/documentation/cpp/api.* Leap_Classes. html.

[88] Guna, J., et al., *An analysis of the precision and reliability of the leap motion sensor and its suitability for static and dynamic tracking.* Sensors, 2014. **14**(2): p. 3702-3720.

[89] Kean, S., J.C. Hall, and P. Perry, *Microsoft's kinect sdk.* Meet the Kinect, 2011: p. 151-173.

[90] Buchmann, V., et al. *FingARtips: gesture based direct manipulation in Augmented Reality*. in *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*. 2004. ACM.

[91] Park, H., H.-K. Jung, and S.-J. Park, *Tangible AR interaction based on fingertip touch using small-sized nonsquare markers.* Journal of Computational Design and Engineering, 2014. **1**(4): p. 289-297.

[92] Baldauf, M., et al. *Markerless visual fingertip detection for natural mobile device interaction*. in *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*. 2011. ACM.

[93] Lee, T. and T. Hollerer. *Handy AR: Markerless inspection of augmented reality objects using fingertip tracking*. in *Wearable Computers, 2007 11th IEEE International Symposium on*. 2007. IEEE.

[94] Wachs, J.P., et al., *Vision-based hand-gesture applications.* Communications of the ACM, 2011. **54**(2): p. 60-71.

[95] Chun, W.H. and T. Höllerer. *Real-time hand interaction for augmented reality on mobile phones*. in *Proceedings of the 2013 international conference on Intelligent user interfaces*. 2013. ACM.

[96] Lee, J.Y., G.W. Rhee, and D.W. Seo, *Hand gesture-based tangible interactions for manipulating virtual objects in a mixed reality environment.* The International Journal of Advanced Manufacturing Technology, 2010. **51**(9-12): p. 1069-1082.

[97] Lu, G., et al., *Immersive manipulation of virtual objects through glove-based hand gesture interaction.* Virtual Reality, 2012. **16**(3): p. 243-252.

[98] Vatavu, R.-D., Ș.-G. Pentiuc, and C. Chaillou, *On natural gestures for interacting in virtual environments.* Advances in electrical and computer engineering, 2005. **24**(5): p. 72-79.

[99] Bowman, D.A., D. Koller, and L.F. Hodges, *A methodology for the evaluation of travel techniques for immersive virtual environments.* Virtual reality, 1998. **3**(2): p. 120-131.

[100] De Boeck, J., C. Raymaekers, and K. Coninx. *Are existing metaphors in virtual environments suitable for haptic interaction*. 2005. sl 2005.

[101] Mitra, S. and T. Acharya, *Gesture recognition: A survey.* IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2007. **37**(3): p. 311-324.

[102] Motion, L., *Leap motion developer portal*. 2014.

[103] Dardas, N.H. and M. Alhaj, *Hand gesture interaction with a 3D virtual environment.* The International Journal of ACM JORDAN, 2011. **2**(3): p. 186-194.

[104] Yang, J., Y. Xu, and C. Chen. *Gesture interface: Modeling and learning*. in *Robotics and Automation, 1994. Proceedings., 1994 IEEE International Conference on*. 1994. IEEE.

[105] Cooper, H., B. Holt, and R. Bowden, *Sign language recognition*, in *Visual Analysis of Humans*. 2011, Springer. p. 539-562.

[106] Kim, J.-S., et al. *Study on intelligent autonomous navigation of Avatar using hand gesture recognition*. in *Systems, Man, and Cybernetics, 2000 IEEE International Conference on*. 2000. IEEE.

[107] Chen, Q., et al. *Accessing Learning Objects in Virtual Environment by Hand Gestures and Voice*. in *Proc. 3rd Annual Scientific Conference of LORNET Research Network (I2LOR-06), TELUQ-UQAM University in Montreal, Canada*. 2006.

[108] Krum, D.M., et al. *Speech and gesture multimodal control of a whole earth 3d virtual environment*. in *Eurographics-IEEE Visualization Symposium 2002*. 2002. Citeseer.

[109] Shao, L., *Hand movement and gesture recognition using Leap Motion Controller.*

[110] Kerefeyn, S. and S. Maleshkov, *Manipulation of virtual objects through a LeapMotion optical sensor.* International Journal of Computer Science Issues (IJCSI), 2015. **12**(5): p. 52.

[111] Khundam, C. *First person movement control with palm normal and hand gesture interaction in virtual reality*. in *Computer Science and Software Engineering (JCSSE), 2015 12th International Joint Conference on*. 2015. IEEE.

[112] de Paiva Batista, R.M., *Navigating Virtual Reality Worlds with the Leap Motion Controller.* 2016.

[113] Liang, H., et al., *Hand gesture-based interactive puppetry system to assist storytelling for children.* The Visual Computer, 2016: p. 1-15.

[114] Nabiyouni, M., B. Laha, and D.A. Bowman. *Poster: Designing effective travel techniques with bare-hand interaction*. in *3D User Interfaces (3DUI), 2014 IEEE Symposium on*. 2014. IEEE.

[115] Prazina, I. and S. Rizvić. *Natural interaction with small 3D objects in virtual environments*. in *Submitted to 20th Central European Seminar on Computer Graphics*. 2016.

[116] Kammergruber, F., A. Ebner, and W.A. Günthner. *Navigation in virtual reality using Microsoft Kinect*. in *12th international conference on construction application of virtual reality, Taipei, Taiwan*. 2012.

[117] Kim, J.-O., M. Kim, and K.-H. Yoo, *Real-time hand gesture-based interaction with objects in 3D virtual environments.* Int. J. Multimed. Ubiquitous Eng, 2013. **8**(6): p. 339-348.

[118] Vultur, O.M., Ș.G. Pentiuc, and A. Ciupu. *Navigation system in a virtual environment by gestures*. in *Communications (COMM), 2012 9th International Conference on*. 2012. IEEE.

[119] Phung, S.L., A. Bouzerdoum, and D. Chai. *A novel skin color model in ycbcr color space and its application to human face detection*. in *Image Processing. 2002. Proceedings. 2002 International Conference on*. 2002. IEEE.

[120] Raees, M., et al., *Image based recognition of Pakistan sign language.* Journal of Engineering Research, 2016. **4**(1).

[121] Piyathilaka, L. and S. Kodagoda. *Gaussian mixture based HMM for human daily activity recognition using 3D skeleton features*. in *Industrial Electronics and Applications (ICIEA), 2013 8th IEEE Conference on*. 2013. IEEE.

[122] Jalal, A., S. Kamal, and D. Kim, *A depth video sensor-based life-logging human activity recognition system for elderly care in smart indoor environments.* Sensors, 2014. **14**(7): p. 11735-11759.

[123] Jalal, A., et al., *Robust human activity recognition from depth video using spatiotemporal multi-fused features.* Pattern recognition, 2017. **61**: p. 295-308.

[124] Jalal, A., S. Kamal, and D. Kim. *Shape and motion features approach for activity tracking and recognition from kinect video camera*. in *Advanced Information Networking and Applications Workshops (WAINA), 2015 IEEE 29th International Conference on*. 2015. IEEE.

[125] Jalal, A., S. Kamal, and D. Kim. *Individual detection-tracking-recognition using depth activity images*. in *Ubiquitous Robots and Ambient Intelligence (URAI), 2015 12th International Conference on*. 2015. IEEE.

[126] Wu, H., et al. *Human activity recognition based on the combined SVM&HMM*. in *Information and Automation (ICIA), 2014 IEEE International Conference on*. 2014. IEEE.

[127] Jalal, A. and Y. Kim. *Dense depth maps-based human pose tracking*

and recognition in dynamic scenes using ridge data. in *Advanced Video and Signal Based Surveillance (AVSS), 2014 11th IEEE International Conference on*. 2014. IEEE.

[128] Pears, Alan (June 1998). "Chapter 7: Appliance technologies and scope for emission reduction". Strategic Study of Household Energy and Greenhouse Issues (PDF). Sustainable Solutions Pty Ltd. Department of Industry and Science, Commonwealth of Australia. p. 61. http://www.energyrating.gov.au/document/report-strategic-study-household-energy-and-greenhouse-issues. Retrieved on 26-06-2017.

[129] *Australian Greenhouse Office (May 2005). «Chapter 5: Assessing lighting savings».* Working Energy Resource and Training Kit: Lighting. https://web.archive.org/web/20070415151053/http://www.greenhouse.gov.au/lgmodules/wep/lights/training/training9.html. Retrieved on 26-06-2017.

[130] Australian Greenhouse Office. (May 2005). 'Chapter 5: Assessing lighting savings'. Working Energy Resource and Training Kit: Lighting. Retrieved 17-03-2017.

[131] "Low-Light Performance Calculator". http://www.scopecalc.com/ Retrieved on 03-06-2017.

[132] How to use a lux meter (Australian recommendation)' (PDF), April 2010. Sustainability Victoria. http://www.sustainability.vic.gov.au/ Retrieved on 09-05-2017.

[133] "Illumination. - 1926.56". Regulations (Standards - 29 CFR). Occupational Safety and Health Administration, US Dept. of Labor. www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=STANDARDS&p_id=10630, Retrieved on 09-05-2017.

Mr. Rehman

He is currently pursuing Ph. D. degree in 3D Interaction from the University of Malakand, Pakistan. He received his M. Phil in Virtual and Augmented Reality and BS degree in Computer Science from University of Malakand in 2006 and 2016 respectively. His research interest includes 3D interaction, virtual and augmented reality, and machine learning.

Dr. Ullah

He is currently working as Associate Professor in the department of Computer Science & IT, University of Malakand, Pakistan. He received his PhD degree in Computer Science in specialty in Robotics from University of Evry Val d'Essonne, France in 2011. He also earned MS degree of Computer Science with specialty in Virtual Reality and Intelligent Systems from the same University in 2007. He received MSc degree of Computer Science and BSc with major in Mathematics and Computer Science from University of Peshawar in 2001 and 1998 respectively. Dr. Sehat Ullah has research interest in virtual and augmented reality systems, machine learning, image processing and haptic systems.

Mr. Raees

He is serving as assistant professor of computer science in the department of Higher Education, Khyber pakhtunkhwa, Pakistan. He received his M.Phil degree from University Of Malakand in 2015 and is currently pursuing Ph.D. in flexible 3D interaction from the same university. His research interest includes 3D interactions, computer vision and machine learning.

Are Instructed Emotional States Suitable for Classification? Demonstration of How They Can Significantly Influence the Classification Result in An Automated Recognition System

M. Magdin*, F. Prikler

Constantine the Philosopher University in Nitra, Faculty of Natural Sciences , Department of Computer Science, Tr. A. Hlinku 1, 949 74 Nitra, Slovakia

Received 7 February 2018 | Accepted 7 March 2018 | Published 16 March 2018

Abstract

At the present time, various freely available or commercial solutions are used to classify the subject's emotional state. Classification of the emotional state helps us to understand how the subject feels and what he is experiencing in a particular situation. Classification of the emotional state can thus be used in various areas of our life from neuromarketing, through the automotive industry (determining how emotions affect driving), to implementing such a system into the learning process. The learning process, which is the (mutual) interaction between the teacher and the learner, is an interesting area in which individual emotional states can be explored. In this pedagogical-psychological area several research studies were realized. These studies in some cases demonstrated the important impact of the emotional state on the results of the students. However, for comparison and unambiguous classification of the emotional state most of these studies used the instructed (even constructed) stereotypical facial expressions of the most well-known test databases (Jaffe is a typical example). Such facial expressions are highly standardized, and the software can recognize them with a fairly big percentage, but this does not necessarily point to the actual success rate of the subject's emotional classification in such a test because the similarity to real emotional expression remains unknown. Therefore, we examined facial expressions in real situations. We have subsequently compared these examined facial expressions with the instructed expressions of the same emotions (the Jaffe database). The overall average classification score in real facial expressions was 94.58%.

Keywords

Instructed Stimuli, Real-life Stimuli, Facial Expression, Emotion, Face Detection.

DOI: 10.9781/ijimai.2018.03.002

I. Introduction

Face recognition and subsequently classification of emotional state is a very important research area of learning process in recent days. This fact was confirmed in study [2], where result of this study shows that emotions may have a significant effect on human trust because the necessary part of social communication between people is the skill to correctly express their emotions. Expressions in face are for all people not only easy movements of face parts but also are the most efficient means to express emotion [10], [16]. In the mutual communication we need correct understanding of emotional state. Therefore we need not only perceive correctly change of muscles in face. We must also know how decode this ostensible movement of muscle in the right way for the following classification of the emotion that is represented by the facial expression. [17], [20], [42]. In present face recognition is an important area not only in research but also in various applications using human-computers interface. For normal human is not a problem to recognize various expressions in face (without delay), however recognition expression of face by computer in real time is still a big problem [32], [28]. The past decade has witnessed many new developments in facial expression analysis due to its wide application in robotic vision [37], [24], forensics [5], affective computing, man-machine interaction [29], [18] and even medical diagnosis [24]. Therefore we can say that the correct classification of emotional state is an important role in various applications and industries of our life, for example: robust tools for behavioral people research, speech processing and speech recognition or access control of people by building monitoring, etc. [25]. Expression of emotions in face is the most natural way for people to express not only their emotions but also their intentions. This is a relatively simple task for a human from view point of detection in real time. However, for a computer, as automatic recognition system, this task is not simple [36]. We can identify the emotional state in several ways, also on the bases of voice analysis, gesture analysis, analysis of handwritten texts, etc. The different studies in last years have shown that many characteristic features needed to classify the emotional state are expressed especially by the expressions of the face [28] in real time or emotion recognition from facial images, e.g. [6]. [34] and [41]. In these research studies researchers reported on significant differences of examined subjects. Recently "EMOTHAW" (EMOTion recognition from HAndWriting and drAWing) was designed and created. EMOTHAW is a first publicly available database (https://sites.google.com/site/becogsys/emothaw) for recognition of emotional states on the basis handwriting [27]. Recognition and classification of emotional state by using multiple ways (so-called as multimodal recognition of

* Corresponding author.

E-mail address: mmagdin@ukf.sk

emotions) is currently considered in the research area as a perspective area, especially because of the big amount of data that can be acquired concurrently on the basis of different physiological signals. These physiological signals are a natural expression of the human body and therefore they can be successfully used to classify the emotional state. Such physiological signals may be electroencephalogram (EEG), temperature or electrocardiogram (ECG). Through them, we can classify an emotional state such as happiness, sadness or anger [39]. On the basis of different sensors we can capture these signals and we can identify individual difference factors. These factors significantly affect the ability to classify emotions (see e.g. [21] for an overview).

Multimodal face recognition is based on capture of individual facial features and observing of each difference. The parts of the face that can be captured are: facial features, eyelids, eyebrows, lips, nose, chin, etc. For this purpose, we can use the so-called "multi-point masks". Individual points on this mask represent the extracted areas from which the classification of the emotional state is performed [7]. Because of the amount of individual parts on this mask that can be scanned simultaneously and the amount of output data thus produced, there have been developed various methods for detecting, extracting and subsequently classifying the emotional state either in real time or using stored face images. The effectiveness of these methods is evaluated using facial image databases with different expressions [25]. The first step to be able to respond to emotions is affect recognition that focuses on identifying emotions and other affective phenomena on the subject. The evaluation of the affective state is usually done according to an emotional model that suits the particular application. One of the simplest models is the one described by Ekman, which is composed of six discrete primitive emotions, namely anger, fear, sadness, surprise, disgust and happiness [12]. Other alternative models include Plutchik's Wheel of Emotion [35] and Russell's Circumplex Model [38], which locates emotions in a 2D space defined by the arousal (or activation), and valence (or positiveness). The latter was extended in [31] by adding a third dimension (dominance) to avoid overlapping of certain emotions [3].

From academic year 2016/2017 in our department of computer science a project has been realized: Modeling the behavior of users based on data mining with support from IBM Bluemix. Our aim in this project is the design and creation of a complex automatic system for detection, face features extraction and classification of the emotional state. In this moment our final application is capable of detecting face in real time, to extract various face features on the basis of the use of multi-point mask and the classification of the emotional state by Ekman scale. The data obtained from the web camera (using data mining techniques) is used as a basis for user behavior modeling in Moodle environment. According to the literature analysis of the evaluation of the emotional state, there are various solutions to this problem; however, all solutions allude to the fact - evaluating the amount of mined data of various participants in real time. In this publication, we point out the results of the solution that has significantly influenced our further research direction.

In the section Material and methods, we are focusing on researches dealing with face recognition using real-time webcam, followed by extraction of individual facial features and classification of the emotional states of students. In the Calculation chapter, we describe the methodology of the two experiments we used. In Experiment 1, we focused primarily on determining the degree of classification success using constructed expressions found in the Japanese models database called Jaffe. This is the most used database for determining emotional states due to precisely tagged images with facial expressions of individual subjects. The results of Experiment 1 are listed in the Results and Discussion section of Experiment 1. Based on the results obtained and their evaluation, we have come to the realization of Experiment 2, which evaluates real-life expressions of emotion. Results and recommendations are listed in the Results and Discussion section.

II. Methodology

In our daily life, during performing various activities, we create and generate different types of data in a large quantity. This data contains an important information regarding our activities, style of living, our behavior in different situations and our emotions. Psychologists have gradually classified different types of emotions, such as love, joy, surprise, anger, sadness and fear. The classification model of Paul Ekman is considered a reference model. It contains emotional states as happiness, anger, disgust, fear, sadness, surprise. Later on, the neutral state was included in these six emotional states [23]. Certain facial expressions are recognizable signals of emotion: a smile signals happiness, crying signals sadness, a nose scrunch signals disgust [14], [15], [9], [26] is the cornerstone of a dominant theory of emotion. These facts are an important prerequisite for various emotion research, or in affective neuroscience, and in a range of applications. A survey of emotion researchers found that a large majority accepted the thesis that certain faces signal specific emotions [13], [11].

In 1975 a method of coding for the face using the action units (FACS) was proposed [16]. FACS was containing 46 points within the geometric face model. These facial parts were the basis for the design and implementation of multi-point masks. These facial parts and facial points have been designed not only for the detection and extraction of the necessary areas of interest, but with their changes, it was possible to classify the emotional state. Each emotion was defined on the basis of a facial point change. Over time, this extraction-classification method was enhanced by the Gaussian classifier, by the Hiden Markov Model (HMM) or the Bayes classifier [8]. The detection phase has been progressively improved. Instead of the robust Viola-Jones detector, the use of the Support Vector Machine (SVM) method has begun. This method completely replaced all three phases of the recognition process: detection, extraction, and classification [40], [43].

These three phases of the recognition process are important for various area researches. The results from recognition process can be used in various fields of our life as investigation of impact of some products on human behavior (neuromarketing), impact of advertisements on human choosing, etc. Emotion recognition accuracy (ERA) from faces has been conceptualized as a performance measure of emotional intelligence [29]. Currently for classification of emotional state, several research groups use in most cases various databases of human faces. However these databases contain only instructed emotional state [22]. These images are organized to individual databases and contain people's faces, which differ depending on the lighting, the angle of the face rotation, the noise in the images and others. These differences are the basis for determining the percentage of success rate of face recognition and subsequent classification of the emotional state [33]. The problem is often the low count of images in the database. Therefore, it is questionable whether the standard algorithms used would work in a realistic situation and then emotional state would be classified with the same percentage of success rate. We can claim that, basing on the fact that a lot of research uses a method in which users mimic the emotional state of a person in the picture from the database, certain face expressions are highly consistent with images that contain databases and achieve a very high percentage of success classification [11]. The software then does not recognize the real states, but the states that are artificially induced. Several experts have pointed out this problem in their work [1], [11].

Based on their opinion, we are inclined to solve the proposed pilot project by submitting the realized evaluation solution, both from the point of proper facial detection under different lighting conditions, the distance of the subject or the rotation of the subject's face, as well as determining the rate of success of the emotional state classification of the subject.

The success rate of the face detection in the image was discussed in the paper titled "Real Time Facial Expression Recognition Using Webcam and SDK Affectiva" [44], where the proposed solution was able to detect the face in real time with 100% success rate in frontal face situations. We are having problems only in the case when face is rotated in left or right, having a total average success rate of 84,27%. Detection was tested using 6 robust databases on a total of 9252 frames:

- Bao Face Database,
- CMU/VASC Image Database,
- Caltech Faces 1999 Database,
- NIST Mug-shot Images Database,
- Yale Face Database,
- Vision Group of Essex University Face Database.

In the following section, we present an experiment that determined the success rate of the emotional state classification of the subject under study (Jaffe database).

III. Measures and Procedure of Experiment 1 – Determining the Degree of Success of the Emotional Status Classification Using the Jaffe Database

The methods that we can use to analyze emotional states have a same basis as methods used for face recognition. Analysis of emotional state is standard divided into three basic phases [18]: detecting the face or face part on the image, extracting the area of interest, and classifying the emotional state. The use of a suitable classifier always depends on what the desired output will be. Classifiers may not be priority designed for phase of classification of the emotional state. We can use these classifiers appropriately also to filter the properties we are interested. At present there are a lot of different methods for detection, extraction and classification (about 200) and for the experienced programmer their use is not an easy process. On the other hand, however, we can use available libraries, apps and SDKs that are already verified, when we are designing and creating an automatic recognition system. Such a SDK solution is also a library from Affectiva. Face detection is performed using the Viola-Jones detector. Histogram of Oriented Gradient (HOG) is extracted from the image area of interest defined by the face orientation points. Support Vector Machine Algorithm (SVM), trained on 10,000 manually encoded face images collected from around the world is used to generate a score from 0 to 100 for each face action. For details on the training and testing system, see [44]. The use of such a solution brings the advantage that in the final phase of the recognition process (classification) we do not have to deal with the so-called "training set". A variety of methods are used to verify the overall functionality of the systems and also to determine the degree of classification success. The most commonly used method, however, is when the subject has to emulate the emotional expression and then the system must correctly classify it. This is done using images from available facial databases with corresponding emotional expressions (according to Ekman's classification). The entire testing process is performed under the supervision of a psychologist to rule out the possibility of incorrect classification or incorrectly instructed expression. More about this classification can be found in [4].

The biggest problem of different solutions is in the listed databases. Researchers often use only one database to determine the percentage of success rate of the recognition system, which ultimately results comparable to other top-level systems. Such testing is quite trivial and inconsistent [6].

We realized the solution may be evaluating using individual databases. These databases contain various images with different light conditions, face rotation, or with several captured faces in the image. This result is for us a comprehensive assessment of the rate of face detection success [44]. The percentage of success rate of the proposed solution was verified as follows (Fig. 1): on one side of a table, we placed a monitor on which the individual images from the respective databases were gradually displayed; on the other side of the table, we placed a laptop with an application for face recognition and classification of emotional state at a distance of 1m; the laptop contained a webcam that captured the image of the monitor; the distance of 1m is the standard distance to capture the object by a webcam.

Fig. 1. Verification of the percentage of success rate of the proposed solution.

To simplify the entire detection process and subsequent classification, we removed the step in which the subject would have to emulate emotions. Because these databases contained only the different faces of the subjects without the precisely defined type of emotional state, we were looking for a database that would contain the rated images. Such database is Jaffe, which contains 214 Japanese model photos (marked KA, KL, KM, KR, MK, NA, NM, TM, UY, YM). It is important to note that these models were asked to reproduce the desired emotion: anger, disgust, fear, happy (joy), neutral, sadness and surprise (Fig. 2). For this reason, we can say that those were forced expressions of the face where the monitored subject (in this case the Japanese model) was fully aware of the emotion she was trying to imitate according to instructions. The test set of images includes 10 people who were photographed during a particular emotion expression 3 to 4 times.

Fig. 2. Faces selected from the tested images with an instructed emotional state (Left to Right - Anger, Disgust, Fear, Happiness, Neutral, Sadness and Surprise)

We've continuously recorded the output from our classification software, which expressed the intensity of the classified emotion in percentage (the classifier is considered successful if it reaches a recognition rate of more than 50%). Testing for each of the Japanese models (Fig. 3) was repeated 5 times (but we have always achieved the same result, confirming that the software correctly detected the face in the same way and also classified the emotional state with the same result).

Fig. 3. Demonstration of facial expression classification (emotional state - fear).

IV. Results and Discussion – Experiment 1

As we can see from Tables I and II, the software often classified a completely different emotional state with high percentage, as was expected.

TABLE I. The Rate of Success of Classification (in Percentage) of the Emotional State of the KM Model

Requested Emotions (Model name: KM)		Recognized Emotion by the Software						
		An	Di	Fe	Ha	Ne	Sa	Su
	An	0,00	1,75	1,34	0,00	94,32	0,00	2,59
	Di	0,00	61,78	0,00	0,00	34,46	0,00	3,76
	Fe	0,00	16,00	33,61	0,00	32,97	5,43	11,99
	Ha	0,00	0,14	0,38	92,16	0,07	0,23	7,02
	Ne	0,00	1,09	31,12	0,00	50,84	0,00	16,95
	Sa	0,00	1,47	11,72	0,00	72,63	1,40	12,78
	Su	0,00	0,00	0,00	0,00	100,00	0,00	0,00

Explanation of Table I:

An – Anger
Di – Disgust,
Fe – Fear,
Ha – Happy (Joy),
Ne – Neutral,
Sa – Sadness,
Su – Surprise.

When classifying the model labelled as KM, the software rated an emotional state anger as a neutral expression (94.32%), also incorrectly determined the emotional state of sadness (72.63%) and surprise (100%) as neutral. However, after consulting with a psychology specialist, we have come to the conclusion that these expressions are problematic to interpret by the model KM, and that's why the software classified them as another type of emotional state (Fig. 4).

Fig. 4. Example of a problematically instructed emotional state (left Anger, right Sadness).

The situation was similar with the model labelled MK. The software classified anger as very neutral (97.48%), also misinterpreted disgust (91.14%), sadness (97.93%) and surprise (100%) as neutral expressions (Table II).

TABLE II. The Rate of Success of the Classification (in Percentage) of the Emotional State of the MK Model

Requested Emotions (Model name: MK)		Recognized Emotion by the Software						
		An	Di	Fe	Ha	Ne	Sa	Su
	An	0,00	1,45	0,00	0,00	97,48	0,00	1,07
	Di	0,00	2,66	0,00	0,00	91,14	4,03	2,17
	Fe	0,00	0,53	50,29	0,00	42,05	0,36	6,77
	Ha	0,00	0,00	0,00	99,84	0,05	0,00	0,11
	Ne	0,00	1,56	0,00	0,00	97,03	0,00	1,41
	Sa	0,00	1,19	0,00	0,00	97,93	0,00	0,88
	Su	0,00	0,00	0,00	0,00	100,00	0,00	0,00

After consulting a specialist in the psychology department, we also came to the conclusion that these expressions are misinterpreted by MK models and therefore classified by the software as another type of emotional state (Fig. 5).

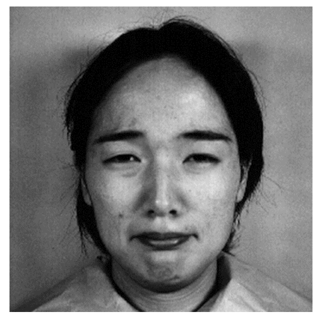


Fig. 5. Example of a problematically instructed emotional state - Sadness.

For the model labeled UY (See Table III), the software has always evaluated (albeit to a minimal extent) each of the expressions as disgust (0.43-0.45%). Anger was classified as neutral state (87.31%), disgust as neutral state (80.16%), sadness as neutral state (87.76%) and surprise as neutral state (51.82%).

TABLE III. The Rate of Success of Classification (in Percentage) of the Emotional State of the UY Model

Requested Emotions (Model name: UY)		Recognized Emotion by the Software						
		An	Di	Fe	Ha	Ne	Sa	Su
	An	0,00	0,43	11,84	0,00	87,31	0,00	0,42
	Di	0,00	0,44	18,97	0,00	80,16	0,00	0,43
	Fe	0,00	0,43	49,73	0,00	49,28	0,00	0,56
	Ha	0,00	0,43	33,60	64,33	0,11	0,98	0,55
	Ne	0,00	0,43	35,76	0,00	63,27	0,00	0,54
	Sa	0,00	0,43	11,44	0,00	87,76	0,00	0,37
	Su	0,00	0,45	47,10	0,00	51,82	0,00	0,63

After consulting a psychology specialist, we also came to the conclusion that these expressions are misinterpreted by the model labelled as UY, and therefore classified by the software as another type of emotional state. As an example, we are showing the type of emotional state –anger, which was supposed to be expressed by the model (Fig. 6). However, according to the psychologist, this is a neutral expression, which is also confirmed by the recognition results.

Fig. 6. An example of a problematically induced emotional state - Anger.

This incorrect classification of the emotional states in these 3 models caused the average rate of classification to fall to 30.01% (Table IV). In other cases, we did not detect any errors or deviations and the software correctly detected the emotional states.

TABLE IV.
Average Rate of Emotional Status Classification (in Percentage)

		Recognized Emotion by the Software						
		An	Di	Fe	Ha	Ne	Sa	Su
Requested Emotions	An	0,00	1,42	1,32	0,00	95,37	0,07	1,82
	Di	0,00	27,42	1,90	0,00	66,59	0,50	3,59
	Fe	0,00	19,43	33,23	0,00	37,68	2,52	7,14
	Ha	0,03	2,12	11,54	59,03	23,37	1,47	2,44
	Ne	0,00	1,53	17,33	0,00	76,97	0,02	4,14
	Sa	1,79	3,04	6,09	0,00	82,49	2,58	4,00
	Su	0,53	17,70	22,20	8,74	38,27	1,72	10,83

The average value was achieved by adding values of the diagonal of the table and we divided this value with the number of emotional expressions (number 7). The emotional state of Anger was not classified. It had a very low classification value of less than 0.1%. Low values were also classified for sadness (2.58%), surprise (10.83), disgust (27.42%) and fear (33.23%). The classifier was successful in only two cases: happiness (59.03%) and neutral expression (76.97%). This result is very interesting, because even after repeating 5 times the measurement (to remove any classifier error) we have always achieved the same results. There are two explanations for us:

E1: While this test method allows for real-time trouble-free face detection, it does not provide an adequate evaluation option for determining the degree of success of the emotional classification. Such a method of evaluation results in the loss of information essential for determining the degree of success of a classification, in particular due to the ambient light conditions.

E2: Every person has the emotions written in the face (typical example is UY model testing, where disgust reaches 0.43-0.45%) and those, whether to a greater or lesser extent, cause a possible classification error and also act as a measurement deviation.

As with Japanese models we are talking about instructed expressions, in order to get a clear answer, we have set the following hypothesis.

H1: Between the instructed expressions and emotional states (naturally evoked) there are differences in the degree of classification success.

We conducted an experiment to accept or reject the hypothesis.

V. Measures and Procedure of Experiment 2 – Determining the Degree of Success of Classifying the Emotional State with the Help Of Students

The experimental sample consisted of 10 students (to preserve consistency with the original Jaffe sample), both men and women aged 20-25. Students have given us written permission so that we can capture them in different situations, thus enabling us to classify the natural emotional state. As a result, during the summer semester of the academic year 2016/2017, students were doing their natural activities while we could capture them from different camera distances (maximum distance when the system is capable of recognizing face of student and realizing classification of the emotional state is 7.5 meters) without the students having a clue about it. So they were being monitored for example during test writing, test answering (Fig. 7), announcement of exam results etc. In this way, we were able to classify all 6 (or 7 - neutral states) emotional states several times, thus achieving the total number of 210 classified emotional states (explanation: 3 recurrent classification of 7 states with 10 students = 210 classification).

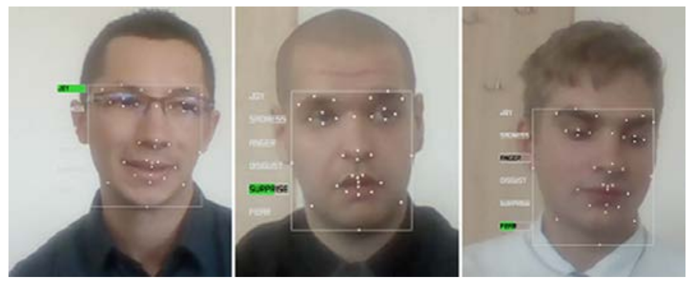


Fig. 7. Selection of student responses and classified emotional states during the test response.

We have chosen this method to be able to confront unequivocally the results obtained with the Jaffe classification results. The entire course of the experiment was under the supervision of a psychologist who subsequently determined whether or not it was actually a classified emotional state. Table V shows the desired emotions that we received by observing student emotions and classification by the software.

TABLE V. Required Emotions and Emotions Recognized by Software

		Recognized Emotion by the Software						
		An	Di	Fe	Ha	Ne	Sa	Su
Requested Emotions	An	96,30	5,30	6,30	1,90	0,60	0,30	1,20
	Di	0,00	89,63	6,70	5,20	10,40	0,00	2,80
	Fe	2,90	5,00	98,78	2,90	5,70	8,60	11,40
	Ha	0,00	0,00	6,10	99,60	0,00	11,10	10,10
	Ne	7,10	0,00	1,10	1,10	99,36	5,10	3,10
	Sa	0,00	2,00	1,10	1,40	10,00	81,25	0,80
	Su	3,00	3,40	8,10	1,10	1,50	0,10	97,12

The overall average classification score was 94.58%. We determined it as follows: We added all the percentage values of the results for the individual expressions on the diagonal in table V and divided the number of emotional expressions (number 7).

VI. Results and Discussion – Experiment 2

As we can see from Table V, we have achieved completely different results than in the first experiment, where we subjected the classification of the emotional state of the subjects' photographs from the Jaffe database to the instructed expressions. Experiment number

2 has shown that the software has been able to classify the faces of students under different lighting conditions (a total of 3 recurring classification) and therefore, if real-time face detection conditions are appropriate, the software can seamlessly classify the subject's emotions as well. At the same time, we can accept the H1 hypothesis, because there is a demonstrable difference in classification between the instructed expressions and emotional states that are naturally induced.

As it has already been highlighted in the previous sections of our contribution, to verify the percentage of success rate of the proposed automatic emotional classification system a dataset of instructed expressions is often used. The question, however, is how really these instructed expressions are similar to the emotions of human that are expressed in real life (to the best knowledge of the authors, this assumption has never been tested). For this reason, Abramson [1] as well as we in this study examined the faces of the students in real situations. Our aim was classification and evaluation of the emotional state of fear and anger in real situations, and following, to compare them with the instructed expressions from dataset. From the results we can see the significant deviations between the requested (instructed) expressions and the expressions that were captured in the real situation and subsequently classified. In the case of classification of emotional states in real situations, it has been not only detected the faces of the students but also their emotional states, being classified with a high percentage of success. These results suggest that there are significant differences in classification between the instructed and emotional states in real time situation. Therefore, we propose searching for other (more sophisticated) options to determine the percentage of success rate of systems for recognition and classification of the emotional state and not to rely only on the traditional methods of comparing the instructed expressions. Our findings also point to the fact that in real life the classification of the emotional state is strongly dependent on the information that the human individual processes from the surrounding environment. From the results, it is clear that the most striking qualitative difference between the instructed and real emotional states is in the case of emotional state - anger. An emotional state of anger can be detected in a real facial situation with a 96.30% of success rate, while in the case of instructed expressions it obtained a 0% of success rate in classification from database Jaffe. Similarly, it was the case also of other classified emotions. This novel finding suggests that instructed stimuli (anger, disgust, fear, sadness or surprise)can convey substantial ambiguity in classification process from database pictures.

VII. Conclusion

At present, recognition problematic (detection and extraction) is one of the fundamental areas of scientific research in the intelligent systems. The use is really wide - from access security (laptops, controlled entry to the room), through character recognition for document scanning (OCR) to virtual reality (Second Life game). However, the implementation of a recognition system, for example, in the education process, opens up a new area of pedagogical-scientific research - the classification of the emotional state of the user. Classification thus became the last (third) phase of the recognition system. Classifying a specific emotional state of the user brings us new opportunities, a new perspective on the issue, a better understanding of what is happening to the user within a particular stage of the learning process.

In the publication, we pointed out the determination of classification success degree with our proposed solution. Normally, the method of determining the classification level as described in Experiment 1 is used as a standard, as is the detection of the subject's face and the subsequent classification of the instructed face expression. However, due to the negative results from Experiment 1, we were looking for another method of determining the classification level to confirm or reject the hypothesis. The hypothesis has not been confirmed, so we can conclude, based on the results of Experiment 2, that there really are large differences in the degree of classification between instructed (artificially induced) emotional expressions and natural (uncontrolled) emotional states. This is a positive finding for us because we intend to permanently link the system to the LMS Moodle e-learning environment and determine the emotional status of the students. In this way, we could then understand how students feel during the test period, determine what they are experiencing and look for a common solution to their problems together with them.

Acknowledgment

This work was supported by the IBM Slovakia [C-16-103/0001-01, Pilot project with Constantine the Philosopher University in Nitra: Modelling the behaviour of users based on data mining with support of IBM Bluemix].

References

[1] L. Abramson, I. Marom, R. Petranker & H. Aviezer,"Is fear in your head? A comparison of instructed and real-life expressions of emotion in the face and body," in *Emotion, 17*(3), 2017, 557-565. doi:10.1037/emo0000252

[2] W.N.W. Ahmad, N.M. Ali, "A Study on Persuasive Technologies: The Relationship between User Emotions, Trust and Persuasion", in *International Journal of Interactive Multimedia and Artificial Intelligence*, Volume in Press, Issue in Press, 2018, 1-5. doi: 10.9781/ijimai.2018.02.010

[3] P. Arnau-González, M. Arevalillo-Herráez & N. Ramzan, "Fusing highly dimensional energy and connectivity features to identify affective states from EEG signals," in *Neurocomputing, 244*, 2017, 81-89. doi:10.1016/j.neucom.2017.03.027

[4] K. Bahreini, R. Nadolski & W. Westera, "Towards multimodal emotion recognition in e-learning environments," in *Interactive Learning Environments, 24*(3), 2016, 590-605. doi:10.1080/10494820.2014.908927

[5] Y. Balasubramanian, K. Sivasankaran & S.P. Krishraj, "Forensic video solution using facial featurebased synoptic video footage record," in *IET Computer Vision, 10*(4), 2016, 315-320. doi:10.1049/iet-cvi.2015.0238

[6] V. Bettadapura, G. Schindler, T. Ploetz & I. Essa, "Augmenting bag-of-words: Data-driven discovery of temporal and structural information for activity recognition," Paper presented at the *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* 2013, 2619-2626. doi:10.1109/CVPR.2013.338

[7] A. Bhattacharya, D. Choudhury & D. Dey, "Edge-enhanced bi-dimensional empirical mode decomposition-based emotion recognition using fusion of feature set," in *Soft Computing,* 2016, 1-15. doi:10.1007/s00500-016-2395-4

[8] I. Cohen, N. Sebe, F.G. Cozman, M.C. Cirelo & T.S. Huang, "Learning bayesian network classifiers for facial expression recognition using both labeled and unlabeled data," Paper presented at the *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1*, 2003, I/595-I/601.

[9] D.T. Cordaro, D. Keltner, S. Tshering, D. Wangchuk & L.M. Flynn, "The voice conveys emotion in ten globalized cultures and one remote village in bhutan," in *Emotion, 16*(1), 2016, 117-128. doi:10.1037/emo0000100

[10] C. Darwin & F. Darwin, "The expression of the emotions in man and animals. *The expression of the emotions in man and animals,* 2009, 1-401. doi:10.1017/CBO9780511694110

[11] M.A. DiGirolamo & J.A. Russell, "The emotion seen in a face can be a methodological artifact: The process of elimination hypothesis," in *Emotion, 17*(3), 2017, 538-546. doi:10.1037/emo000024

[12] P. Ekman, "An argument for basic emotions," in *Cognition and Emotion, 6*(3-4), 1992, 169-200. doi:10.1080/02699939208411068

[13] P. Ekman, "What scientists who study emotion agree about," in *Perspectives on Psychological Science, 11*(1), 2016, 31-34. doi:10.1177/1745691615596992

[14] P. Ekman & D. Cordaro, "What is meant by calling emotions basic," in *Emotion Review, 3*(4), 2011, 364-370. doi:10.1177/1754073911410740

[15] P. Ekman & W.V. Friesen, "Felt, false, and miserable smiles," in *Journal of*

Nonverbal Behavior, 6(4), 1982, 238-252. doi:10.1007/BF00987191

[16] P. Ekman & W.V. Friesen, *Unmasking the face: A guide to recognizing emotions from facial clues*. Prentice-Hall, Englewood Cliffs, NJ (1975)

[17] N.L. Etcoff & J.J. Magee, "Categorical perception of facial expressions," in *Cognition, 44*(3), 1992, 227-240. doi:10.1016/0010-0277(92)90002-Y

[18] B. Fasel & J. Luettin, "Automatic facial expression analysis: A survey," in *Pattern Recognition, 36*(1), 2003, 259-275.

[19] S. Fleck & W. Straßer, "Smart camera based monitoring system and its application to assisted living," Paper presented at the *Proceedings of the IEEE, 96*(10), 2008, 1698-1714. doi:10.1109/JPROC.2008.928765

[20] D. Goren & H.R. Wilson, "Quantifying facial expression recognition across viewing conditions," in *Vision Research, 46*(8-9), 2006, 1253-1262. doi:10.1016/j.visres.2005.10.028

[21] U. Hess & P. Thibault, "Darwin and emotion expression," in *American Psychologist, 64*(2), 2009, 120-128. doi:10.1037/a0013386

[22] M.S. Hossain & G. Muhammad, "Audio-visual emotion recognition using multi-directional regression and ridgelet transform," in *Journal on Multimodal User Interfaces, 10*(4), 2016, 325-333. doi:10.1007/s12193-015-0207-2

[23] M.A. Jarwar & I. Chong, "Exploiting IoT services by integrating emotion recognition in web of objects," Paper presented at the *International Conference on Information Networking,* 2017, 54-56. doi:10.1109/ICOIN.2017.7899474

[24] D. Jayatilake, T. Isezaki, Y. Teramoto, K. Eguchi & K. Suzuki, "Robot assisted physiotherapy to support rehabilitation of facial paralysis," Paper presented at the *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 22*(3), 2014, 644-653. doi:10.1109/TNSRE.2013.2279169

[25] S.M. Lajevardi & Z.M. Hussain, "Feature selection for facial expression recognition based on optimization algorithm," Paper presented at the *Proceedings of 2009 2nd International Workshop on Nonlinear Dynamics and Synchronization, INDS 2009,* 2009, 182-185.

[26] R.W. Levenson, "Basic emotion questions," in *Emotion Review, 3*(4), 2011, 379-386. doi:10.1177/1754073911410743

[27] L. Likforman-Sulem, A. Esposito, M. Faundez-Zanuy, S. Clemencon & G. Cordasco, "EMOTHAW: A novel database for emotional state recognition from handwriting and drawing," Paper presented at the *IEEE Transactions on Human-Machine Systems, 47*(2), 2017, 273-284. doi:10.1109/THMS.2016.2635441

[28] K. Matsuno, C. Lee, S. Kimura & S. Tsuji, "Automatic recognition of human facial expressions," Paper presented at the *IEEE International Conference on Computer Vision,* 1995, 352-359.

[29] J.D. Mayer, P. Salovey, D.R. Caruso & G. Sitarenios, "Emotional intelligence as a standard intelligence," in *Emotion*, 1(3), 2001, 232-242. doi:10.1037/1528-3542.1.3.232

[30] D. McDuff, R. El Kaliouby & R.W. Picard, "Crowdsourcing facial responses to online videos," Paper presented at the *IEEE Transactions on Affective Computing, 3*(4), 2012, 456-468. doi:10.1109/T-AFFC.2012.19

[31] A. Mehrabian, "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament," in *Current Psychology, 14*(4), 1996, 261-292.

[32] N. Mehta & S. Jadhav, " Facial emotion recognition using log gabor filter and PCA," Paper presented at the *Proceedings - 2nd International Conference on Computing, Communication, Control and Automation, ICCUBEA 2016,* 2017, doi:10.1109/ICCUBEA.2016.7860054

[33] G. Muhammad & M.F. Alhamid, "User emotion recognition from a larger pool of social network data using active learning," in *Multimedia Tools and Applications, 76*(8), 2017, 10881-10892. doi:10.1007/s11042-016-3912-2

[34] J.C. Mundt, A.P. Vogel, D.E. Feltner & W.R. Lenderking, "Vocal acoustic biomarkers of depression severity and treatment response," in *Biological Psychiatry, 72*(7), 2012, 580-587. doi:10.1016/j.biopsych.2012.03.015

[35] R. Plutchik, (2001). *The Nature of Emotions.* *Am. Scientist*, vol. 89, 344

[36] N. Rathee, A. Vaish & S. Gupta, "Adaptive system to learn and recognize emotional state of mind," Paper presented at the *Proceeding - IEEE International Conference on Computing, Communication and Automation, ICCCA 2016,* 2017, 32-36. doi:10.1109/CCAA.2016.7813726

[37] L.D. Riek & R.N.M. Watson, "The age of avatar realism: When seeing shouldn't be believing," Paper presented at the *IEEE Robotics and Automation Magazine, 17*(4), 2010, 37-42. doi:10.1109/MRA.2010.938841

[38] J.A. Russell, "Affective space is bipolar," in *Journal of Personality and Social Psychology, 37*(3), 1979, 345-356. doi:10.1037//0022-3514.37.3.345

[39] C. Torres-Valencia, M. Álvarez-López & Á. Orozco-Gutiérrez, "SVM-based feature selection methods for emotion recognition from multimodal data," in *Journal on Multimodal User Interfaces, 11*(1), 2017, 9-23. doi:10.1007/s12193-016-0222-y

[40] P. Visutsak, "Template-based affine registration of autistic brain images," Paper presented at the *IJCCI 2015 - Proceedings of the 7th International Joint Conference on Computational Intelligence, 2,* 2015, 188-192.

[41] Y. Yang, C. Fairbairn & J.F. Cohn, "Detecting depression severity from vocal prosody," Paper presented at the *IEEE Transactions on Affective Computing, 4*(2), 2013, 142-150. doi:10.1109/T-AFFC.2012.38

[42] A.W. Young, D. Rowland, A.J. Calder, N.L. Etcoff, A. Seth & D.I. Perrett, "Facial expression megamix: Tests of dimensional and category accounts of emotion recognition," in *Facial expression recognition: Selected works of Andy Young*, 2016, 68-110. doi:10.4324/9781315715933

[43] M. Zhang, D. Lee, A. Desai, K.D. Lillywhite & B.J. Tippetts, "Automatic facial expression recognition using evolution-constructed feature," in. Paper presented at the *10th International Symposium on Visual Computing*, ISVC 2014, 2014, 282-291

[44] M. Magdin, F. Prikler. Real Time Facial Expression Recognition Using Webcam and SDK Affectiva, International Journal of Interactive Multimedia and Artificial Intelligence, (2017), http://dx.doi.org/10.9781/ijimai.2017.11.002

M. Magdin

He works as a professor assistant at the Department of Computer Science. He deals with the theory of teaching informatics subjects, mainly implementation interactivity elements in e-learning courses, face detection and emotion recognition using a webcam. He participates in the projects aimed at the usage of new competencies in teaching and also in the projects dealing with learning in virtual environment using e-learning courses.

F. Prikler

was student at the Department of Computer Science. He deals with the theory of teaching informatics subjects, mainly face detection and emotion recognition using a webcam. At present he works as programmer in the company Muhlbauer Technologies.

LA UNIVERSIDAD
EN INTERNET

Rectorado
Avenida de la Paz, 137
26006 Logroño (La Rioja)
t (+34) 941 21 02 11

www.unir.net