International Journal of Interactive Multimedia and Artificial Intelligence

September 2019, Vol. V, Number 6
ISSN: 1989-1660

unir LA UNIVERSIDAD EN INTERNET

"We must address, individually and collectively, moral and ethical issues raised by cutting-edge research in artificial intelligence and biotechnology, which will enable significant life extension, designer babies, and memory extraction."

Klaus Schwab

Special Issue on Use Cases of Artificial Intelligence, Digital Marketing and Neuroscience

INTERNATIONAL JOURNAL OF INTERACTIVE MULTIMEDIA AND ARTIFICIAL INTELLIGENCE
ISSN: 1989-1660 –VOL. 5, NUMBER 6

Dr. Monique Janneck, Lübeck University of Applied Sciences, Germany
Dr. Carina González, La Laguna University, Spain
Dr. Mohammad S Khan, East Tennessee State University, USA
Dr. David L. La Red Martínez, National University of North East, Argentina
Dr. Juan Francisco de Paz Santana, University of Salamanca, Spain
Dr. Héctor Fernández, INRIA, Rennes, France
Dr. Yago Saez, Carlos III University of Madrid, Spain
Dr. Guillermo E. Calderón Ruiz, Universidad Católica de Santa María, Peru
Dr. Giuseppe Fenza, University of Salerno, Italy
Dr. José Miguel Castillo, SOFTCAST Consulting, Spain
Dr. Moamin A Mahmoud, Universiti Tenaga Nasional, Malaysia
Dr. Madalena Riberio, Polytechnic Institute of Castelo Branco, Portugal
Dr. Juan Antonio Morente, University of Granada, Spain
Dr. Holman Diego Bolivar Barón, Catholic University of Colombia, Colombia
Dr. Manik Sharma, DAV University Jalandhar, India
Dr. Sara Rodríguez González, University of Salamanca, Spain
Dr. Elpiniki I. Papageorgiou, Technological Educational Institute of Central Greece, Greece
Dr. Edward Rolando Nuñez Valdez, Open Software Foundation, Spain
Dr. Juha Röning, University of Oulu, Finland
Dr. Luis de la Fuente Valentín, Universidad Internacional de La Rioja - UNIR, Spain
Dr. Paulo Novais, University of Minho, Portugal
Dr. Giovanny Tarazona, Francisco José de Caldas District University, Colombia
Dr. Sergio Ríos Aguilar, Corporate University of Orange, Spain
Dr. Fernando López, Universidad Internacional de La Rioja - UNIR, Spain
Dr. Mohamed Bahaj, Settat, Faculty of Sciences & Technologies, Morocco
Dr. Javier Martínez Torres, Centro Universitario de la Defensa de Marín, Escuela Naval Militar, Spain
Dr. Abel Gomes, University of Beira Interior, Portugal
Dr. Edgar Henry Caballero Rúa, Inforfactory SRL, Bolivia
Dr. Víctor Padilla, Universidad Internacional de La Rioja - UNIR, Spain
Ing. María Monserrate Intriago Pazmiño, Escuela Politécnica Nacional, Ecuador
Dr. José Manuel Saiz Álvarez, Tecnológico de Monterrey, México
Dr. Alejandro Baldominos, Universidad Carlos III de Madrid, Spain

Editor's Note

Towards an Intelligent Society: Advances in Marketing and Neuroscience

Artificial Intelligence (AI) is becoming the main character of what is now being called the fourth industrial revolution. Its role is gaining importance in everyday life, and the reason is straightforward: it offers a wide range of possibilities to make life easier in many different aspects.

However, AI is also at the center of a debate since some studies anticipate a radical transformation of the future industry, where some scenarios involve the destruction of millions of jobs. However, it is generally acknowledged that many new, more specialized, job positions will be created, as it has happened throughout history with other technological revolutions (see The Future of Work special issue in Nature [1]). Nevertheless, uncertainties remain on the net impact on employment. While this debate is certainly useful, at this point we find it better to focus on a general objective assessment of how AI is impacting the world and the society.

From a global perspective, a clear statement can be made: Artificial Intelligence can have an immense positive impact on societies. Some of this impact is already unveiling in recent years and is particularly observable in the fields of health and medicine [2, 3], where AI is turning into a key player at the time of diagnosing diseases at an early stage or developing new medicines and specialized treatment. Personalized medicine is probably the biggest breakthrough of the coming years, and AI is taking an active role to push this field forward.

However, medicine is not the only field where AI can enhance the process of personalization and customization. Marketing is certainly another good scope of application, where intelligent software can help knowing the target audience and offering them what they need in response. In this sense, we can already find intelligent devices that are able to make predictions about our behaviors [4]. From a more practical perspective, AI can facilitate the sales process [5, 6, 7], dealing with most routinely procedures such as information tasks or documentation handling, which become streamlined and turn cheaper.

Additionally, AI enable brands to offer a commitment to their customers [8, 9]. This commitment will trigger some kind of emotional response in the customer depending on the perceived quality and the surrounding circumstances. In fact, AI can provide a software with means to detect the feelings arising during an interaction or engagement with customers, and determining whether these feelings are positive, negative or neutral.

In summary, with tools such as those outlined, AI would be capable to allow [10]:

- Knowing your tastes, desires and expectations as a consumer, as well as predicting your needs.
- Analyzing your behavior and consumer habits when browsing the Internet.
- Studying the emotional response during an interaction.
- Anticipating trends.
- Offering to selected customers the products or services they demand in a timely manner.
- Using the most effective channels to enhance the consumer experience to the best possible.
- Customizing communications to enhance the customers feelings.

All of these achievements are easy to reach when virtual assistants are able to retrieve a complete picture of a customer's behavior, tastes and way of interacting. By these means, customers are not offered a generic experience but rather a unique result fitting their needs.

Since all data is not available in digital formats, the speed in which they are generated, processed and analyzed is dizzying. This velocity is one of the key aspects when it comes to applying AI to marketing, and especially when monetizing publications or services in social networks or the Internet.

This Special Issue focuses in cases that explore the relationship between Artificial Intelligence and marketing, as well as neuroscience. AI can be combined with specific neuroscience techniques to achieve a more successful and profitable neuromarketing. For this Special Issue, we have found that descriptions of successful use cases are highly valuable to help researchers identify fields where novel applications of AI can enhance the outcome of digital marketing and neuroscience.

The first paper of the Special Issue is entitled "Data and Artificial Intelligence Strategy: A Conceptual Enterprise Big Data Cloud Architecture to Enable Market-Oriented Organizations" and is written by C. Moreno, R. Alberto-Carrasco and E. Herrera-Viedma, a group of researchers from the University of Granada and from the Complutense University of Madrid, both in Spain. In their paper, they present an AI cloud architecture capable to help companies to switch from a descriptive to a prescriptive usage of data, bringing agility and many possibilities to acquire new marketing insights to these companies.

The second article has been written by A.Goli, H. Khademi-Zare, R. Ravakkoli-Moghaddam and A. Sadeghieh, researchers of the Yazd University and the University of Tehran in Iran, and is entitled "An Improved Artificial Intelligence Based on Gray Wolf Optimization and Cultural Algorithm to Predict Demand for Dairy Products: A Case Study". This paper combines different AI techniques to carry out the prediction of dairy product demand in Iran, testing the improvement of neural networks with the help of novel meta-heuristic algorithms such as gray wolf optimization and cultural algorithm.

The following research work, "The Promotion of Graduate Programs through Clustering Prospective Students" has been carried out by R.M. Cantón-Croda, D.E. Gibaja-Romero and F.R. Castillo-Villar, a group of researchers of UPAEP, Mexico. In their work, authors perform clustering, which is a well-known AI technique to tackle the problem of the promotion of academic programs, a crucial problem particularly interesting for private universities, suggesting how a deep analysis can be used to design better promotion strategies.

The following paper analyzes the state of emotional state of the users. Specifically, the work of M. Magdin, T. Sulka, J. Tomanová and M. Vozár, researchers from Constantine the Philosopher University in Nitra (Slovakia) is entitled "Voice Analysis Using PRAAT Software and Classification of User Emotional State." Their paper proposes the classification of the emotional state of a person based on the voice track analysis, a useful application of AI with potential direct benefits to marketing.

The following three works have a point in common: they belong to the field of AI applied to medicine and neuroscience. The first of them, entitled "A User-centered Smartphone Application for Wireless EEG and its role in Epilepsy," has been written by S. Ahufinger, P. Balugo, M.M. González, H. González and P. Herrero, researchers of Polytechnic University of Madrid and San Carlos Clinical Hospital, both in Spain. The paper presents the design and evaluation of a wireless EEG smartphone application resulting from a user-centered design,

DOI: 10.9781/ijimai.2019.07.005

that fits the clinical and research needs, and explore its application to epilepsy diagnosis, a field of a large medical interest.

Moreover, C. Pruenza, M.T. Solano, J. Díaz, R. Arroyo and G. Izquierdo, researchers from the Knowledge Engineering Institute, the University Hospital Quirónsalud Madrid, the Ruber Juan Bravo Hospital and the Vithas NISA Hospital, all of them in Spain, have partnered to author the following, entitled "Model for prediction of progression in Multiple Sclerosis". In their paper, they develop a personalized prediction model of three stages of the multiple sclerosis disease, which can be used as a clinical decision support system, using AI and big data techniques.

The next paper has been published by J.M. Lombardo, M.A. Lopez, M. Lopez, M. León, F. Miron, J. Arranbarri and D. Alvarez, a group of researchers from the FIDESOL Foundation in Spain, and is entitled "Natural interaction Technologies, virtual reality and artificial intelligence for gait disorders." Parkinson's disease is the most common degenerative disorder after Alzheimer's disease, and in this paper authors propose a tool that can provide health staff with means to analyze, evaluate and monitor the progress of patients' disorders, as well as the design of custom rehabilitation sessions in patients with Parkinson's disease.

The last paper is a work by K. Aufderhaar, M. Schrepp and J. Thomaschewski entitled "Do Women and Men Perceive User Experience Differently?". In this paper, authors study how women and men differ when rating the user experience of different websites using standardized questionnaires. Interestingly, they conclude that gender is not a significant factor, while personal attitudes and preferences are.

This Special Issue has gathered a diverse set of applications of Artificial Intelligence to the fields of marketing, healthcare and neuroscience. Research of such applications is expected to keep growing in the upcoming years, revolutionizing the way in which we understand marketing and medicine, and opening new possibilities which we might not be able to imagine as of today.

The future of marketing and neuroscience looks exciting, and Artificial Intelligence is at the core of it.

F. Mochón and A. Baldominos

References

[1] "The Future of Work", Nature 550, 315 (2017). https://www.nature.com/news/the-future-of-work-1.22840

[2] Hamet, P. and Tremblay, J. (2017). "Artificial intelligence in medicine", Metabolism 69, pp. S36–S40.

[3] He, J., Baxter, S.L., Xu J., Xu J., Zhou, X. and Zhang, K. (2019). "The practical implementation of artificial intelligence technologies in medicine", Nature Medicine 25, pp. 30–36.

[4] Yang, Y. and Siau, K.L. (2018). "A Qualitative Research on Marketing and Sales in the Artificial Intelligence Age", Proceedings of the Thirteenth Midwest Association for Information Systems Conference. Saint Louis, Missouri May 17-18, 2018.

[5] Columbus, L. (2018). "10 Ways Machine Learning Is Revolutionizing Sales", Forbes. Available online: https://www.forbes.com/sites/louiscolumbus/2018/12/26/10-ways-machine-learning-is-revolutionizing-sales/#3c97d8813fd. (Last visited: 2019/07/05).

[6] Zaledonis, L. "Why AI Will Be Your New Best Friend in Sales", Quotable (Salesforce). Available online: https://www.salesforce.com/quotable/articles/why-AI-will-be-your-new-best-friend-in-sales/. (Last visited: 2019/07/05).

[7] Syam, N. and Sharma, A. (2018) "Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice", Industrial Marketing Management 69, pp. 135–146.

[8] Gacanin, H. and Wagner, M. (2019) "Artificial Intelligence Paradigm for Customer Experience Management in Next-Generation Networks: Challenges and Perspectives", IEEE Network 33, 2, pp. 188–194.

[9] Vieira, A. and Sehgal, A. (2017) "How Banks Can Better Serve Their Customers Through Artificial Techniques", In: Digital Marketplaces Unleashed, pp. 311–326. Springer.

[10] Milgrom, P.R. and Steven Tadelis, S. (2018). "How Artificial Intelligence and Machine Learning Can Impact Market Design", NBER Working Paper No. 24282. Issued in February. https://www.nber.org/papers/w24282.

TABLE OF CONTENTS

OPEN ACCESS JOURNAL

ISSN: 1989-1660

Data and Artificial Intelligence Strategy: A Conceptual Enterprise Big Data Cloud Architecture to Enable Market-Oriented Organisations

Caio Moreno[1], Ramón Alberto Carrasco[1], Enrique Herrera-Viedma[2]*

[1] Department of Management and Marketing, Complutense University, Madrid 28223 (Spain)
[2] Andalusian Research Institute in Data Science and Computational Intelligence, University of Granada, Granada 18071 (Spain)

Received 11 April 2019 | Accepted 25 April 2019 | Published 22 June 2019

Abstract

Market-Oriented companies are committed to understanding both the needs of their customers, and the capabilities and plans of their competitors through the processes of acquiring and evaluating market information in a systematic and anticipatory manner. On the other hand, most companies in the last years have defined that one of their main strategic objectives for the next years is to become a truly data-driven organisation in the current Big Data context. They are willing to invest heavily in Data and Artificial Intelligence Strategy and build enterprise data platforms that will enable this Market-Oriented vision. In this paper, it is presented an Artificial Intelligence Cloud Architecture capable to help global companies to move from the use of data from descriptive to prescriptive and leveraging existing cloud services to deliver true Market-Oriented in a much shorter time (compared with traditional approaches).

Keywords

Market-oriented Organisations, Big Data, Cloud Architecture, Artificial Intelligence Strategy, Data Supermarket.

DOI: 10.9781/ijimai.2019.06.003

I. Introduction

MARKET-ORIENTED (MO) enterprises have as its main objective to perfectly understand the needs of its clients to satisfy them better than the competition and, in this way, to have a competitive advantage over the said competition. In this aspect a key role is played by the data, in such a way that converted into timely knowledge and transmitted to the entire organization to work in coordination, will achieve this common goal [1]-[3].

Precisely, Business Intelligence (BI) is a business philosophy that bases decision-making on data that has been opportunely converted into knowledge process in which, Artificial Intelligence (AI) plays a key role [4][5]. In this way, there have been authors who have defined a formal framework for the effective implementation of Business Intelligence (BI) in a MO company [6].

However, the emergence of Big Data [7] has made these models obsolete, and market-oriented companies do not know how to adapt given the wide variety of products and services that composes the so-called Big Data ecosystems.

The main objective of this work, therefore, is to define a formal framework that allows Market Orientation to be effective in the context of Big Data, including the necessary components.

To achieve this goal, in section II it is introduced market orientation, and formal architectures to achieve it in a BI framework. In section III a review of this proposed formal frameworks is made, adding the appropriate components that allow carrying out Market Orientation in a Big Data context. Finally, in sequence, some conclusions and future work are presented.

* Corresponding author.
E-mail address: viedma@decsai.ugr.es

II. Market-Oriented Enterprise Strategy Based on Data

The data-driven business strategy has been in existence for several decades. An example of the theoretical framework of strategic planning is showed in Fig. 1, which has three components based, to a greater or lesser extent, on the collection and analysis of data:

1. **Specification of objectives**: Normally, increase the value of the company, present and future, based on doing the same with the value of the clients.
2. **Decision Making (DM)**: Based on the objectives specified through the generation and evaluation of the different strategies.
3. **Monitoring of the results**: Quantifying to what extent the objectives have been reached after the DM.

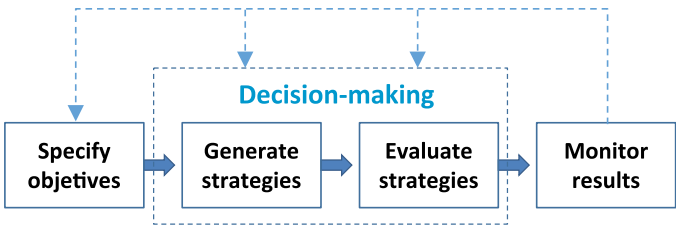


Fig. 1. The formal process of strategic planning (based on [1]).

In the eighties and nineties of the twentieth century, companies were in a general situation of lack of differentiation, since they offered products/services (P/S) of very similar quality with differences not perceptible for clients. At this juncture, Market Orientation, proposal, arises as the generation in the entire organization of market intelligence about the current and future needs of customers, the dissemination of information by all departments and the ability to respond to it throughout the organization [2]. In this definition three dimensions are identified:

- **Generation of market intelligence.** It consists of the responsibility of the entire organization to obtain a marketing information system on the present and future needs of customers, as well as distributors, suppliers, lobbyists, competitors and macro environment in general.
- **Dissemination of information.** Among the different areas of the organization in order for them to work in common in the same objective.
- **Response to information.** In such a way that this knowledge obtained and disseminated becomes actions that will result in obtaining competitive advantages in the organization.

Market-Oriented (MO) companies are committed to understanding both the needs of their **customers** and the capabilities and plans of their **competitors**, through the processes of acquiring and evaluating **market information** in a systematic and anticipatory manner [3]. Therefore, a key concept for MO is, undoubtedly, Business Intelligence (BI) that can be defined as [4][5]: the process of converting data into knowledge and this into actions or decisions to create the competitive advantage of the business. In Fig. 2 it is possible to observe a global scheme of the different components that encompasses the concept of BI focused on MO, proposed by Stone and Woodcock [6].

The different components are explained in more detail below:

1. **Data.** It would include the internal data of the organization such as: the contents in the operative processes of the Customer Relationship Management (CRM) and Enterprise Resource Planning (ERP) systems; those related to bidirectional communications with customers and society in general commonly supported in collaborative and social CRM; third-party data; qualitative data including those from typical secondary sources; etc.
2. **Data Warehouse (DW)**. They are the databases specifically designed for analysis and they would include much of the raw data discussed, through the so-called ETL (Extraction, Transformation and Loading) processes [8]. Data marts are a portion of the DW with a departmental purpose; in our case, the one that interests us is the marketing data mart that would coincide, to a large extent, with the concept widely used in the past of marketing information systems.
3. **Insight Generation**. It would include, among others, the typical models of the relational strategy such as valuation, identification, recruitment, retention and client development. In large part, this component is supported by Data Mining or Data Science of which modeling phase is, in turn, supported largely by Artificial Intelligence (AI), and more specifically, by Machine Learning (ML) systems.
4. **Action**. This component is related to business decisions, which is in TD processes, based on the knowledge discovered in the previous component (with visual representation, scorecards, qualitative

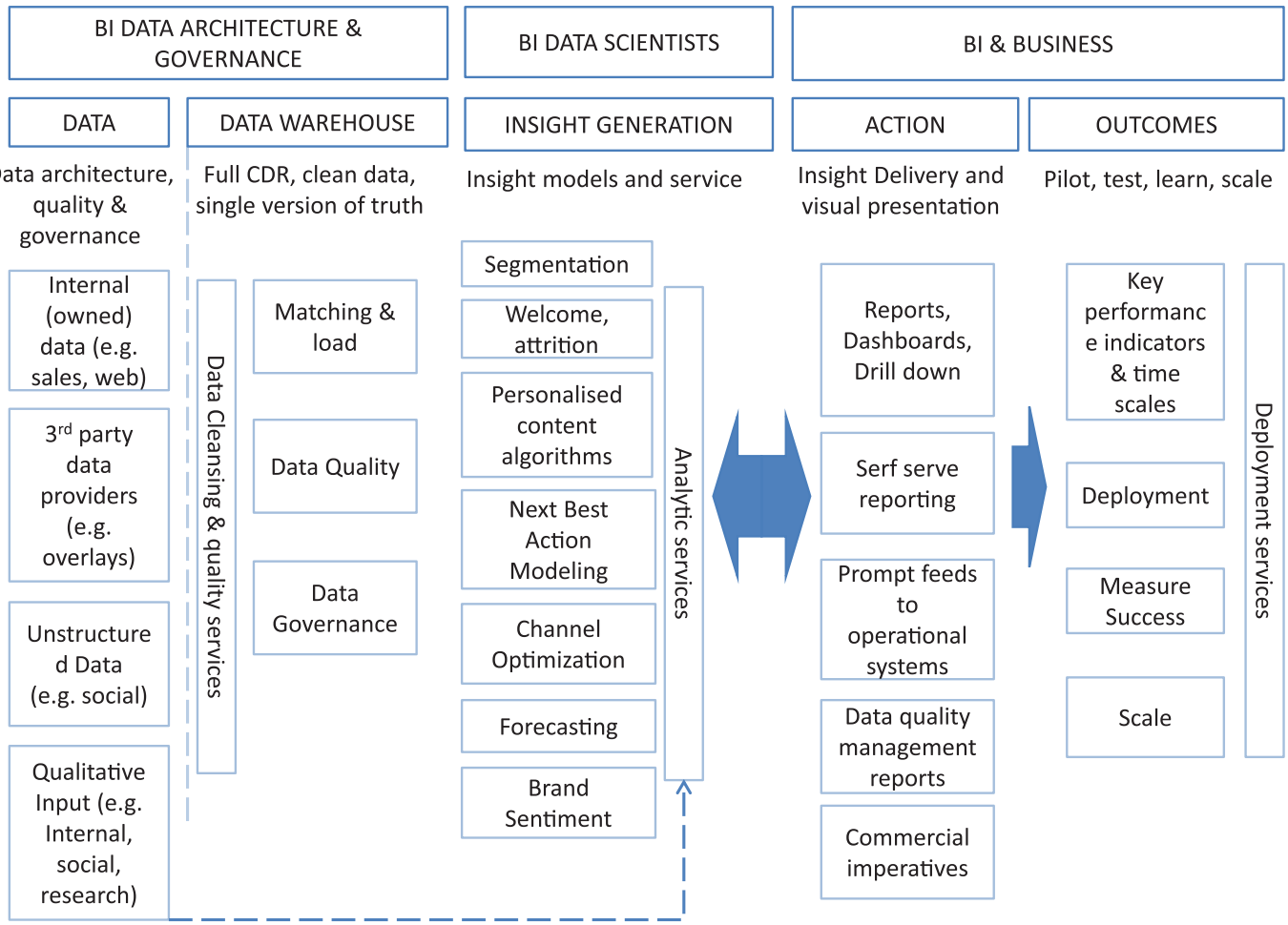


Fig. 2. BI applied to the MO (based on [6]).

reports, etc.). Many of these decisions will end up in the form of specific operations in the corresponding CRM systems.

5. **Outcomes**. In such a way that it is possible to discern to what extent it has been successful in the actions carried out.

III. Definition of a Conceptual Framework for Market Oriented Organizations Based on an Artificial Intelligence Cloud Architecture

Most of the modern organizations have invested, for many years, in a conceptual framework similar to the one proposed by Stone and Woodcock [6], and explained in the previous section, as the way to implement their Market Orientation. However, with the increasing amount of data volume, with a variety of formats (unstructured, semi-structured and structured) that are produced (and therefore they have to be processed) at high velocity, organizations are now obligated to materialize a new Big Data paradigm [9] capable to adapt to these new challenges. Although, there are several proposals for the use of Big Data adapted to certain sectors [10][11], the new components of this MO architecture, that solves these Big Data challenges, have not been fully formalized until now.

Thus, in this section, it is proposed a new formal framework for companies that want to adapt to a Big Data MO strategy. For this purpose, the formal framework, presented in Fig. 2, is redefined by adding the necessary components. The new proposed framework can be seen in Fig. 3. The changes introduced are explained below:

A. Data

In this architectural data layer, a fundamental new source for companies, that comes from the called Internet of Things (IoT), also named Internet of Everything, is incorporated. It is a new technology paradigm envisioned as a global network of machines and devices capable of interacting with each other [12]. A practical example of the use of IoT data is the one called smart trains, where the railway industry is exploiting the opportunities to use IoT data. This new data will enable predictive maintenance, smart infrastructure, advanced monitoring of assets, video surveillance systems, energy efficiency, etc. [13].

Special emphasis must be placed on the usefulness of the market data that this new technology will bring for MO companies [14]. Therefore, the value proposition of IoT data in MO companies is to receive real-time data coming directly from connected objects on the internet, capable to provide rich market insights to improve business outcomes.

B. Data Management Solutions for Analytics

Another addition to the architecture is this new layer called Data Management Solutions for Analytics (DMSA) instead of the DW layer. Although the DW component is still valid at present for the functions indicated in section II, it is clear that the large volume of data, which Big Data implies with such data, produced in real-time and with a variety of formats (more and more, among other factors, by the greater implantation of IoT sensors, as seen in the previous point), cannot be included in the conventional DW with the conventional ETL processes discussed.

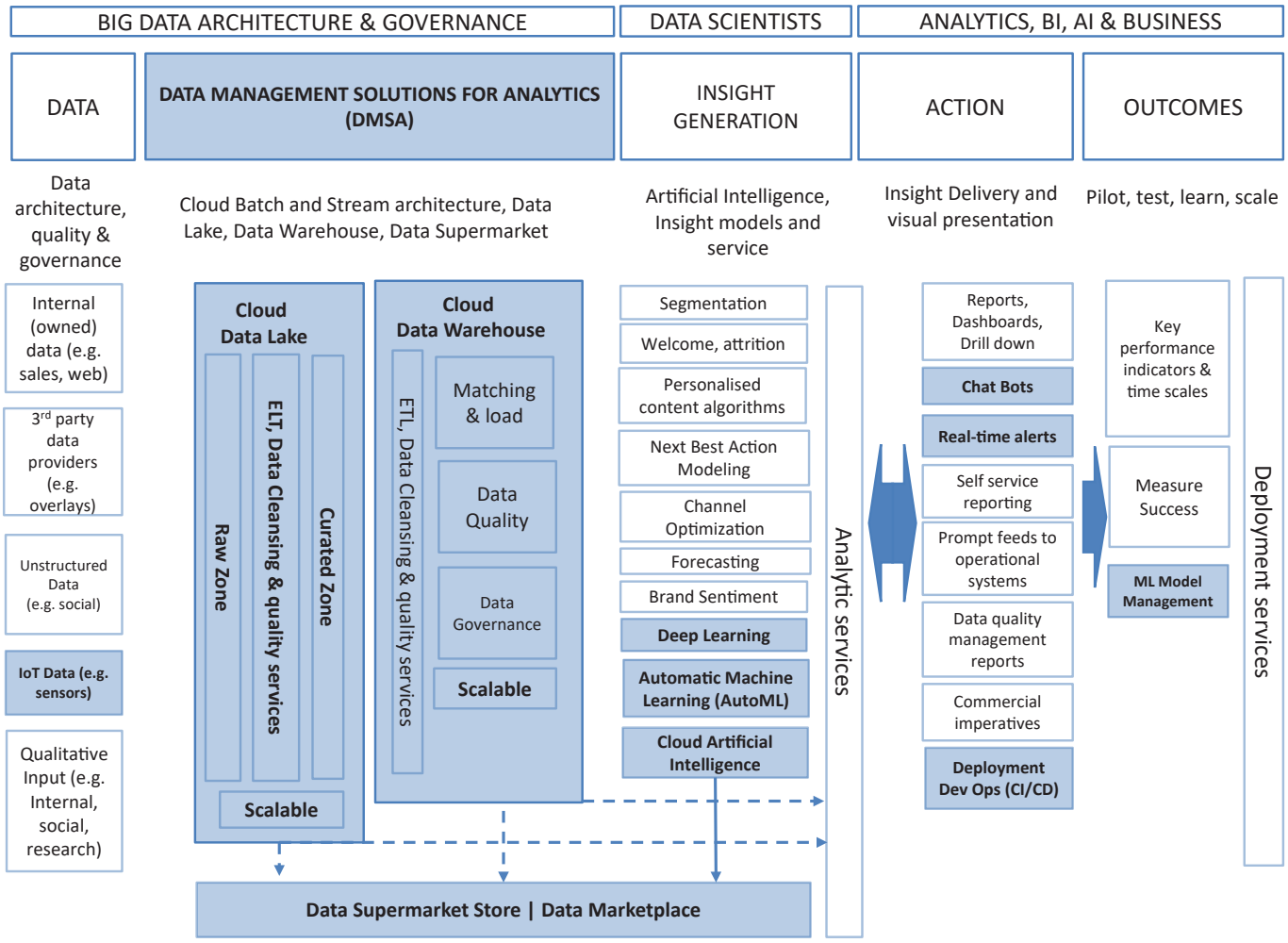


Fig. 3. A conceptual Big Data MO strategy architecture.

This nomenclature has been used by Gartner that defines DMSA as "*a complete software system that supports and manages data in one or many file management systems, most commonly a database or multiple databases*" [15]. Gartner also evaluates the different software products that support the DMSA. This evaluation includes only the main products and many of them are specifically focused on the cloud (Google, Amazon AWS, Microsoft, Alibaba Cloud…) [16]. In effect, the cloud allows the ingestion of this large volume of data.

According to the National Institute of Standards and Technology (NIST) Cloud Computing is [17]"*a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction*". Cloud providers, according to Iorga & Karmel [18], provides "*economies of scale, cutting-edge technology advancements, and a higher concentration of expertise enabling cloud providers to offer state-of-the-art cloud ecosystems that are resilient, self-regenerating, and secure—far more secure than the environments of consumers who manage their own systems*". Cloud is a key weapon to allow companies to adjust to market demand; the data and the market will grow, and the data architecture will have to grow to adapt. Most of the main cloud providers (Microsoft, Google and Amazon) in the last years invested heavily to create state-of-the-art data and AI services to enable small, medium and large organizations to use the cloud to transform their business into MO companies.

At present, there are two main pieces included in the DMSA: The Data Lake and the DW in which the cloud component will be fundamental. They are explained in more detail below:

A **Data Lake** is a centralized data repository that allows enterprises to collect a larger volume and variety of data and store all structured, semi-structured and unstructured data at any scale without the rigidity and overhead of traditional DW architectures [19]. Very often this data repository runs on a cloud provider in which case it is usually called Cloud Data Lake. Data Lakes have a high degree of flexibility and scalability in such a way that it allows MO companies to ingest all the market data in general, including that of their own customers and competitors. This is possible thanks to the characteristics of these repositories [20][21] that scale as much as possible; plug-in disparate data sources; acquire high-velocity data; store in native format; do not worry about original data schema; run massively parallel SQL queries; and allow advanced algorithms like deep learning that will power real-time decision analytics (as it will be seen in the Insight Generation layer).

In a Data Lake there are two main levels:

- **Raw Zone.** This tier has the capacity of storing data as-is, i.e., raw data, without having to first structure these data (see Fig. 4). Therefore, raw data are data not classified when they are stored. As a result, data preparation, cleansing, and transformation tasks are eliminated. The meaning of this storage philosophy is that [21]: "*Storing data in its rawest form enables us to find answers from the data for which we do not know the questions yet; whereas a traditional data warehouse is optimized for answering questions that we already know*". At a technological level, this layer is a folder stored thanks to distributed file systems such as HDFS (Hadoop File System).
- **Curated Zone.** Where the result of different types of analytics, using Big Data processing engines, is stored (see Fig. 5). Only high-value data are stored as files inside the Curated Zone in a folder (often in HDFS); it means data that passed quality data checks, i.e., data cleansing, data transformation and data enrichment [21]. This data, available in the curated zone, can be used to ad-hoc dashboards and visualizations, real-time analytics, and machine learning to guide better decisions.

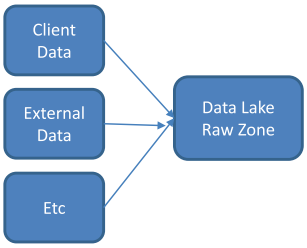


Fig. 4. Plug in disparate data sources into the Cloud Data Lake: Raw Zone.

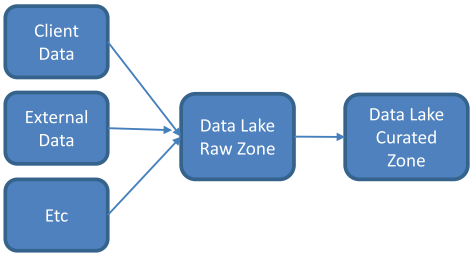


Fig. 5. Plug in disparate data sources into Cloud Data Lake: Raw Zone, clean, transform and enrich and store in the Curated Zone.

About the other main component of the DMSA, i.e., the typical DW, it still plays a fundamental role in the enterprise data architecture of any MO organization and it is a key component of the data landscape. However, in a Big Data environment, it often becomes in a **Cloud Data Warehouse**. The reasons are multiple [22]: to reduce cost, increase security, simplify maintenance, and to make possible unlimited and easy DW growth. As mentioned, data is growing in an exponential form, and MO organisations need to adapt using modern Cloud DW architectures. Parallel database systems [23] have been available for organizations for many years and provided enterprise capabilities to run large SQL queries on a large amount of data for many business users in traditional reporting and analytics systems. This new Cloud DW must: maintain the capability to run large SQL queries in a large amount of data, run in the cloud, use MPP (Massive Parallel Processing) [24] with storage and compute decoupled and allow unlimited growth and high SQL query performance.

Other advantages of using the cloud in this layer are that MO companies can program their cloud infrastructure using automatic scaling capabilities. One of the main benefits is cost reduction. **Automatic scaling** allows services to go to sleep during times of low load and can handle unexpected traffic spikes.

As seen in section II, DWs are loaded by ETL processes. Having expanded the storage philosophy with the Data Lakes explained, it is necessary to also extend the philosophy of these processes incorporating **ELT (**Extract, Load, and Transform) processes where data is extracted from the source, then loaded into a landing area in the Data Lake, transforming it where it sits in the Data Lake and then loading it into the Curated Layer of the Data Lake or DW. When the data is extracted from the source into the landing area (Raw Zone), it is a raw copy, meaning the column "names" is kept the same as in the source database and data is not converted. In most cases, in Big Data projects, it is preferable to use ELT instead of ETL. ELT has the benefit of minimizing the processing on the source, since no transformation is being done, which can be extremely important if the source is a production system where you could be impacting the user experience as opposed to a copy of the source (via replication, database snapshot, etc.). The negative side of this approach is that it may take longer to get the data into the target system data lake or data warehouse, also with the landing area there is an extra step in the process, and more disk space will be needed for the landing area.

C. Insight Generation

As described in section II, this layer is supported largely by AI and ML. The new proposed architecture adds two new key elements for MO companies to increase the monetary value of the current data asset. On one hand, some components that will imply a change in the AI model making process conception that will allow increasing their power and the industrialization of them: Cloud AI and AutoML (Automated machine learning). On the other hand, there is a fundamental element that is the incorporation of the new AI models that are revolutionizing the business world, with special mention to Deep Learning algorithms due to the repercussion that it will have in the MO companies. These commented elements are explained below in more detail:

- **Cloud AI.** As seen, AI plays a fundamental role in the Insight Generation layer. Thus, Cloud providers invested a considerable amount of money and time to build research teams to create state-of-art AI services and more specifically Cloud ML services [25]. Some of these important providers are [26]: Google Cloud Machine Learning Platform, Microsoft Cognitive Services, Amazon Machine Learning; and IBM Watson Analytics. Having said that, it is proposed that MO companies select from the available Cloud AI services offers the services that will enable them to differentiate from their competitors instead of building their own (including the typical models presented in Fig. 2: customer segmentation, channel optimization, brand sentiment, etc.). This approach will reduce the cost of hiring highly skilled, expensive and scarcest experts, also reducing the time to market of new insights generation solutions and spending less time/money in AI research. Most of these AI services will not even require minimum AI skills and will be available in the Cloud for software developers to integrate using APIs (Application Programming Interfaces). It should be noted that the suppliers of these new Cloud AI services also have the advantage of being able to train the ML algorithms with much more data that the company itself could have. This implies that, in many cases, they will be much more efficient models. A real example could be for a MO company to use cloud text analytics services to better understand what customers are writing about companies' products and services in the social media or to automatic review customer's comments in a website or in a contact center.
- **Automatic Machine Learning (AutoML)** is a software that can run in the cloud or in a local computer that enables developers with limited ML expertise to train high-quality models. With the shortage of specialized professional [27], the use of AutoML tools [28][29] in MO companies can increase the capacity of the insight generation team.
- **Deep Learning.** They are an evolution of artificial neural networks composed of multiple processing layers to learn representations of data with multiple levels of abstraction [30]. In MO organizations Deep Learning can play an important role in improving existing ML models to forecast time series data (sales), monitor the brand sentiment, understanding customer behavior using text analytics/ NLP (Natural Language Processing), detect customers using face recognition techniques, etc.[31][32].

Most of MO companies in the last decades used the Insight Generation process to increase company value, reduce costs, understand better their customers and competitors, and provide relevant insights for the business. Our purpose in this new framework is to add the **Data Supermarket**, a place to commercialize the data products generated by the insight generation process to other consumers exchanging them for a monetary value [33]. A typical insight generation process is developed by a multidisciplinary data scientist team, where raw data is converted into insights and data products. Our proposed contribution is to define the Data Supermarket as a key element in the Big Data monetisation strategy where data products created by this Big Data Scientist team, will be shipped as data products in a form of services or products to internal or external consumers. Smart Steps from Telefonica is a successful example of building data products and providing them to external companies. Smart Steps is an insight solution that uses anonymized and aggregated mobile network data to provide useful insights [33]. Therefore, the principal goal of the Data Supermarket is to generate new revenue to the company and enable the organization to increase penetration in new markets, not explored before. Organizing the data into one single repository and creating a data product catalog is also a benefit for the end business users. Data will be democratized to the business users with a proper data definition [34].

In Fig. 6 the lifecycle to build data products and sell in the Data Supermarket is showed. As mentioned, the Data Supermarket concept is based on a normal supermarket. In daily life people can buy multiple products in a single place, they are all available in the supermarket. The same concept can be transposed to data. Using this same idea, it is proposed that data will come from different data suppliers (source systems), ingested/moved to the Raw Zone (landing zone) where the raw data is organized in a catalog (raw area). The next step is to transform the raw data into different data products according to the internal and external business user's needs. Like the experience of going to a food supermarket, where all the products are gathered and arranged in an organized manner.

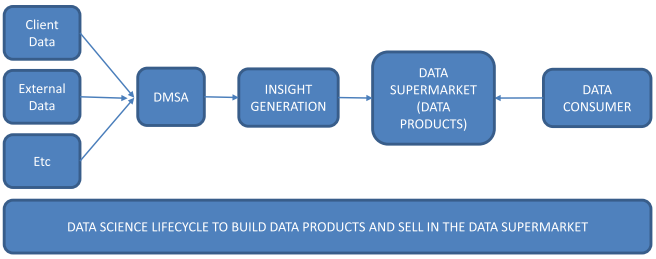


Fig. 6. Data science lifecycle to build data products and sell in the data supermarket.

The benefit of a data supermarket is the synergy of having best practices applied to activities like data collection, data transformation, data storage and data consumption. All the complexity of dealing with data is hidden from the end business user and presented in the Data Supermarket in a better user experience. The presentation or Data Supermarket store will handle the important tasks of data access, data security and data commercialization (free or paid) for internal or external users. The data supermarket store is the broker responsible for selling data products and for allowing market-oriented companies to profit from the data products available in the company data asset. The future data supermarket/data marketplace could become an important key asset of the modern MO organization.

D. Action

The MO scheme requires to be continuously adapting to changes in this market precisely (behavior of customers, competitors, etc.). It is for this reason that this Action layer should facilitate this continuous adaptation. In this layer, some components were added with the aim of automating many of their tasks and therefore drastically reduce the staff that performs them. This staff will be largely replaced by automated agents to perform some specific tasks with more velocity and accuracy. Thus, it is proposed the addition of new components, real-time alerts and chatbots, which are described below:

- **Real-time alerts system.** A MO company needs to be able to notify in real time, to their decision makers, the essential information for decision-making. The real-time notification system will communicate with the decision maker through the best notification option available and desired by the user. The most common

notification mechanisms/systems are SMS, Phone Calls, e-mail, WhatsApp, iOS/Android mobile notification, desktop alert, mass notification, led wall boards, drones, autonomous cars and robots. These systems provide only the essential information that needs to be monitored in real time; and in case that certain rules happen the required actions can be taken consequently.

- **Chatbots.** Answering to clients is another important task that MO companies need to accomplish, and there is a technological capability that can help organizations with this requirement: the chatbots. A chatbot is a software aimed at simulating the conversation of a human being [35][36], i.e. that can interact with humans by text or voice, responding queries using sophisticated natural language processing and speech recognition techniques, they can also retrieve historical information from the DMSA layer and respond personalized questions like what is my current bank balance, what is my national insurance number, how many days do I still have available for annual leave, etc. The insurance company Norwich Union, an Aviva company, is a real example of using chatbots as an automated customer service representative. These virtual agents (chatbots) were designed to help with general queries regarding products [37].
- **DevOps (Development Operations)** [38]-[40]**.** In order to enable MO companies to better serve their customers and compete more effectively in the market, it is proposed the usage of DevOps, which is "*a combination of cultural philosophies, practices, and tools that increase an organization's ability to deliver applications and services at high velocity*"[41]. This component is especially important for the development of technical services of companies since it is a methodology that allows its automation. In this way, the knowledge obtained from the corresponding layer can be implemented more quickly.

E. Outcomes

Outcomes layer includes the continuous motorization of the Key Performance Indicators (KPI) in order to check the effectiveness of the actions. It should be noted that many of these business actions are proposed or executed directly by ML models, so it is necessary to control these automatic decision makers. Thus, the knowledge discovered by these AI techniques is not universal, and its results will be degraded as the behavior of the market changes (consumers, competitors, etc.). Therefore, it is considered essential to incorporate constant monitoring and management of these ML models:

- **ML Model Management** [42]. It is a set of techniques and software that are used to manage ML models. It automatizes the ML lifecycle [43] in MO organizations. It helps organizations to measure the results of the ML models and guarantee the historical of all the ML iterative process. In this way, when it is identified that a ML model is no longer useful, it will be automatically replaced by others.

IV. Concluding Remarks and Future Work

The market is constantly evolving; thus, MO organization has a huge advantage compared to non-market-oriented companies [44]. This difference has been increased in those MO companies that have effectively used the data for decision making, incorporating AI techniques to increase sales, reduce costs, improve customer satisfaction, develop new value propositions, etc.

Because of the emergence of Big Data and AI, every company will have to invest in a new data and AI strategy, learn how to leverage existing internal data, buy new data from other companies and extend new capabilities to become data-driven companies [45]. This journey is mandatory in the digital era that we all live, there is no way a company will survive without shifting to a data culture, the market is evolving rapidly, and new products will disrupt companies and change entire industries.

In the context that there are many disoriented companies on how to adapt their classic conceptual architectures and take advantage of the great potential of this immense volume of data, this work proposes a conceptual Big Data Science Cloud Architecture, that aims to set the foundational building blocks for a modern MO organisation, capable to support Data and AI use cases and provide the flexibility necessary to add or remove capabilities according to market needs.

Companies that adopt the proposed formal framework will bring agility and many possibilities to get new marketing insights and to innovate and create new P/S individually adapted to the needs of each of its customers [46].

As an additional advantage, companies that know how to make their own decisions based on data in the context of Big Data, can consider new lines of business based on the commercialization of data and knowledge. For this reason, it has been incorporated into our formal framework the Data Supermarket component dedicated to selling Data Products that are powerful weapons that will drive innovation and additional revenue for organizations.

As future work, we plan to focus on topics such as: IoT, AutoML, Chatbots, Robotic Process Automation (RPA), Cloud, Data, AI Strategy and Data Supermarket applied to MO companies in different industries and measure the results achieved in short, medium and long term. We also plan to include automatic TD [47]-[49] applied to MO companies in the formal framework proposed.

Acknowledgment

This paper has been elaborated with the financing of FEDER funds in the Spanish National research project TIN2016-75850-R from Spanish Department for Economy and Competitiveness.

References

[1] J. S. Armstrong, "The value of formal planning for strategic decisions: Review of empirical research", *Strategic management journal*, vol. 3, no. 3, pp. 197-211, 1982.

[2] A. K. Kohli, B. J. Jaworski, "Market orientation: the construct, research propositions, and managerial implications", *Journal of marketing*, vol. 54, no. 2, pp. 1-18, 1990.

[3] J. C. Narver, S. F. Slater, "The effect of a market orientation on business profitability", *Journal of Marketing*, vol. 54, no. 4, pp. 20-35, 1990.

[4] S. Negash, P. Gray, "Business intelligence", Handbook on decision support systems, pp. 175-193, Springer, Berlin, Heidelberg, 2008.

[5] DWI: The Data Warehouse Institute, URL (consulted 2019): https://tdwi.org/Home.aspx.

[6] M. D. Stone, N. D. Woodcock, "Interactive, direct and digital marketing: A future that depends on better use of business intelligence", *Journal of research in interactive marketing*, vol. 8, no. 1, pp. 4-17, 2014.

[7] A. G. Picciano, "Big data and learning analytics in blended learning environments: Benefits and concerns". *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 2, no. 7, pp. 35-43, 2014.

[8] R. Kimball, M. Ross, W. Thornthwaite, J. Mundy, B. Becker, "The data warehouse lifecycle toolkit". John Wiley & Sons, 2008.

[9] A. Gandomi, M. Haider, "Beyond the hype: Big data concepts, methods, and analytics", *International Journal of Information Management*, vol. 35, no. 2, pp. 137-144, 2015.

[10] A. Baldominos, F. de Rada, Y. Saez, Yago, "Big Data Analytics Solution for Intelligent Healthcare Management", *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no 7, pp. 13-20, 2018.

[11] D. J. Bodas-Sagi, J. M. Labeaga, "Big Data and Health Economics: Opportunities, Challenges and Risks", *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no 7, pp. 47-52, 2018.

[12] L. Atzori, A. Iera, G. Morabito. "The Internet of Things: A survey", *Computer Networks*, vol. 54, no. 15, pp. 2787-2805, 2010.
[13] P. Fraga-Lamas, T. Fernandez-Carames, L. Castedo, "Towards the Internet of smart trains: A review on industrial IoT-connected railways", *Sensors*, vol. 17, no. 6, pp. 1457, 2017.
[14] B. Nguyen, L. Simkin, L, "The Internet of Things (IoT) and marketing: the state of play, future trends and the implications for marketing", *Journal of marketing management*, vol. 33, pp. 1-6, 2017.
[15] Gartner: Data Management Solutions for Analytics, URL (consulted 2019): https://www.gartner.com/reviews/customers-choice/data-warehouse-solutions
[16] Gartner: Magic Quadrant for Data Management Solutions for Analytics, URL (consulted 2019): https://www.gartner.com/doc/3898487/magic-quadrant-data-management-solutions
[17] P. Mell, T. Grance, "The NIST definition of cloud computing", *NIST Special Publication*, pp. 800-145, 2011.
[18] M. Iorga, A. Karmel, "Managing risk in a cloud ecosystem", *IEEE Cloud Computing*, vol. 2, no. 6, pp. 51-57, 2015.
[19] J. Dixon, "Pentaho Hadoop and Data Lakes." URL (consulted 2019): http://jamesdixon.wordpress.com/2010/10/14/pentaho-hadoop-and-data-lakes
[20] D. E. O'Leary, "Embedding AI and crowdsourcing in the big data lake. *IEEE Intelligent Systems*, vol. 29, no. 5, pp. 70-73, 2014.
[21] P. Pasupuleti, B. S. Purra, "Data Lake development with big data". *Packt Publishing Ltd*, 2015.
[22] A. LaPlante, B. Sharma, "Architecting Data Lakes", O'Reilly Media, 2014.
[23] D. J. DeWitt, J. Gray, "Parallel database systems: The future of high-performance database processing". University of Wisconsin-Madison Department of Computer Sciences Technical Report, 1992.
[24] W. H. Inmon, D. Linstedt. "A Brief History of Big Data in Data Architecture: a Primer for the Data Scientist", Morgan Kaufmann, Boston, 2015.
[25] R. Arulmurugan, K. R. Sabarmathi, H. Anandakumar, "Classification of sentence level sentiment analysis using cloud machine learning techniques", *Cluster Computing*, pp. 1-11, 2017.
[26] A. L'heureux, K. Grolinger, H. F. Elyamany, M. A. Capretz, "Machine learning with big data: Challenges and approaches". *IEEE Access*, vol. 5, pp. 7776-7797, 2017.
[27] T. H. Davenport, D. J. Patil, "Data Scientist", *Harvard business review*, vol. 90, no. 5, pp. 70-76, 2012.
[28] F. Hutter, Balázs Kégl, R. Caruana, I. Guyon, H. Larochelle, et al.. "Automatic Machine Learning (AutoML)". ICML 2015 Workshop on Resource-Efficient Machine Learning, 32nd International Conference on Machine Learning, Lille, France, Jul 2015.
[29] R. S. Olson, J. H. Moore, "TPOT: A Tree-based Pipeline Optimization Tool for Automating Machine Learning". Proceedings of the Workshop on Automatic Machine Learning, in PMLR, pp. 66-74, 2016.
[30] Y. LeCun, Y. Bengio, G. Hinton, "Deep learning", *Nature*, vol. 521 no. 7553, pp. 436-444, 2015.
[31] Z. C. Lipton, "The mythos of model interpretability", *ArXiv e-prints*, pp. 1-3, 2016.
[32] W. Samek, T. Wiegand, K. R. Müller, "Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models", *arXiv preprint*, 2017.
[33] Telefonica. "Smart Steps". URL (consulted 2019): https://www.business-solutions.telefonica.com/en/products/big-data/business-insights/smart-steps/
[34] B. Bengfort, J. Kim, "Data Analytics with Hadoop: An Introduction for Data Scientists." O'Reilly Media, Inc, 2016.
[35] B. A. Shawar, E. Atwell, "A comparison between Alice and Elizabeth chatbot systems". University of Leeds, School of Computing research report, 2002.
[36] J. Weizenbaum, "ELIZA---a computer program for the study of natural language communication between man and machine", *Communications of the ACM*, vol. 9, no. 1, pp. 36-45, 1966.
[37] M. C. Jenkins, R. Churchill, S. Cox, D. Smith, "Analysis of user interaction with service oriented chatbot systems", in International Conference on Human-Computer Interaction, pp. 76-83, 2007.
[38] M. Walls, "Building a DevOps culture", O'Reilly Media, Inc, 2013.
[39] E. Mueller, "What's a DevOp? The Agile Admin". URL (consulted 2019): http://theagileadmin.com/2010/10/08/whats-a-devop/
[40] D. Samovskiy, "The rise of DevOps"; URL (consulted 2019): http://www.somic.org/2010/03/02/the-rise-of-devops/
[41] Amazon AWS. What is DevOps. URL (consulted 2019): https://aws.amazon.com/devops/what-is-devops/
[42] M. Vartak, H. Subramanyam, W. E. Lee, S. Viswanathan, S. Husnoo, S. Madden, M. Zaharia, "Model DB: a system for machine learning model management." In Proceedings of the Workshop on Human-In-the-Loop Data Analytics, pp. 14, ACM, 2016.
[43] M. Zaharia, A. Chen, A. Davidson, A. Ghodsi, S. A. Hong, A. Konwinski, et al, "Accelerating the Machine Learning Lifecycle with MLflow" *Data Engineering*, vol. 39, 2018.
[44] V. Kumar, E. Jones, R. Venkatesan, R. P. Leone, "Is market orientation a source of sustainable competitive advantage or simply the cost of competing?", *Journal of marketing*, vol. 75, no. 1, pp. 16-30, 2011.
[45] M. Ye, G. Li, "Internet big data and capital markets: a literature review", *Financial Innovation*, vol. 3, no.6, pp. 1-18, 2017.
[46] R. A. Carrasco, M. F. Blasco, J. García-Madariaga, E. Herrera-Viedma, "A Fuzzy Linguistic RFM Model Applied to Campaign Management", *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 4, pp. 21-27, 2019.
[47] J.A. Morente-Molinera, G. Kou, K. Samuylov, R. Ureña, E. Herrera-Viedma. "Carrying out consensual Group Decision Making processes under social networks using sentiment analysis over comparative expressions", *Knowledge-Based Systems*, vol. 165, pp. 335-345, 2019.
[48] F.J. Cabrerizo, M.R. Ureña, W. Pedrycz, E. Herrera-Viedma. "Building consensus in group decision making with an allocation of information granularity", *Fuzzy Sets and Systems*, vol. 255, pp.115-127, 2014.
[49] Y. Liu, Y. Dong, H. Liang, F. Chiclana, E. Herrera-Viedma. "Multiple attribute strategic weight manipulation with minimum cost in a group decision making context with interval attribute weights information", *IEEE Trans. Syst. Man and Cyb: Systems*, Accepted, 2019.

Caio Moreno

Caio Moreno is a Big Data Architect, Data Scientist and Data and AI Solution Architect with 18 years of professional experience, working for companies like AOL (America Online), Pan American Health Organization, IT4biz (CEO and Founder), Hitachi Vantara, Pentaho, and many companies and governments in multiple countries and continents. He is currently a part-time Ph.D. Student at Complutense University of Madrid and Advanced Analytics Manager at Avanade London Office (An Accenture and Microsoft owned company). He holds a master's degree in Business Intelligence and Data Mining at the Complutense University of Madrid, a specialization in Data Science at University Camilo José Cela (UCJC)/U-TAD, an MBA at Getulio Vargas Foundation (FGV-SP) and a Specialization in Software Development/Java at The Federal University of Technology – Paraná (UTFPR). His main field of research is focused on Data Science, Big Data, IoT, Cloud and AutoML (Automated Machine Learning) applied to Marketing and in the financial sector. He has been a professor at The State University of Western Paraná (UNIOESTE) – Brazil (2010-2012) and EOI (Escuela de Organización Industrial) – Spain (2016-2017). Also, a public speaker in many international events in Brazil, Europe, Africa and USA. Fluent in English, Spanish and Portuguese. He is married and has 2 daughters.

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 professor at the University of Granada, Granada, Spain (2004-2013) and currently at the Complutense University, Madrid, Spain (from 2013).

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 modeling, and aggregation of information, information retrieval, bibliometric, digital libraries, web quality evaluation, recommender systems, and social media. In these topics he has published more than 250 papers in ISI journals and coordinated more than 22 research projects. Dr. Herrera-Viedma is Vice-President of Publications of the IEEE SMC Society and an Associate Editor of 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.

An Improved Artificial Intelligence Based on Gray Wolf Optimization and Cultural Algorithm to Predict Demand for Dairy Products: A Case Study

Alireza Goli[1], Hassan Khademi Zare[1]* , Reza Tavakkoli-Moghaddam[2], Ahmad Sadeghieh[1]

[1] Department of Industrial Engineering, Yazd University, Saffayieh, Yazd (Iran)
[2] School of Industrial Engineering, College of Engineering, University of Tehran, Tehran (Iran)

Received 20 December 2018 | Accepted 12 January 2019 | Published 8 March 2019

Abstract

This paper provides an integrated framework based on statistical tests, time series neural network and improved multi-layer perceptron neural network (MLP) with novel meta-heuristic algorithms in order to obtain best prediction of dairy product demand (DPD) in Iran. At first, a series of economic and social indicators that seemed to be effective in the demand for dairy products is identified. Then, the ineffective indices are eliminated by using Pearson correlation coefficient, and statistically significant variables are determined. Then, MLP is improved with the help of novel meta-heuristic algorithms such as gray wolf optimization and cultural algorithm. The designed hybrid method is used to predict the DPD in Iran by using data from 2013 to 2017. The results show that the MLP offers 71.9% of the coefficient of determination, which is better compared to the other two methods if no improvement is achieved.

Keywords

Artificial Intelligence, Gray Wolf Optimization, Cultural Algorithm, Regression, Time-series Based Neural Network.

DOI: 10.9781/ijimai.2019.03.003

I. Introduction

Demand prediction includes determining the volume of sales or market trends for a particular product and determining its sales in the entire market. By forecasting demand in a company, it is possible to determine the company's market share and the required planning for performance improvement. Therefore, one of the earliest stages of budget planning is demand prediction. With demand prediction, the income of the firm and its costs are based on the predicted income. Therefore, if sales forecasting is not highly accurate, it will affect other monetary and financial variables [1].

Since prediction is one of the key tools in manufacturing system planning, various researchers have well attempted to develop this area. In many ways, demand prediction is focused on behavioral patterns of demand in the past. In these methods, the use of historical data is important [2].

Various prediction models were used in the literature, such as linear or polynomial regression techniques, moving average, box Jenkins models and structural models. However, these models had weaknesses that did not allow to consider complex and non-linear affecting factors. In recent decades, artificial intelligence has been used to detect relationships between complex and nonlinear variables [3]. Accordingly, due to the high ability of neural networks to learn complex and non-linear relationships, they have been widely used in various fields of science [4].

* Corresponding author.
E-mail address: Hkhademiz@yazd.ac.ir

Hybrid artificial intelligence (HAI) tools are used to establish a better relationship between the input and output values of the training data. Recent studies have shown that new researches in the field of neural networks are focused on using HAI [5]. The growth of HAIs in different domains indicates their ability to achieve better results compare to traditional methods. Several main related research in this filed are explained below.

II. Related Works

Khalafallah (2008) [6] has used artificial neural network to predict US housing demand. Khashei and Bijari (2011) [7] have developed a hybrid neural network with ARIMA technique to improve the performance of the neural network. Slimani et al. (2015) [8] have examined Moroccan supermarket demand prediction using neural networks. They have used a MLP neural network.

Liu et al. (2017) [9] in their study, developed the artificial neural network in order to predict the periodic demand for products. Rangel et al. (2017) [10] have predicted demand for drinking water for the next 24 hours in the city of Barcelona through a variety of neural networks. In order to improve these neural networks, the training phase has been improved with the help of genetic algorithm. Evaluations carried out by the error index indicate that the use of the genetic algorithm improves prediction performance up to 1.5%. In 2018, Anand and suganthi [11] proposed an ANN that is optimized with hybrid genetic algorithm and particle swarm optimization (GA-PSO) to predict the electricity demand of the state of Tamil Nadu in India. The obtained results showed that the proposed method have higher accuracy than single optimization models such as PSO and GA.

Obedinia et al. [12] in 2018 prepared a hybrid forecasting engine based on ANN and shark smell optimization (SSO) algorithms. Dosdoğru et al. [13] in 2018 used different meta-heuristics likes Ant Lion Optimization (ALO) and Firefly Algorithm [14] to improve the performance of the ANN. According to the results, hybrid ANN models provided a remarkable solution for other forecasting problems.

This research is focused on the prediction of dairy product demand (DPD) by using artificial intelligence. In this regard, MLP, is used as demand prediction tools. Since this tool has limitations and defects, the most recent meta-heuristic algorithms have been used in order to improve such limitations.

The main contribution of this research is to provide an integrated framework based on statistical tests, time series neural network and HAIs with the help of the most recent meta-heuristic algorithms. The used algorithms include gray wolf optimization (GWO) and cultural algorithm (CA). The work is organized as follows: Section III introduces the methodology, section IV presents the structure of hybrid methods. In Section V, the case study is introduced, and in Section VI, the results of implementation of the hybrid prediction methods are presented; Section VII concludes the paper with some suggestion for future researches.

III. Methodology

In MLP, the total weighted input and bias are given to the neural network through the transfer function in order to generate outputs. The results are saved in a layered feed forward topology called the feed forward neural network [3]. MLP networks include an input layer, one or more hidden layers, and an output layer. Each layer has a number of processing units, and each unit is fully linked with weighted connections to units in the next layer. The N input converts to the L output through nonlinear functions. The units of the neural network are calculated by Eq. (1).

$$x_0 = f\left(\sum_h x_h w_{h0}\right) \tag{1}$$

In Eq. (1), f is the activation function, x_h is the activation of the h_{th} hidden layer node and w_{ho} is the interconnection between the h_{th} hidden layer node and the o_{th} output layer node. The most common activation function that is used in the literature is the sigmoid function that is based on Eq. (2).

$$x_0 = \frac{1}{1+\exp\left(-\sum_h x_h w_{h0}\right)} \tag{2}$$

The quality of the output of the MLP is specified by comparing the predicted output (y) and real output (o). The main three criteria of evaluation prediction quality is presented in Eqs. (3-5).

$$MSE = \frac{1}{n}\sum_i^n \left(y_i - o_i\right)^2 \tag{3}$$

$$R^2 = 1 - \frac{\sum_i \left(y_i - o_i\right)^2}{\sum_i \left(y_i - \bar{y}\right)^2} \tag{4}$$

$$MAE = \frac{1}{n}\sum_i^n \left|y_i - o_i\right| \tag{5}$$

The aim of the MLP is to reduce the MSE by adjusting the connections between the layers. The weights are adjusted using the gradient descent back propagation (BP) algorithm. This algorithm requires training data. During the training process, the MLP starts with a random set of initial weights and then continues a convergence to least error (optimum or local) is met [15].

The BP algorithm has a deep independency to the number of hidden layers. Moreover, this algorithm is so slow converging in big data analysis. For these reasons, it is suggested to use a meta-heuristic algorithm instead of the classic learning method, which has a fast convergence and low independency to the number of hidden layers. The method of implementing a meta-heuristic algorithm in the neural networks is explained in the section VI and the structure of proposed hybrid MLP and novel meta-heuristic algorithms is described in Fig. 1.

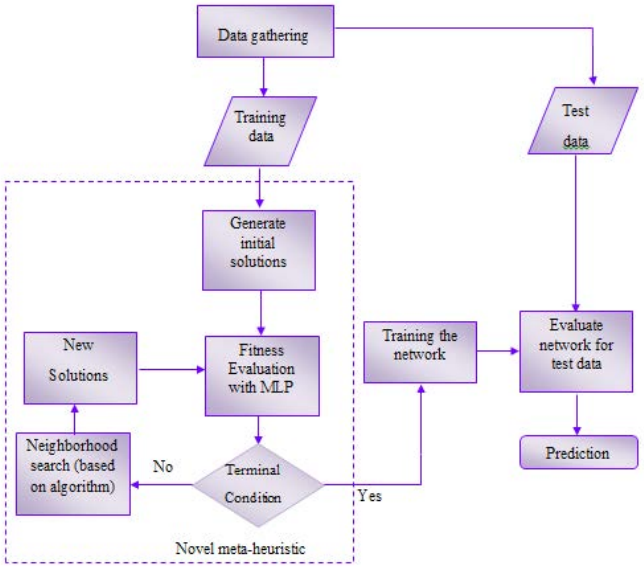


Fig. 1. The structure of the proposed hybrid MLP and novel meta-heuristic algorithms.

A. Prediction Network Optimization

As previously described, the use of traditional methods to optimize the prediction network structure has disadvantages that need to be addressed by a more powerful method. In this regard, it is suggested to use new meta-heuristic algorithms.

In order to implement each prediction network optimization algorithm, a suitable solution representation and a fitness function are required. Each solution representation shows a set of weights corresponding to the neural network architecture. Fig. 2 shows an example solution string for a network architecture with four neurons in the input layer, two neurons in the middle layer and one neuron in the output layer. In this representation, there are 8 (4×2) weights linking the first and second layers (W1) and 2 weights linking the second and third layers (W2). The values of all these weights range from zero to one.

W_1^1	W_2^1	W_3^1	W_4^1	W_5^1	W_6^1	W_7^1	W_8^1	W_1^2	W_2^2

Fig. 2. The structure of the solution string for network optimization algorithms.

The next step after determining the network weights is to obtain a set of predictions (Y), compare them with their counterparts among real observations (O), and measure the difference in terms of the mean squared error [16], which is presented in Eq. (3).

In this research, the GWO and CA are used to optimize the prediction network. In all of these algorithms, the solution representation and fitness function are the same. In section IV, the details of each of these algorithms are expressed.

IV. Applied Novel Meta-Heuristic Algorithm

A. Gray Wolf Optimization

The gray wolf optimization (GWO) algorithm was proposed by Mirjalili et al. (2014) based on their collective hunting [17]. The gray wolf is from the Canadian wolf family. Gray wolves are on top of the food chain and prefer to live in groups. Their groups consist of 5-12 wolves. The wolf alpha is the leader wolf in the group, whose orders must be followed by the group. Alphas are basically responsible for deciding about hunting, sleeping, moving time, etc. The second level is the gray wolf beta. Betas are wolves led by alphas, which help alphas in decision making and other group activities. Beta wolves are probably the best candidate for being alphas and play the role of an assistant for the alpha and a moderator for the group. The lowest grade is the omega gray wolf. Omega wolves have a sacrificial role for other members of the group. They are the last wolves that are permitted to eat. A wolf that is not an alpha, beta or omega is called obedient (or delta). Delta wolves follow alphas or betas and lead omegas. Scouts and sentinels belong to this group.

1. Social Behavior of Wolves

When designing the GWO in social behavior of wolves, the most appropriate solution is called wolf alpha. Then, the second and third best solutions are called wolf beta and gamma respectively. The remaining solutions are assumed to be omega. Therefore, optimization in the GWO algorithm is led by alpha, beta and delta, and omega wolves follow these three categories.

2. Turning Around the Prey

Gray wolves encircle the prey during hunt. For modeling the turning behavior mathematically, Eqs. (6-7) are presented [17].

$$\vec{D} = \left| \vec{C} \,.\, \overrightarrow{X_P}(t) - \vec{X}(t) \right| \tag{6}$$

$$\vec{X}(t+1) = \overrightarrow{X_P}(t) - \vec{A}.\vec{D} \tag{7}$$

Where t represents the repetition, A and C are the vector coefficients, $\vec{X}$ is the vector of the prey position and X represents the position vector of a gray wolf. The vectors A and C are calculated according to Eqs. (8-9).

$$\vec{A} = 2\vec{a}.\vec{r_1} - \vec{a} \tag{8}$$

$$\vec{C} = 2.\vec{r_2} \tag{9}$$

Where $\vec{a}$ decreases linearly from 2 to 0 under the path of repetitions, and r_1 and r_2 are random vectors.

3. Detecting the Position of the Prey

In order to mathematically simulate the hunting behavior of gray wolves, we assume that alphas (the best candidate solution), betas and deltas have enough knowledge about the potential position of the prey. Therefore, we will save the first three best solutions and force other search agents (omegas) to update their position according to the position of the best search agents based on Eqs. (10-16).

$$\overrightarrow{D_\alpha} = \left| \overrightarrow{C_1}.\overrightarrow{X_\alpha} - \vec{X} \right|, \tag{10}$$

$$\overrightarrow{D_\beta} = \left| \overrightarrow{C_2}.\overrightarrow{X_\beta} - \vec{X} \right|, \tag{11}$$

$$\overrightarrow{D_\delta} = \left| \overrightarrow{C_3}.\overrightarrow{X_\delta} - \vec{X} \right|, \tag{12}$$

$$\vec{X}_1 = \overrightarrow{X_\alpha} - \overrightarrow{A_1}.\left(\vec{D}_\alpha\right), \tag{13}$$

$$\vec{X}_2 = \overrightarrow{X_\beta} - \overrightarrow{A_1}.\left(\vec{D}_\beta\right), \tag{14}$$

$$\vec{X}_3 = \overrightarrow{X_\delta} - \overrightarrow{A_1}.\left(\vec{D}_\delta\right), \tag{15}$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \tag{16}$$

The final position can be in a random location within a circle marked with the alpha, beta, and delta positions in the search space. In other words, alphas, betas and deltas estimate the position of the prey and other wolves update their position randomly around the prey. In summary, the steps of the GWO algorithm are as the algorithm 1.

1: Generate initial search agents Gi (i=1, 2,…., n)
2: Initialize the vector's a, A and C
3: Estimate the fitness value of each hunt agent
  Ga=the best hunt agent
  Gβ=the second best hunt agent
  Gδ=the third best hunt agent
4: Iter=1
5: repeat
6: for i=1: Gs (grey wolf pack size)
 Renew the location of the current hunt agent
End for
7: Estimate the fitness value of all hunt agents
8: Update the value of Ga, Gβ, Gδ
9: Update the vectors a, A and C
10: Iter=Iter+1
11: until Iter>= maximum number of iterations {Stopping criteria}
12: output Ga
13. End

Algorithm 1. The GWO pseudocode [17].

B. Cultural Algorithm (CA)

Cultural algorithms were introduced by Reynolds in 1994 [18]. These algorithms are inspired by evolution of the human culture and influence of a community on its people and its effect on creation of future generations. A cultural algorithm considers the two search spaces in the optimization process. In other words, cultural algorithms are a branch of evolutionary computing, in which there is a component called the "belief space" in addition to the population space, which is based on the Darwinian genetic theory. This algorithm uses the domain of knowledge for the search process. In fact, in the belief space, a group of elite people (or those who are more attentive) are selected, and in population space, there will be a way that is considered for influencing this selected group on the individuals. The addition of knowledge is effective in improving the efficiency of evolutionary algorithms and makes the search process more intelligent. Cultural algorithms from this point can be viewed as an extension of genetic algorithms. This algorithm has different knowledge in its belief space, which helps the search process. The components of the cultural algorithm include: a) the population space, b) the belief space, c) the admission function, and d) the effect function, which are explained below.

1. Population Space

The population is a cultural algorithm is almost same to this in the genetic algorithm and is generated randomly.

2. Belief Space

The belief space of a cultural algorithm is divided into different parts. These parts represent the different areas of knowledge possessed by the population from the search space. In the belief space, successful people's generalized experiences are gained from the demographic space and improved the generations. Thus, in fact, people's cultural information is modeled in the belief space. These experiences will transfer and affect all future generations. This space is actually effective in pruning the population. Every person is a particle in the search space, which is used by the belief space to move away people from undesirable areas and push them toward promising and near-solution areas. Various knowledge is formed to shape the belief space, including situational knowledge, normative knowledge (criterion), topographic knowledge (pivotal component), history knowledge, and range knowledge; the belief space and temporary knowledge are calculated by Eq. (17) [19].

$$y(t)=\{S(t),N(t)\} \tag{17}$$

Where y is the space of belief, S(t) is positional knowledge and N(t) is normative knowledge.

3. Positional or Situational Knowledge

We consider the best solutions of each generation (or current generation) as a conditional particle. A solution in the CA is considered as a target. Situational knowledge here includes the best component and according to Eq. (18), belief (culture) atmosphere will be updated.

$$y(t+1)=\begin{cases}\min\limits_{x_i(t)} f(x_i(t)) & \min\limits_{x_i(t)} f(x_i(t)) \le f(y(t)) \\ y(t) & otherwise\end{cases} \tag{18}$$

4. Normative Knowledge (Criterion)

This knowledge source stores a set of good and promising ranges that are extracted from a set of fine particles for each dimension. If the space is competitive, norms take a proper shape and [20]. This knowledge is applied according to the following equation. Here, n represents the number of dimensions of the problem, and each X is defined Eq. (19-20) [21].

$$N=(X_1,X_2,X_3,\ldots,X_n\} \tag{19}$$

$$X_i=\left[l_i \quad u_i \quad L_i \quad U_i\right] \tag{20}$$

Where l_i and u_i are the upper and lower limit of the ith dimension, respectively, and L_i and U_i are the values of the competency function in that range. According to this knowledge, the search space is gradually becoming smaller and closer to good regions.

5. Admission Function

A variety of static and dynamic admission functions can be used in this algorithm. In this paper, the dynamical admission function is used to vary people's effect on culture in each repetition, and the number of population members is derived from Eq. (21) [21].

$$n=\left[\frac{n_s\times\gamma}{t}\right] \tag{21}$$

Where n_s is the population size, γ is a random number between 0 and 1, and t is the number of the algorithm iteration.

6. Effect Function

The belief atmosphere affects the population space by using the mutation operator. This effect is possible in two ways: measure of mutation and direction of mutation. The effect function of Eqs. (22 - 23) is used to change the responses [18].

$$x_{ij}^{'}=\begin{cases}x_{ij}(t)+\sigma_{ij}\left|N(0,1)\right| & x_{ij}(t)<y_j(t)\\ x_{ij}(t)-\sigma_{ij}\left|N(0,1)\right| & x_{ij}(t)>y_j(t)\\ x_{ij}(t)+\sigma_{ij}N(0,1) & otherwise\end{cases} \tag{22}$$

$$\sigma_{ij}(t)=\alpha\left[x_j^{max}(t)-x_j^{min}(t)\right] \tag{23}$$

In summary, the benefit of the cultural algorithm can be shown as the algorithm 2.

1. Begin
2. t = 0;
3. Initialize Population POP (t);
4. Initialize Belief Space BLF (t);
5. Repeat
6. Evaluate Population POP (t);
7. Adjust (BLF (t), Accept (POP (t)));
8. Adjust (BLF (t));
9. Variation (POP (t) from POP (t-1));
10. until termination condition achieved
11. End

Algorithm 2. The CA pseudocode [21]

V. Case Study Description

To study the trend of demand for dairy products in Iran [22], the database containing the monthly sales of dairy products of Pegah Golpayegan Dairy Company during a period of 60 months from 2013 to 2018 was collected. The trend of demand during these 60 months is displayed in Fig. 3. As shown in this figure, the demand for dairy products does not follow any specific pattern. Thus, to make a prediction about the demand in the coming months, it is necessary to identify a set of determinant factors and then apply an artificial intelligence tool on the collected data. After the examination of the literature, the authors identified 12 variables that can affect the demand for dairy products. These variables are listed in Table I.

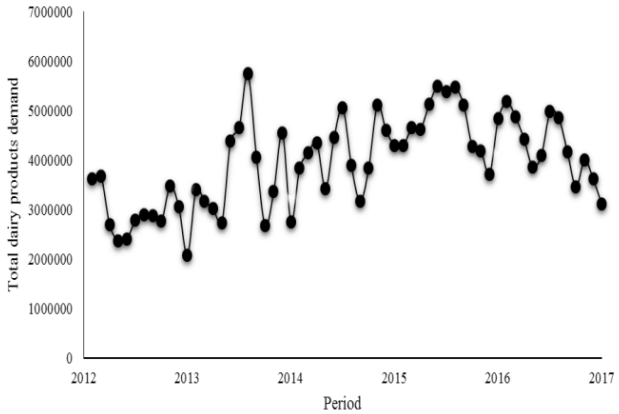


Fig. 3. Total demand for dairy products from 2013 to 2018.

TABLE I. List of Variables on Dairy Products Demand

Variable	Symbol	Period	Description
Inflation	IN	Monthly	General price increase rates
Consumer price index	CPI	Monthly	Average price levels of products portfolio purchased by consumers
Producer Price Index	PPI	Monthly	The average price of goods and services that producers pay for the production
The dairy products price index in urban areas	DPI	Monthly	Measuring the changes in the price of dairy products that are consumed by households in urban areas
Gini coefficient	GC	Monthly	Measuring the distribution of income within the country over a period of time
Average milk price	AMP	Yearly	Milk is the primary ingredient in the production of all dairy products and its price influences the price of other dairy products
Population	POP	Annually	The total population of the Iran which is reported annually
Average number of households	ANH	Annually	Measuring the distribution of the individuals among the households
Average annual gross income	AAGI	Annually	The average income which the fixed costs are not reduced it.
Annual cost percentage for dairy products	ACPD	Annually	The percentage of households' average income that pays for dairy products
Legatum prosperity index	IPI	Annually	This index determines the welfare level based on the economic situation (such as GDP), health, education, etc.
Industrial production index	IPI	Annually	Indicates the change in the number of goods and services produced by the enterprises.

VI. Numerical Results

First, it is necessary to analyze the set of indexes that could affect the demand for dairy products. In this regard, it is necessary to measure the correlation of each of the indexes with the DPD. Therefore, the Pearson correlation coefficient was used. This coefficient is always between -1 and 1. If the Pearson correlation coefficient of X and Y is close to 1 or -1, respectively, it means that the relationship between X and Y is significant, and as close as 0, it means the insignificance of the relationship between these two variables. The Pearson correlation coefficient between each of the 12 indexes and the DPD was obtained using SPSS 16 software and is shown in Table II.

TABLE II. The Pearson Correlation Test for the Identified Variables

Parameter	Pearson correlation	Significance	Covariance
IN	-0.685	0.00	-6.79E6
GC	0.594	0.00	6.2E3
CPI	0.548	0.00	6.97E6
PPI	0.490	0.00	6.44E8
DPI	0.040	0.765	1.1E7
AMP	0.466	0.00	9.9E5
POP	0.552	0.00	6.6E5
ANH	-0.595	0.00	-3.5E5
AAGI	-0.006	0.962	-7.8E4
ACPD	-0.607	0.00	1.1E5
LPI	-0.465	0.00	-2.7E5
IPI	0.138	0.293	1.8E7

In Table III, at 95% confidence level, the variables with a significance value of more than 0.05 were statistically insignificant on the DPD and their trend was not correlated with the demand for dairy products in the last five years. Accordingly, the dairy products price index in urban areas (DPI) and average annual gross income (AAGI) and industrial production index (IPI) did not have a significant correlation with the demand for dairy products.

TABLE III. Regression Analysis for the Statistically Significant Variables

Variable	β	Std. error	T value	P value
Constant	1.235E8	2.609E8	1.473	0.00638
CPI	-5.504E4	7.518E4	-2.732	0.00467
IN	-7.396E4	3.812E4	-2.640	0.00058
PPI	7.2923E3	3.134E4	1.233	0.00817
AMP	-5.226E5	3.499E5	-3.494	0.00141
POP	8.857E5	5.318E5	3.665	0.00102
ANH	-2.351E7	2.101E7	-2.119	0.00268
LPI	-4.047E5	1.189E6	-1.341	0.00735
GC	-1.164E8	1.741E8	-2.668	0.00507

The results of Table II show that the highest correlation was between inflation and the DPD. Inflation has led to a sharp drop in dairy consumption in Iran. This is also well illustrated in the ACPD index. Despite increasing the annual cost percentage for dairy products in some periods, the consumption of dairy products dropped in the same periods due to increasing inflation in the price of dairy products. On the other hand, the DPD correlation with these variables has its own complexity as the indexes affect each other. For example, the population of the country has increased over the past five years, which has led to a higher demand for dairy products, resulting in a positive correlation between the DPD and the population in Iran. For this reason, and despite the interaction of different indexes on each other, it is necessary to use an intelligent tool to predict the demand for dairy products.

The data between the years 2013 and 2017 was used to construct a multi variable regression (MVR) as shown in Eq. (24), where total dairy demand is the dependent variable while the independent variables are categorized in Table III as statistically significant variables.

$$\begin{aligned}\hat{Y}_t = {} & 1.23\text{E}8 - 5.50\text{E}4\ CPI_t - 7.39\text{E}4\ IN_t + 7.29\text{E}3\ PPI_t \\ & - 5.22\text{E}5\ AMP_t + 8.857\text{E}5\ POP_t \\ & - 2.351\text{E}7\ ANH_t - 4.047\text{E}5\ LPI_t \\ & - 1.164\text{E}8\ GC_t \end{aligned} \quad (24)$$

The MVR equation is given in Eq. (24) while the regression coefficient (β) of each variable, standard errors, T values and two tailed p values of the coefficients were obtained from SPSS MVR analysis and are summarized in Table IV. It should be noted that the major impact on the demand may vary from country to country; in other words, a variable in correlation with the total dairy demand in a country may have no significant correlation in another country. Therefore, it is recommended to determine important statistical variables that affect the total dairy demand of a particular country before performing future predictions.

After determining the statistically significant variables correlating with the DPD, these variables were considered as inputs of the prediction network; then, the demand for dairy products was considered as the output of the prediction network. Fig. 4 shows the scheme of the proposed network prediction.

In this research, two kinds of artificial neural network models including ANN1 and ANN2 were used. For the ANN1 (multilayer perceptron), the total dairy demand was modeled as a function of the statistically significant variables, as shown in Fig. 4. In order to

evaluate the ANN1 for the future months, the future values of the statistically significant variables were determined by using the ANN2, which is a time series based neural network. In this kind of ANN, one variable in the month "t" was modeled as a function of its values in the past years [15]. The concept of the ANN2 is shown in Fig. 5.

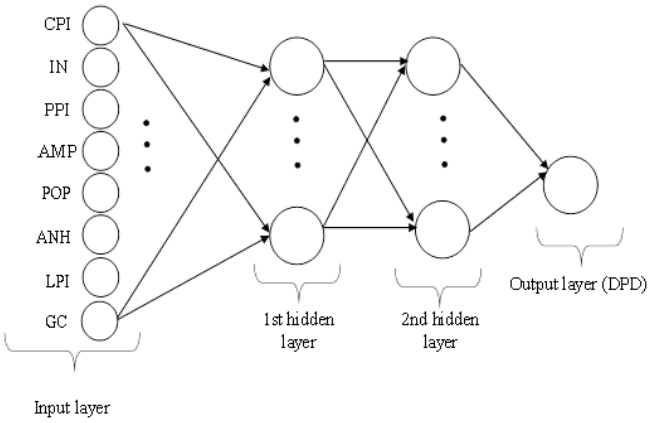


Fig. 4. The proposed prediction network.

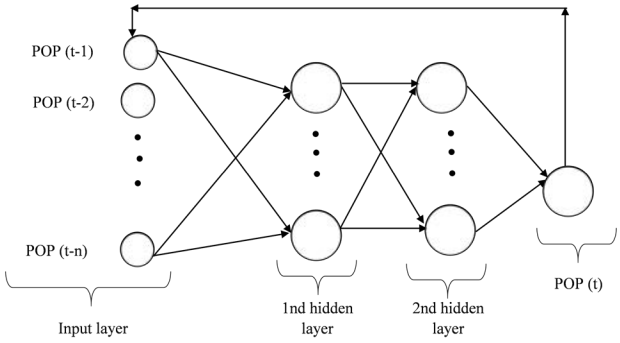


Fig. 5. The proposed time series based neural network.

A. Significant Variables Prediction

In order to predict the statistically significant variables determined as the input variables of the prediction network, an MLP network was considered for each of the variables. The structure of this network is in accordance with Fig. 5. Moreover, the value of n is fixed to 5. In other words, the value of each variable per month is a function of its own value in the last 5 months. Eqs. (25-32) show the function of each variable.

$$CPI_t = f\left(CPI_{t-1}, CPI_{t-2}, CPI_{t-3}, CPI_{t-4}, CPI_{t-5}\right) \quad (25)$$

$$IN_t = f\left(IN_{t-1}, IN_{t-2}, IN_{t-3}, IN_{t-4}, IN_{t-5}\right) \quad (26)$$

$$PPI_t = f\left(PPI_{t-1}, PPI_{t-2}, PPI_{t-3}, PPI_{t-4}, PPI_{t-5}\right) \quad (27)$$

$$AMP_t = f\left(AMP_{t-1}, AMP_{t-2}, AMP_{t-3}, AMP_{t-4}, AMP_{t-5}\right) \quad (28)$$

$$POP_t = f\left(POP_{t-1}, POP_{t-2}, POP_{t-3}, POP_{t-4}, POP_{t-5}\right) \quad (29)$$

$$ANH_t = f\left(ANH_{t-1}, ANH_{t-2}, ANH_{t-3}, ANH_{t-4}, ANH_{t-5}\right) \quad (30)$$

$$LPI_t = f\left(LPI_{t-1}, LPI_{t-2}, CPI_{t-3}, LPI_{t-4}, LPI_{t-5}\right) \quad (31)$$

$$GC_t = f\left(GC_{t-1}, GC_{t-2}, GC_{t-3}, GC_{t-4}, GC_{t-5}\right) \quad (32)$$

After trying different neural network topologies, the optimal neural network topology was found as 5-5-5-1 (5 inputs, 5 neurons in the first and second hidden layers and 1 output) for ANN2 and ANN1; with the activation functions of sigmoid function for the input layer.

In the ANN2 neural network designed for each variable, the data for the last 12 months (2017) were considered as test data while the rest were considered as training data. Then, the amount of each variable was foreseen for the following 12 months (2018). The prediction process was initially calculated using the available data (the target data) for the forecasting each month between 2014 and 2017. For these predictions, the coefficient of determination (R2) and mean square error [16] were calculated and are presented in Table IV.

TABLE IV. The Prediction Results of the Significant Variables

	Train		Test		Total	
Variable	R^2	MSE	R^2	MSE	R^2	MSE
CPI	0.909	0.236	0.932	0.324	0.901	0.272
IN	0.924	0.314	0.920	0.317	0.922	0.315
PPI	0.953	0.207	0.946	0.216	0.957	0.213
GC	0.841	0.407	0.834	0.419	0.839	0.409
AMP	0.881	0.501	0.796	0.561	0.855	0.529
POP	0.903	0.367	0.884	0.433	0.896	0.401
ANH	0.916	0.298	0.909	0.371	0.912	0.318
LPI	0.871	0.513	0.866	0.524	0.869	0.519

Table IV shows the results of forecasting each of the variables affecting the DPD for 2018 in three categories. The first category refers to the training data, the second part refers to the test data, and the last part refers to the total data for five years. The results of Table IV show that the MLP was able to create an approximately 90% adaptation between Targets and Predictions. This indicates the performance of the MLP in the time series prediction with irregular structure. On the other hand, the MSE error rate in the training data in all the variables was less than that in the test data. The reason for this is the sudden change in any of the variables in the last year. This is a clear indication of the variables reported annually (the AMP, POP, ANH, LPI, GC). The forecast error rate was higher for the variables reported annually than for those reported on a monthly basis (the CPI, IN, PPI). The data reported on an annual basis, the information of the last 5 months (the last 5 months) is equal to each other; thus, the forecast for the following period would be very close to that for the last five months. This is while the amount of these types of variables changed in the following years. Therefore, the forecast error would be higher for the annually reported variables. In Fig. 6 and 7, a comparison was made between the prediction of PPIs reported on a monthly basis and the AMP reported annually.

The predicted value of each significant variable will be used as inputs of the prediction network to estimate the DPD in 2018. For example, in Fig. 6 and 7, the predicted value is shown for each of the months of 2018 with respect to the PPI and AMP indexes.

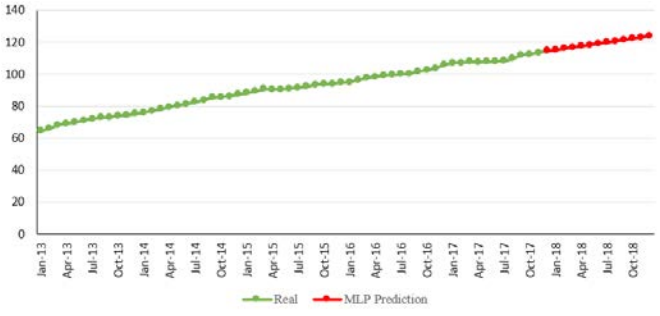


Fig. 6. The PPI prediction for 2018.

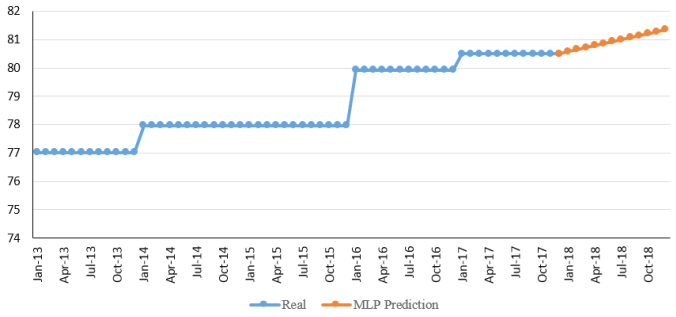


Fig. 7. The AMP prediction for 2018.

As shown in Fig. 6, since the PPI data were reported on a monthly basis, the MLP neural network well identified the time series trend of this variable. For this reason, according to Table IV, the MSE was 0.231 in predicting this variable. In Fig. 7, the forecast is shown for the AMP variable, which was reported annually. The value of this variable was quite stable over a year but jumped at the beginning of the following year. This mutation was not precisely predictable by the MLP, which is why the MSE value for the AMP was 0.529. Despite this, the MLP's neural network was able to match 85% of the actual data and predictions.

B. DDP Prediction Using Hybrid MLP

In order to improve the performance of the MLP neural network, we optimized this tool with various meta-heuristic algorithms such as the GWO and CA. As described in Section IV, each of these algorithms was used in the training section and attempted to determine the bias weights in order to minimize the MSE value. Since each algorithm has its own specific method for optimizing the MSE, the results of the prediction of the combined neural networks with each of the algorithms were different. In Table V, the values of the R^2, MSE and MAE are presented for each MLP based prediction.

TABLE V. The Prediction Results of the MLP-based Methods

	MAE	MSE	R2
MLP-GWO	**0.0941**	**0.0127**	**0.9476**
MLP -CA	0.1681	0.0337	0.8520
Pure MLP	0.3147	0.0679	0.7191

In Table V, the pure MLP was a multi-layer perceptron neural network, in which the BP was used to train the network. In this kind of neural network, the coefficient of determination was obtained as 71.9%. Moreover, the MSE index was 0.067. The use of the meta-heuristic algorithms to improve the MLP increased the coefficient of determination up to 94.7%. Comparison of the proposed algorithms for improvement of the MLP-GWO yielded the best results, and then the CA algorithm performed better than other algorithms.

The dispersion of the best fitted method (MLP-GWO) outputs to the target values is shown in Fig. 8. Moreover, Fig. 9 represents the MLP- GWO predictions with the trend of the DDP.

In Fig. 8, the horizontal axis represented the normalized output of the MLP- GWO, and the vertical axis represented the normalized target values. Moreover, the Y = X line indicated the ideal prediction. In other words, the distance from the points to this line represented the prediction error of each available data. In Fig. 9, the trends of the DDP over the last five years were dictated and projected by the MLP- GWO. Moreover, Fig. 9 shows that the MLP was able to identify and predict the overall trend of the DDP by using its input variables. The biggest error in the forecast was for Jan-2016. In this period, there was a great change in the demand. By analyzing Fig. 7, it is seen that the average price of milk in this period increased sharply. For this reason, the MLP predicted an increase in the demand by analyzing the input variables; however, this prediction was accompanied by some error.

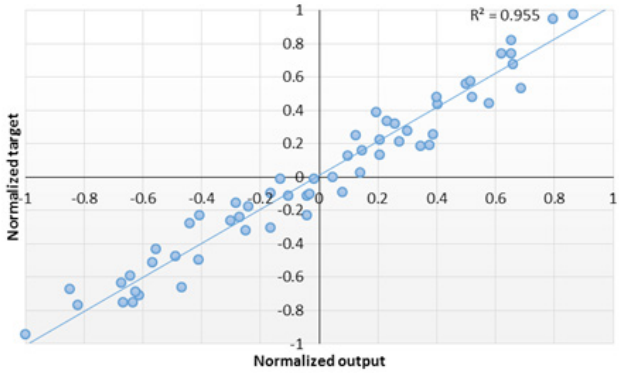


Fig. 8. Compliance between the normalized MLP- GWO output and targets.

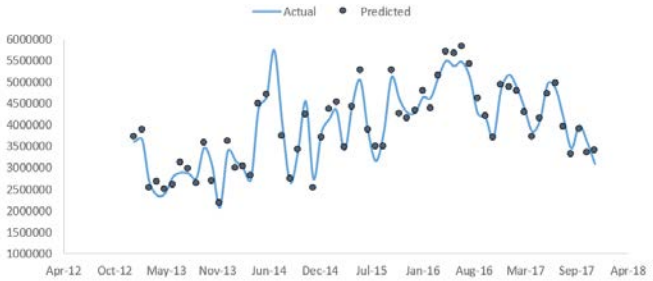


Fig. 9. The actual demand and best fitted prediction based on the MLP.

VII. Conclusion and Future Researches

The key innovation of this research was to improve artificial intelligence with the help of the newest meta-heuristic algorithms such as the gray wolf algorithm and cultural algorithm. Artificial intelligence was used to provide a comprehensive framework to predict the demand for dairy products. In this framework, a series of socio-economic variables that seemed to be effective on the DDP were first identified. Then, using the Pearson correlation coefficient, ineffective variables were eliminated. The results showed that the variables such as the index of dairy prices in urban areas (DPI), annual average gross income (AAGI) and industrial production index (IPI) did not have a significant correlation with the DDP. Then, the statistically significant variables were predicted using a time-based neural network. In the next step, the combination of each of the GWO and CA with MLP was applied. The summary of the results showed that the forecast error decreased by 1.8 times in the hybrid method based on the MLP. In order to improve this research, it is suggested to use hybrid algorithms such as hybrid GWO-PSO to improve artificial intelligence predictions. It is also recommended to evaluate the ability of these new algorithms and compare it with that of the pure GWO and PSO algorithms.

References

[1] A. Kochak, S. Sharma, Demand forecasting using neural network for supply chain management, International Journal of Mechanical Engineering and Robotics Research, 4 (2015) 96.

[2] R.G. Perea, E.C. Poyato, P. Montesinos, J.A.R. Díaz, Optimisation of water demand forecasting by artificial intelligence with short data sets, Biosystems Engineering, (2018).

[3] S. Torabi 1, E. Hassini, Multi-site production planning integrating procurement and distribution plans in multi-echelon supply chains: an interactive fuzzy goal programming approach, International Journal of Production Research, 47 (2009) 5475-5499.

[4] A. Mostafaeipour, A. Goli, M. Qolipour, Prediction of air travel demand using a hybrid artificial neural network (ANN) with Bat and Firefly algorithms: a case study, The Journal of Supercomputing, (2018) 1-24.
[5] M. Tkáč, R. Verner, Artificial neural networks in business: Two decades of research, Applied Soft Computing, 38 (2016) 788-804.
[6] A. Khalafallah, Neural network based model for predicting housing market performance, Tsinghua Science & Technology, 13 (2008) 325-328.
[7] M. Khashei, M. Bijari, A novel hybridization of artificial neural networks and ARIMA models for time series forecasting, Applied Soft Computing, 11 (2011) 2664-2675.
[8] Slimani, I. El Farissi, S. Achchab, Artificial neural networks for demand forecasting: Application using Moroccan supermarket data, in: Intelligent Systems Design and Applications (ISDA), 2015 15th International Conference on, IEEE, 2015, pp. 266-271.
[9] Y.-C. Liu, I.-C. Yeh, Using mixture design and neural networks to build stock selection decision support systems, Neural Computing and Applications, 28 (2017) 521-535.
[10] H.R. Rangel, V. Puig, R.L. Farias, J.J. Flores, Short-term demand forecast using a bank of neural network models trained using genetic algorithms for the optimal management of drinking water networks, Journal of Hydroinformatics, 19 (2017) 1-16.
[11] A. Anand, L. Suganthi, Hybrid GA-PSO Optimization of Artificial Neural Network for Forecasting Electricity Demand, Energies, 11 (2018) 728.
[12] O. Abedinia, N. Amjady, N. Ghadimi, Solar energy forecasting based on hybrid neural network and improved metaheuristic algorithm, Computational Intelligence, 34 (2018) 241-260.
[13] A.T. Dosdoğru, A. Boru, M. Göçken, M. Özçalıcı, T. Göçken, Assessment of Hybrid Artificial Neural Networks and Metaheuristics for Stock Market Forecasting, Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 27 (2018) 63-78.
[14] S.M. Seyedhosseini, M.J. Esfahani, M. Ghaffari, A novel hybrid algorithm based on a harmony search and artificial bee colony for solving a portfolio optimization problem using a mean-semi variance approach, Journal of Central South University, 23 (2016) 181-188.
[15] M.E. Günay, Forecasting annual gross electricity demand by artificial neural networks using predicted values of socio-economic indicators and climatic conditions: Case of Turkey, Energy Policy, 90 (2016) 92-101.
[16] M. Montajabiha, A. Arshadi Khamseh, B. Afshar-Nadjafi, A robust algorithm for project portfolio selection problem using real options valuation, International Journal of Managing Projects in Business, 10 (2017) 386-403.
[17] S. Mirjalili, S.M. Mirjalili, A. Lewis, Grey wolf optimizer, Advances in engineering software, 69 (2014) 46-61.
[18] R.G. Reynolds, An introduction to cultural algorithms, in: Proceedings of the third annual conference on evolutionary programming, World Scientific, 1994, pp. 131-139.
[19] C.-J. Lin, C.-H. Chen, C.-T. Lin, A hybrid of cooperative particle swarm optimization and cultural algorithm for neural fuzzy networks and its prediction applications, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 39 (2009) 55-68.
[20] Y. Sun, L. Zhang, X. Gu, A hybrid co-evolutionary cultural algorithm based on particle swarm optimization for solving global optimization problems, Neurocomputing, 98 (2012) 76-89.
[21] H.M. Song, M.H. Sulaiman, M.R. Mohamed, An application of grey wolf optimizer for solving combined economic emission dispatch problems, International Review on Modelling and Simulations, 7 (2014) 838-844.
[22] S. Khademolqorani, A hybrid model for the prioritization of municipal projects in Iran, International Journal of Applied Optimization Studies, 1 (2018) 1-10.

Alireza Goli

Alireza Goli received his Bachelor and Master Degree in Industrial Engineering from Golpayegan University of technology (Iran, 2013) and Isfahan University of Technology (Iran, 2015) respectively. Now, he is a Ph.D. candidate in Industrial Engineering at Yazd University. He is working as an active Editor/ Reviewer in different reputed journals. He has published more than 30 papers. His current research interests include supply chain management, artificial intelligence, portfolio management, meta-heuristic algorithms.

Hasan Khademi Zare

Hasan Khademi Zare is a professor of Industrial Engineering at College of Engineering, Yazd University in Iran. He obtained his Ph.D. in Industrial Engineering from Amirkabir University of Technology in Iran. Professor Hassan Khademi Zare has published 3 books, more than 500 papers in reputable academic journals and conferences.

Reza Tavakkoli-Moghaddam

Reza Tavakkoli-Moghaddam is a professor of Industrial Engineering at College of Engineering, University of Tehran in Iran. He obtained his Ph.D. in Industrial Engineering from Swinburne University of Technology in Melbourne (1998), his M.Sc. in Industrial Engineering from the University of Melbourne in Melbourne (1994) and his B.Sc. in Industrial Engineering from the Iran University of Science and Technology in Tehran (1989). He is the recipient of the 2009 and 2011 Distinguished Researcher Awards and the 2010 and 2014 Distinguished Applied Research Awards at University of Tehran, Iran. He has been selected as National Iranian Distinguished Researcher for 2008 and 2010. Professor Tavakkoli-Moghaddam has published 4 books, 17 book chapters, more than 790 papers in reputable academic journals and conferences.

Ahmad Sadeghieh

Ahmad Sadeghieh is a professor of Industrial Engineering at College of Engineering, Yazd University in Iran. He obtained his Ph.D. in Industrial Engineering from Cardiff University in England. Professor Ahmad Sadeghieh has published 2 books, more than 300 papers in reputable academic journals and conferences.

The Promotion of Graduate Programs through Clustering Prospective Students

Rosa-María Cantón-Croda*, Damián-Emilio Gibaja-Romero*, Fernando-Rey Castillo-Villar

Deanship of Graduate Studies in Engineering and Businesses, UPAEP-University Puebla, Puebla (Mexico)

Received 25 January 2019 | Accepted 10 June 2019 | Published 1 July 2019

Abstract

The promotion of academic programs, particularly at graduate levels, emerges as a response to market changes. In general, graduate programs are not a first order necessity which makes necessary the right promotion of such programs guarantee the attraction of prospective students, which enroll in some of them, which is essential for the financial sustainability of universities. Notably, the last one is a crucial problem for private universities. In this paper, we analyze the prospective students that enroll in a private to design better promotion strategies by using on data gathered by online sources. Specifically, we use clustering techniques to define marketing strategies based on segments of students. We find that age and city are crucial to promoting graduate programs while marital status and sex does not impact the decision of students in the university that we analyze.

Keywords

Clustering Methods, Decision Making, Machine Learning.

DOI: 10.9781/ijimai.2019.07.001

I. Introduction

NOWADAYS, the Internet's massification simplifies the way that consumers get information about academic programs at any education level. At the same time, online platforms, like social networks and own web-pages, allow that universities interact more directly with the society, and particularly with their prospective students. Through these interactions, universities get and generate data that summarizes students' preferences and needs. Hence, it is possible for universities the alignment of their online marketing strategies to what prospective students need or are interested [2]. For example, Pham et al. [1] observe that universities promote their academic programs through YouTube by considering students targeted segments, i.e., people that search for cuisine tutorials receive the publicity of gastronomic academic programs, while people that see space programs receive information about engineering programs. In a different market, Freeman et al. [10] provide empirical evidence about the positive impact of promoting nutrient-poor food through Facebook; and they observe that the social network offers unique digital marketing tools that increase the engagement of consumers on these products.

In general, we observe that digital platforms are structures that facilitate the sharing of information between consumers and providers of education services, and vice versa, even if no transaction is made; hence, the value of platforms relies on being knowledge intermediaries for all agents, regardless of the market's side where they stand [25]. Consequently, digital platforms are structures that universities can use to disseminate advertising about all their academic programs, even if prospective students do not ask directly for information related to education programs [26].

* Corresponding author.

E-mail addresses: rosamaria.canton@upaep.mx (R. M. Cantón-Croda), damianemilio.gibaja@upaep.mx (D. E. Gibaja-Romero).

Although digital platforms facilitate the communication between universities and prospective students, it is essential to remark that the last ones are exposed to advertisements from different products, which exposes them to excess information [7]. Moreover, obtaining a graduate certificate/master's degree/doctoral degree does not represent a first order necessity since people face family obligations or employment commitments when they want to enroll in a graduate program [11]-[12]. Excess of information and other priorities are factors that overshadow graduate programs, i.e., the probability of ignoring graduate programs' advertising is high. Hence, such factors cause a negative impact on the effectiveness of digital promotion strategies [13].

The designing of digital marketing campaigns must identify those features of prospective students that drive their decision to start and admission process, but also enroll in an academic program [6]. In other words, the marketing design must consider data from prospective consumers that become active consumers of the educations service.

The majority of universities are aware of having web pages and social networks profiles to interact with their community (prospective students, students, professors, and the society in general). Also, universities value the importance of storing data generated by the interaction in such platforms [8]. Although the theory emphasizes that owning data is distinct from having relevant information for the planning of marketing strategies [16], not all universities know how to treat information from digital sources [22].

Notably, for private universities, designing effective strategies to promote graduate programs is relevant given the high costs of their academic programs and the fact that such education level represents an opportunity for professional improvement or it is considered a personal achievement [12]. As we mentioned before, it is not a first order necessity that private universities need to understand if they want to attract new students that successfully enroll in the university [4] [6] [17].

In the case of Mexico, private universities face an additional problem. The National Council of Research and Technology

(CONACYT, its acronym by its name in Spanish, Consejo Nacional de Ciencia y Tecnología) provides scholarships to study a program that holds a CONACYT's quality certification, no matter if the university that offers them is private or public. This is a public strategy that pursues the generation of high-skilled human capital, but it also generates incentives, among prospective students, to apply to those graduate programs with the CONACYT's certification. Naturally, the largest population of graduate students belongs to public universities [27]. Consequently, private universities should make an additional effort to promote non-certified graduate programs.

We analyze the promotion's dataset of a Mexican private university that offers graduate programs with and without the CONACYT's certification. Mainly, we restrict our attention to the data generated by the university's web page since it includes those prospective students that enroll in a graduate program of the university. We analyze the dataset through machine learning techniques, particularly clustering algorithms to identify the features that drive students' decision making to enroll at a graduate program. From the marketing perspective, the identification of clusters facilitates the development of promotion campaigns by focusing on groups of graduate programs instead of launching a general campaign, which may confuse, or specific campaigns that generate an excess of information. From a technical point of view, this methodology produces better results when data is not generated by a controlled survey; for these datasets, machine learning provides more explicit logical rules for marketers [28]. We observe that age and the city, where prospective students live, are features under we segment graduate programs. Surprisingly, marriage status and sex do not provide any information about the students that enroll in a graduate program of this university. By the previous results, the design of marketing campaigns must consider segments of age and city.

The paper is organized as follows. In section II, we present a brief literature review about how private universities design marketing campaigns; we focus on machine learning to generate data-driven marketing strategies. Part III discusses the data treatment and the clustering method that we implement to partition data. Section IV presents the implementation of the clustering analysis and the main results. In the last section, we offer the conclusions.

II. Literature Review

Our paper is closely related to the literature that analyzes the selling of educational services. We present a brief review of how marketing and machine learning contribute to promoting universities since our objective is to propose marketing strategies to promote graduate programs.

A. Marketing Private Educational Services

Marketing schools have called the attention of academics and practitioners since governments, in recent years, pursues high-skilled workers generation boosting competition among universities. Some of these policies facilitate the creation of universities, while others provide financial resources to consolidate research activities. Hence, marketing schools is an area of study that emerges from understanding how universities respond to market changes due to the government's intervention; mainly, how they attract new students under public policies that prioritize strategic economic sectors [9, 29]

In New Orleans, for example, school-choice policies promote hierarchies among universities to benefit advantaged students; i.e., the government provides scholarships to those students that enroll in high ranked universities [21]. Thus, public policy generates an "unfair" competition among universities for the attraction of well-qualified students. As a reactive response, universities invest in marketing strategies to avoid disadvantaged students. Lubienski [30] finds a similar result in Michigan, where the government does not discriminate among universities. However, schools invest in marketing strategies to attract better-performing students.

Concerning the promotion of private universities, we recall that the successful attraction and retention of students are crucial tasks for financial sustainability since a significant part of their financial resources come from their students' population. Around the globe, whenever private universities do exist, these academic institutions compete against public universities to attract new students since the last ones offer more accessible tuition fees Turner [19]. Hence, graduate programs, offered by private universities, deal with extra pressure to attract new students since getting a graduate certificate is a complementary achievement [11], and financial support is scarce [12].

To guarantee their financial sustainability, private universities invest in marketing campaigns (digital or not) to attract more students. It is worth to mention that the success of this investment depends in increasing the number of prospective students, but also in generating a significant conversion rate (the percentage of prospective student that enroll to some academic program) [23].

The promotion of graduate programs requires that universities transform their data into valuable information to design effective marketing strategies. Moreover, the data analysis must consider that students do not only find the information shared by a university about their graduate programs; socio-psychological characteristics, and economic aspects as well, play a significant role in how they choose a graduate program and complete the enrollment process [20].

Typically, in the designing of marketing campaigns for graduate programs, marketers focus on showing that people with a graduate certificate increase their life's quality. So, marketing strategies usually show where former students work and how their salary increases when they get a graduate certificate [14] [15]. The previous strategy communicates the discharge profile of graduate programs, which indicates if the graduate programs help prospective students to fulfill specific objectives in their personal or professional lives [3] [4]. However, researchers agree that marketing strategies also should communicate the admission profile since prospective students need to know if the program is suitable for them [5]. For example, Turner [19] finds that universities in the United States use racial diversity as a selling point for their services; which means that promotion strategies show a welcoming environment where prospective students ate welcome regardless of their socioeconomic features. At the same time, universities show the future gains of studying in a multicultural environment. In other words, admission and discharge profiles are necessary for the development of marketing strategies.

B. Market Segmentation and Machine Learning

Market segmentation has been regarded as one of the most important strategic elements of marketing [31]. Canhoto et al. [32] refer to this concept as "the practice of grouping customers" as its primary purpose is to divide the market into different groups of consumers and then to target one or more of them with specific marketing tactics. Market segmentation brings multiple benefits to firms such as a full understanding of the market, accurate predictions of consumer behavior, and identification of new market opportunities [33]. Besides, market segmentation leads to a better allocation of financial resources to those consumer groups that the firm can satisfy [31].

However, it seems that market segmentation has lagged the current needs of marketing practitioners and new technological advances [34] [35]. Notably, the digital revolution has transformed the way consumers communicate their interests and needs so that it challenges the traditional conceptualizations of marketing strategies. Regarding market segmentation, there is a need to explore more about new segmentation variables, data analysis techniques, and segmentation models [35]. In this sense, machine learning techniques help to

improve market segmentation using data from different sources, which is a useful capacity to understand how the data "behave" [18].

For firms, the possibility of discovering well-delimited segments may generate what Kumar et al. [36] call "pockets of growth." The authors assert that dividing the market into segments enables firms to identify lucrative hot spots which may be attended through tailored offers and messages. This customized process may become a sustainable competitive advantage for firms that are competing in dynamic market environments [36]. Unfortunately, there is a substantial gap between the large amounts of information available in the market and the lack of knowledge and skills of marketers to analyze and identify patterns from it [23].

III. Methodology

In this paper, we analyze data of prospective students that ask for information about graduate programs in a private university. The dataset comes from the promotion's webpage of the graduate department.

A. Acquiring Data

Universities can get prospective students' data from different sources, like social networks and specialized survey companies. In this paper, we use the dataset gathered from the university's web page because universities have total control over the information that they share in their web page, the data that they get by interacting with its community (professors, students, and prospective students). The fact that such dataset belongs to the university represents an advantage over data from third parties to generate effective marketing strategies. Note that third parties summarize what prospective students search for, but it is baffling that such datasets point out those students that enroll in the university.

Consequently, the identification of programs that need, or not, marketing campaigns is not direct. On the contrary, the promotion's dataset summarizes the features that make appealing a graduate program for a prospective student, and we can observe if a prospective student enrolls or not in a graduate program. In other words, it is easier to identify if a marketing strategy succeeds or not in promoting a specific program.

B. Prepare the Data

We build our data by deleting all prospective students that do not enroll in any graduate program from the promotion's dataset. The dataset includes features like age, marital status, sex, address, and the graduate program where prospective students enroll. We identify that the university offers 21 Ph.D. programs and 25 masters programs in business, engineering, and social sciences. By using the software R, we split the promotion's dataset into two datasets: the first one, *datasetPHD* only includes data of Ph.D. programs, and the second one *datasetMASTER* summarizes data of masters programs. We remark that we only use the software R to analyze the dataset.

Our primary objective is to segment prospective students to improve the designing of marketing strategies. In the first analysis, we consider the whole *datasetPHD* and *datasetMASTER*. The values that each variable can take are described below

- **Age (A)**. From 27 to 76 in the Ph.D.'s dataset, and from 21 to 63 in the masters' dataset.
- **Marital status (MS)**. In both datasets, we have the values of single, married, divorced, and NC (no answer).
- **Sex (S)**. In both datasets, we find *male* and *female* values.
- **City (C)**. In *datasetMASTER*, we find that students come from 92 cities, while Ph.D. students come from 114 cities. In the appendix, we include all the values that this variable can take at each dataset.
- **Graduate programs (GP)**. This is the endogenous variable of our exercise since our objectives pursue the determination of factors that drive the enrollment of students to an academic program. In the appendix, we present all the graduate programs that university offers.

C. Pre-Process Data

Given the values that each variable can take, we decide to convert qualitative features into factors to simplify the analysis. So, we transform qualitative data into quantitative data by using the function **InsFactor** of Software R. Thus, the variables **GP** runs from one to 21 in *datasetPHD*, and **GP** runs from one to 25 in *datasetMASTER*.

Also, we find that eight Ph.D. programs (*Desarrollo Económico y Sectorial Estratégico, Desarrollo Económico y Sectorial Estratégico online, Ingeniería Mecatrónica, Ingeniería Mecatrónica online, Logística y Dirección de la Cadena de Suministro, Logística y Dirección de la Cadena de Suministro online, Planeación Estratégica y Dirección de Tecnología, Planeación Estratégica y Dirección de Tecnología online*, and eight masters' programs (*Dirección y Mercadotecnia online, Ingeniería Mecatrónica online, Logística y Dirección de la Cadena de Suministro, Logística y Dirección de la Cadena de Suministro online, Planeación Estratégica y Dirección de Tecnología, Planeación Estratégica y Dirección de Tecnología online, Psicología del Trabajo y de las Organizaciones, Sistemas Integrados de Manufactura y Estrategias de Calidad*) own the CONACYT's certification. So, we split the previous datasets into two datasets, one that only considers graduate programs with the CONACYT's certification (*datasetMASTER1* and *datasetPHD1*), and the other includes only graduate programs with no graduate certification (*datasetMASTER2* and *datasetPHD2*).

Finally, we standardized the data since variables take values in different ranges. Through this process, we simplify the comparison between variables.

D. Clustering and Segmentation

Segmentation is a marketing technique that serves to identify the right customers of a product. In our case of analysis, we search for determining the characteristics of students that enroll in a particular program and enhance marketing strategies. However, segmentation processes depend on the data that we have in hand. For example, *demographic segmentation* uses data like age, gender, and income, while *psychographic segmentation* generates a partition of customers based on values and lifestyle [24]. Note that the promotion's dataset allows us to segment prospective students into demographic clusters.

We recall that our dataset comes from the promotion's web-page [28]. Hence, unsupervised learning techniques are appealing to create clusters of prospective students and characterize students enrolled in a specific program. In other words, we segment the students of this university through clustering techniques.

Although it is possible to relax the segmentation by no creating a partition of students, i.e., the construction of non-exclusive clusters, digital marketing supports the identification of well-delimited segments to avoid overlapping advertisement, and hence, an excess of information [7].

Clustering analysis is a tool to describe the features of students that are "close enough." However, it is important to mention the lack of a general agreement in the meaning of "close enough," which implies that results are subjective. Even more, we can find different clusters, even if we use the same dataset since the methods' application relies on the selection of two key aspects:

1. **The way that we measure the similarity between variables.** Remember, we transform all qualitative features into quantitative data by using the factor function **InsFactor**. It is worth to mention that **city** indicates a spatial feature of prospective students;

however, the factor transformation does not care on how far the prospective student is from the university, instead, the city variable summarizes a student's feature. In other words, we only care in which place the prospective students are, and not in the geographical features of the city where the prospective student lives. By the previous discussion, the classical *Euclidean distance* is an appealing measure of similarity between the **GP** variable and the other variables. So, we compute the dissimilarity level through the following formula

$$d(GP, y) = \sqrt{\sum(GP - y)^2}$$

where $y \in \{A, C, Ms, S\}$. As usual, the dissimilarity level runs from zero to infinity, i.e., as the distance between variables increases, they do not share similar features.

2. **The clustering algorithm**. We use cluster analysis to identify groups of prospective students that share similar features when they enroll in an academic program. Since the university pursues the financial sustainability of all its graduate programs, no institutional strategy emphasizes the promotion of a specific graduate program. So, we use agglomerative clustering to identify those academic programs, with a high level of similarity, concerning other variables in the promotion's dataset. We search for clusters through the *k-means algorithm* since we consider that data distribute in a Euclidean space. This algorithm partitions the students' database into *k clusters* in a way that minimizes the within-cluster sum squares.

Despite the subjectivity of the clustering results, it is worth to mention that the mentioned aspects provide clustering methods with the flexibility to design marketing strategies without any bias. This feature arises since the algorithm freely searches for groups of graduate programs with similar characteristics, i.e., no pre-established set, of academic programs, insides in the finding of clusters. In practical terms, the marketer learns the students' behavior from the promotion's dataset.

E. Exploring the Data

First, it is necessary to determine if our datasets admit the creation of clusters, i.e., if elements in our datasets are "close enough." Hence, we explore the dissimilarity level of these elements.

By exploring the data, we observe that graduate programs concentrate, in proportion, almost the same number of students with some specific marital status or sex. In other words, the probability that someone married enrolls, in some graduate program, is the same for all academic programs; similarly, the probability that a woman enrolls in the *master in Dirección de Organizaciones* is equal to the probability that a woman enrolls in the *masters in Logística y Dirección de la Cadena de Suministro*. We find that previous observation hold for enrolled students in Ph.D. programs. In other words, data distribute almost uniformly when we analyze it by the variables **MS** and **S**.

Consequently, we find a single cluster when we measure the distance between programs by considering the marital status and sex. Based on this observation, the marketing proposal relies on promoting all academic programs at the same time, which compromise the success of the campaign by excess information [7] [13]. Therefore, we focus on clustering our datasets through the variables of **city** and **age**.

IV. Clustering Implementation

In this section, we first group graduate programs by considering the variable **city** to exemplify the clustering methodology. Finally, we present the results for the variable **age**.

A. The possibility of Clusters' Generation

First, we use the dissimilarities to check if the datasets admit the possibility two generate different clusters, i.e., we first verify that such variables do not present a uniformly distributed behavior. In Fig. 1 and 2, we observe the heat map of the dissimilarity matrix for Ph.D.'s and master's programs, respectively, when we try to group them by the **city** of precedence. In both cases, we observe that it is possible to group data since the dissimilarity level is low for blocks of graduate programs (the red color illustrates this feature). Even more, the diagonal of both figures indicate us that graduate programs differ concerning the location where their students come from. In other words, it is possible to group graduate programs by city, but the distribution is not uniform.

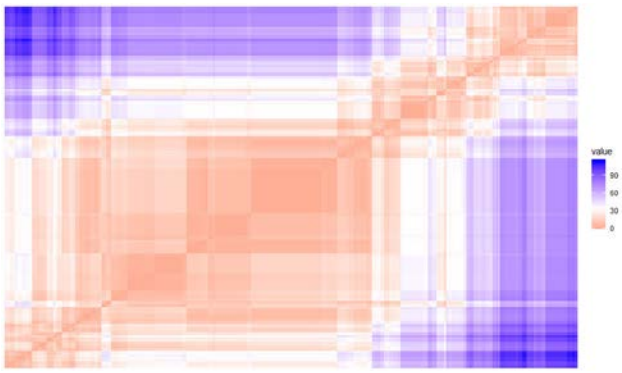


Fig. 1. Dissimilarity's heat map of Ph.D.'s programs concerning variable **C**.

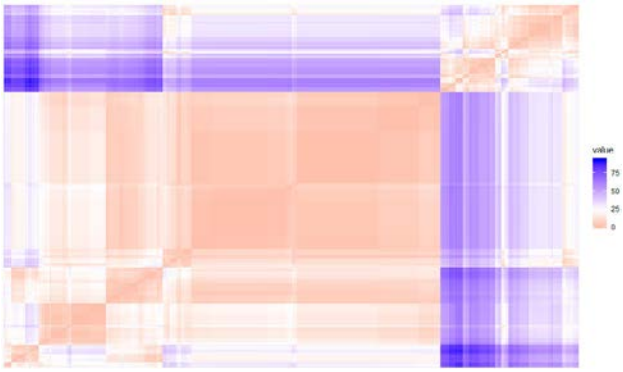


Fig. 2. Dissimilarity's heat map of master's programs for variable **C**.

B. The Optimal Number of Clusters

By the previous section, we observe that graduate programs can be grouped when we consider the variable **City**. However, the dissimilarity matrix does not indicate the optimal number of clusters.

Given that clustering is an unsupervised machine learning technique, the number of clusters is a parameter that we need to indicate as an input for our algorithm. In the literature, we can find a broad discussion about the criteria to determine the optimal number of clusters [33] [37]. In general, the criteria that we use to establish such number depends on the algorithm and the distance that we use to group our data [18] [28].

We need to indicate the number of clusters that we want to compute since clustering is an unsupervised machine learning technique. To do it, remember that we use the *k-means algorithm* to group data concerning the Euclidean distance. Under this setting, the number of clusters that we set indicates the number of centroids that the algorithm needs to find. In other words, the algorithm searches for *k* centroids, which correspond to points that minimize the sum of squares within the cluster.

During the *k*-means algorithm, centroids do not only minimize the sum of squares within data elements in a group; the algorithm also pursues the separation with other clusters. Both criteria establish that the sum of intergroup squares must increase when we establish the number of centroids to find. If such sum decreases, the distance between clusters is not enough to separate the data.

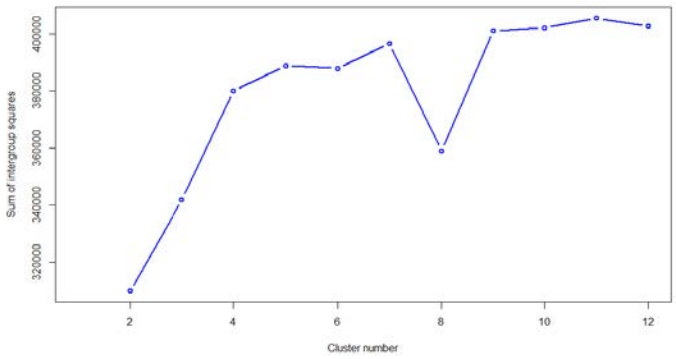


Fig. 3. Sum of intergroup squares concerning the number of clusters for the datasetPHD.

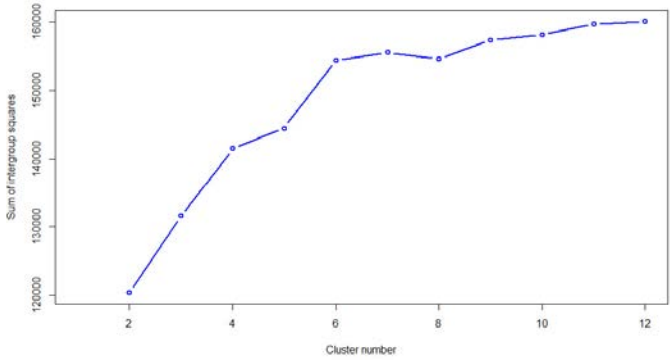


Fig. 4. Sum of intergroup squares with respect to the number of clusters for the datasetMASTER.

For our case, we use this criterion to set the number of *k* centroids, which we can find by computing the *sum of intergroup squares* through the instruction **betweens** in the software R. Below, we write the code that we use to compute the sum of intergroup squares iteratively:

sumbt1 = kmeans(ProgramCity, centers=1)$betweeens

for(i in 2:10) sumbt1[i] = kmeans(ProgramCity, centers = i)$betweenss

Note that the previous code also establishes the sum of intergroup squares as a function of the number of centroids. We illustrate how the sum of intergroup squares changes when the number of clusters increases in Fig. 3 and Fig. 4 for the *datasetPHD* and *datasetmasters*, respectively.

In the case of doctorate programs, we observe that the sum of squares decreases when the number of clusters increases to eight (see Fig. 3). In other words, under this criterion, the optimal number of clusters to partition the data of doctorate programs is seven. Concerning masters programs, we observe that the sum of squares decreases when the number of clusters is eight. Hence, we choose to compute seven clusters for the *datasetMASTER*.

Also, we verify if masters' data overlap if we consider seven clusters. We use the function **fviz** to verify if clusters are disjoint or not. This function uses the Principal Component Analysis to verify if the *k*-means algorithm outputs well-delimited clusters. Although the standardization that such instruction does is not easy to interpret, graphically the function illustrates the region that each cluster cover when we reduce to two dimensions, the dimension of a dataset [38]. Fig. 5 and Fig. 6 illustrate the graphical results of the function **fviz** when we search for seven clusters to group doctorate students and seven clusters to group masters students. When the *k*-means algorithm searches seven clusters in both datasets, in Fig. 5 and Fig. 6, we observe that no cluster overlap.

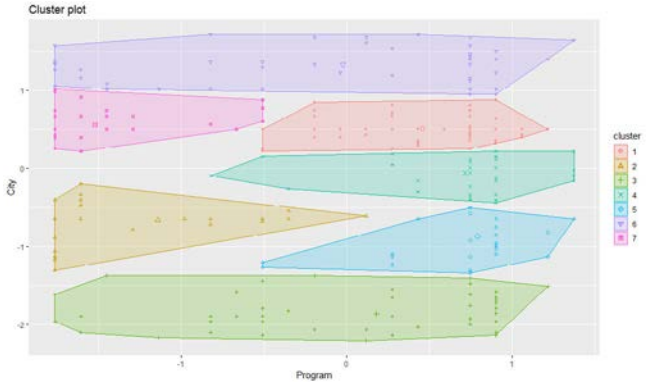


Fig. 5. Partition delimitation of datasetPHD.

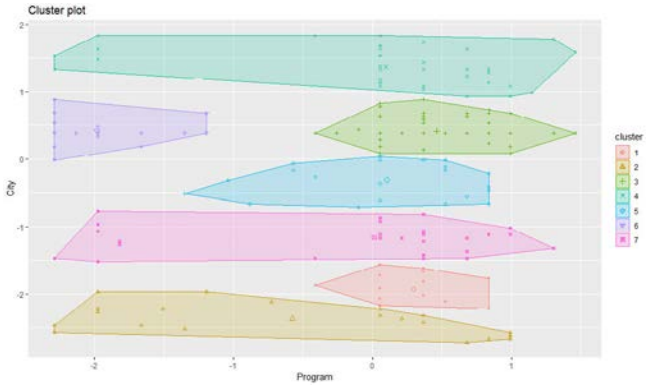


Fig. 6. Partition delimitation of datasetMASTER.

C. Clusters

In the previous sections, we discuss the possibility to cluster data since the dissimilarity level is low. Moreover, in terms of the distance between clusters, we find that seven is the "optimal" number of clusters that we can find. Finally, the Principal Component Analysis ensures the generation of non-overlapped clusters when we apply the *k*-means algorithm.

Table I shows the centroids of the seven clusters that the *k*-means algorithm finds for Ph.D. programs, while Fig. 7 graphically shows these clusters. Remember, we transform the qualitative information into quantitative information with the **InsFactor** instruction. To simplify the analysis, we rounded the centroids value to the nearest integer; in brackets, we include the program and the city that corresponds to each factor. In this table, we observe that *Mecatrónica* and *Planeación* appear twice. In the case of *Mecatrónica*, factor **12** refers to the face-to-face modality, while factor **13** corresponds to the online modality. Hence, students around *(13, 10)* are interested in the online modality of the *Mecatrónica* program, while students grouped around *(12, 103)* searches the face-to-face modality. Concerning the *Planeación* program, this is an element of two centroids given its high demand. So, these clusters indicate the cities where their students live. Specifically, the demand for this program comes around *Apodaca (6)* and *Miguel Hidalgo (63)*. Also, we can observe that *Mecatrónica* and *Planeación* appear twice. In the case of *Mecatrónica*, factor **12** refers to the face-to-face modality, while factor **13** corresponds to the online modality. Hence, students around *(13, 10)* are interested in the online modality of the *Mecatrónica* program, while students grouped around *(12, 103)* searches the face-to-face modality. Concerning the *Planeación* program, this is an element of two centroids given its high demand. So, these clusters indicate the cities where their students live. Specifically, the demand for this program comes around *Apodaca (6)* and *Miguel Hidalgo (63)*.

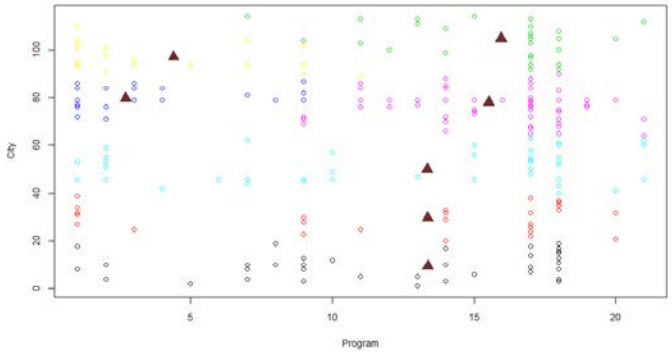


Fig. 7. Cluster visualization of datasetPHD.

TABLE I. The Centroids of the DatasetPHD Clusters

Cluster	Program	City
1	15 (Logística)	79 (Puebla)
2	4 (Finanzas)	45 (Guadalupe)
3	17 (Planeación)	6 (Apodaca)
4	13 (Mecatrónica)	10 (Benito Juárez)
5	17 (Planeación)	63 (Miguel Hidalgo)
6	12 (Mecatrónica)	103 (Tlaxcala)
7	2 (Desarrollo Económico)	80 (Puerto la Cruz)

By data in Table II, we can say that *Logistica* master is the one under students feel the most significant interest, and its students come from cities near to *Río Blanco*, *Torreón*, and *Iztacalco*. This calls our attention since *Logistica* is not the program with the largest population of students. We can conclude that *Logística* represents a point of attraction for students, but they enroll in an academic program near to it. Notably, we observe that *Mercadotecnia* and *Planeación*, the programs with the largest population of students, also belong to the clusters where *Logística* is an element of the centroid, as Fig. 8 illustrates.

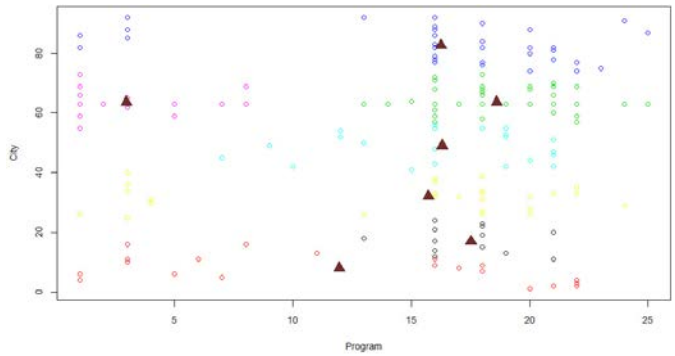


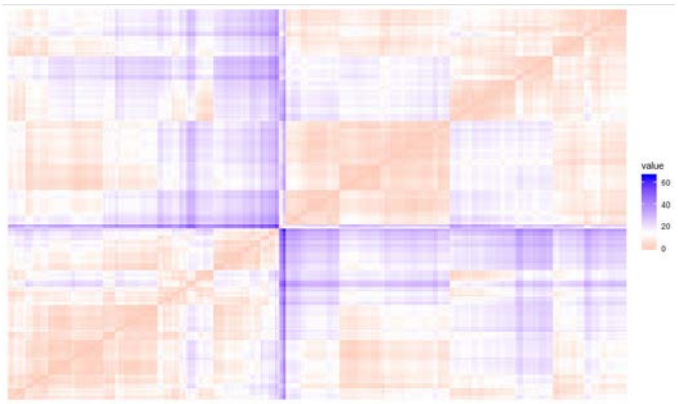
Fig. 8. Cluster visualization of datasetMASTER.

TABLE II. The Centroids of the datasetMASTER Clusters

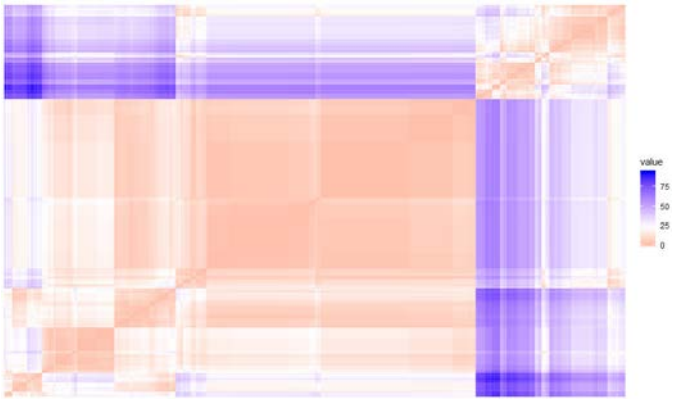
Cluster	Program	City
1	17 (Planeación)	16 (Boca del Río)
2	11 (Ingeniería de Software)	7 (Apan)
3	16 (Logística)	64 (Río Blanco)
4	16 (Logística)	82 (Torreón)
5	16 (Logística)	49 (Iztacalco)
6	2 (Administración de Empresas de Servicio)	63 (Puebla)
7	15 (Mecatrónica)	32 (Cuautla)

D. Clustering Data by Age

We apply the previous procedure to support clustering by using the variable **age**. So, we compute the dissimilarity matrix, whose heath map is illustrated in Fig. 9. In datasets, masters and Ph.D. students, Fig. 9.a and 9.b respectively, it is possible to create clusters when the distance between points is measured concerning the age. It is worth to mention that dissimilarity is more significant than in the city case, which indicates that the variance along with graduate programs increases. In other words, age is a factor that drives the decision to enroll in a graduate program in opposition to marital status.



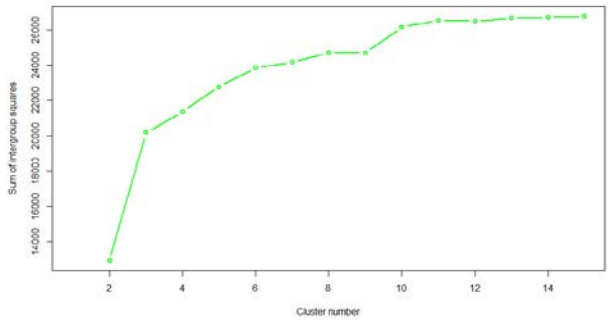
(a) Heat map of datasetPHD



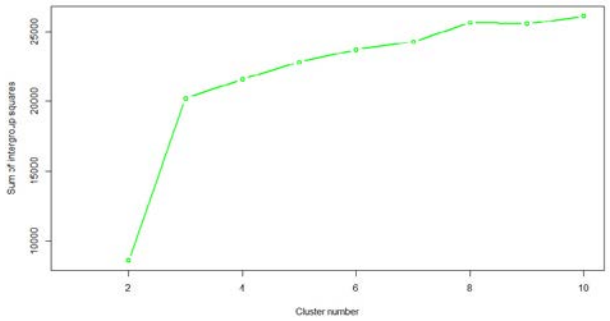
(b) Heat map of datasetMASTERS

Fig. 9. Heat map of the dissimilarity matrix when we measure distances by age.

In both cases, the dissimilarity level does not establish the optimal number of clusters to compute through the *k*-means algorithm. Hence, we use the criterion of the sum of inter-group squares (see Fig. 10). By Fig. 10.a, we observe that seven Ph.D. clusters contribute to differentiating academic programs concerning age, and Fig. 10.b suggests that eight clusters are the optimal number to group masters' programs.



(a) Number of clusters for datasetPHD



(b) Number of clusters for datasetMASTERS

Fig. 10. The sum of intergroup squares as a function of the number of clusters.

For the *datsetPHD*, Table III indicates the centroids of the seven clusters that groups Ph.D. programs.

Fig. 11 graphically shows the clusters of doctorate programs. By Table III, we observe that *Planeación* is the program that drives the decision of Ph.D. students. This is not surprising since *Planeación* has the CONACYT's certificate. However, it is worth to mention five years separate the centroids of clusters 4, 5, 6, and 7; this is a fact of interest since generations of students have different abilities to interact digitally. Usually, we consider that generations are separated by 10-15 years [13], in opposition to our findings. The previous observation may reflect the fast-changing in digital platforms/networks and accessibility to the internet as well. Finally, we note that older students are interested in the *Tecnologías de Informacion* program.

TABLE III. Centroids for Clustering Ph.D. Based on Age

Cluster	Program	Age
1	3 (Organizaciones)	45.11
2	3 (Organizaciones)	34.32
3	21 (Tecnologías de Información)	58.33
4	19 (Manufactura)	46.27
5	18 (Planeación)	40.203
6	18 (Planeación)	31.05
7	16 (Planeación)	35.37

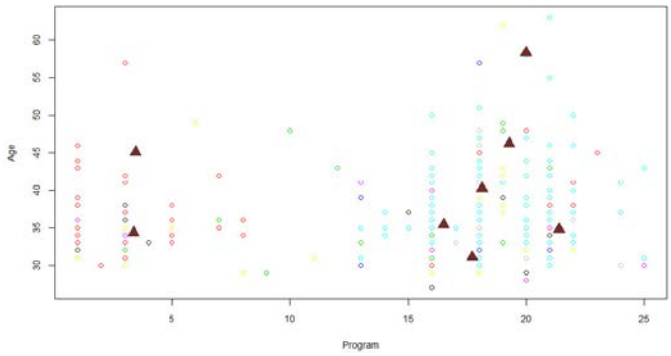


Fig. 11. Clusters of Ph.D. programs by age.

Table IV presents the centroids of the eight clusters that the *k*-means algorithm finds; we can observe them in Fig. 12.

TABLE IV. Centroids for Clustering Masters' Programs Based on Age

Cluster	Program	Age
1	3 (Agronegocios)	36.3
2	15 (Mecatrónica)	33
3	20 (Psicología del Trabajo)	42.41
4	21 (Manufactura)	36.32
5	17 (Logística)	37.17
6	18 (Planeación)	58.2
7	16 (Logística)	47
8	19 (Planeación)	31

We observe that *Planeación* y *Logística* are programs that attract the attention of masters' students. Both programs appear twice as an element of the cluster's centroids due to the online and face-to-face modalities. The online modalities correspond to the clusters of this program with the highest age. Also, it is essential to mention that *Agronegocios* is the only program with no CONACYT's certification that appears in Table IV.

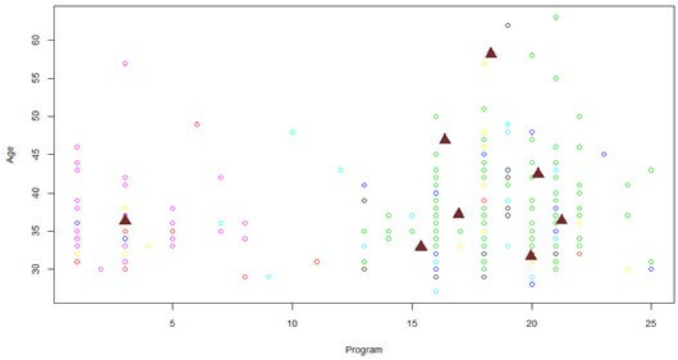


Fig. 12. Clusters of masters 'programs by age.

V. Discussing Marketing Strategies

In the previous sections, we illustrate the application of clustering methods to find similarities among graduate programs of a private university. Given the high costs of personal marketing and the probability to generate excess information, grouping graduate programs contribute to focus on segments of people interested in similar programs.

A. Doctorate Recommendations

For Ph.D. programs, we observe that it is possible to promote them by focusing on seven clusters of customers when we group them by age or city. This analysis indicates that a program, the centroid's element, may serve as an attention focus in the promotion of groups of programs.

When we group Ph.D. programs by city, we observe that *Planeación* is the centroid for programs in engineering, and it attracts students from two cities, *Miguel Hidalgo* and *Apodaca*. Hence, promotion in cities where *Planeacion* is the centroid must prioritize the marketing on this program. Even more, such marketing strategies serve to attract students for the other Ph.D. programs in the cluster.

Programs with a business/economic orientation are grouped in two different clusters. In the first one, *Finanzas* should drive the marketing strategies, while *Desarrollo Económico* is the focus for the designing of marketing strategies for the other cluster.

In the analysis by age, we observe that programs with the CONACYT's certificate (*Planeación* and *Manufactura*) dominate the formation of the cluster. In other words, Ph.D. students search for CONACYT programs, which is intuitive given the economic incentive that such an institution provides. However, *Organizaciones* is a non-certificated program that attracts two segments of students in generations separated by ten years. In the case of *Planeacion*, students' generations separated by five years;, this suggests that the university should take care of how it interacts with its students through online platforms. So, it is necessary to understand the online applications that students in each cluster use to communicate with the university. Since the promotion's dataset does not provide details about students' online habits, we suggest asking for such information. For example, the promotion webpage may include a question such as *from which online platform do you get university's information?*

The promotion's dataset does not provide complete information about the labor situation of students enrolled in *Planeación* and *Organizaciones*. Given that *Planeación* is a research program and *Organizaciones* is a business, we conjecture that their students have academic and management duties, respectively.

B. Masters Recommendations

In the master's dataset, the clustering analysis finds that CONACYT's programs lead the attraction of students since five clusters, of seven, have their centroids in one of these programs when we cluster by the city. Concerning age, we find that seven of eight clusters have their centroid in a program with the CONACYT's certification. Comparing these results with the ones obtained for the Ph.D. case, we observe that the CONACYT's certification drives the decision of students to enroll in a masters' program. Hence, the university must focus their attention on similar programs to *Organizaciones*, *Tecnologías de Información*, and *Agronegocios* to attract students that do not search for the CONACYT's scholarship.

Finally, in the masters' case, it is essential to remark that older students search for the online modality, while younger students are interested in the face-to-face scheme. These observations suggest that older students want flexibility in the way that they take courses, and consequently, marketing strategies must promote such a feature if the promotion's campaigns want to attract older students. On the contrary, marketing strategies must show the benefits of face-to-face classes to increase the interest of younger students.

Given the importance of the CONACYT's certification, Appendix B presents the same analysis for these programs. In this case, the variable age is the one that contributes to differentiating among CONACYT's programs since the distribution concerning city is almost uniform.

VI. Conclusion

The promotion of graduate education represents a significant problem for marketing designers since the attraction of new students, for these programs, highly depends on economic fluctuations. In other words, the demand for these academic programs is characterized by its volatile since getting a graduate certification is not a first order necessity. Hence, the attraction and enrollment of new students require that universities successfully communicate with their prospective students.

Online platforms, like web-pages and social networks, facilitate the communication between universities and their community. So, online interactions, through these platforms, contribute to gathering data that reflects the preferences and needs of the whole community. Consequently, it is possible to observe the features that drive the decision making of students that enroll in a particular program.

The identification of features that determine how prospective students become students of a program is a strategy that enhances the designing of marketing campaigns. In other words, promotion strategies that incorporate such features are more efficient since they identify the objective public more precisely and eliminate unnecessary information that confuses prospective consumers. In an ideal scenario, we can generate specific marketing to promote each program based on these features. However, personal marketing is too costly and difficult to achieve with the technology in hand. In this paper, we use clustering analysis to group academic programs and design marketing strategies based on the features that diminish the dissimilarities between programs. In other words, we propose the development of marketing strategies for groups of programs b considering how close they are with respect to variables like age and city.

Concerning the age variable, we observe that it is possible to group students in generations separated by five years, which may reflect how fast online interactions change. Typically, generations are separated by 15-20 year, which represents a considerable gap in an environment where online interactions change rapidly. We propose marketing strategies that include the online habits of such prospective students. We conjecture that five years of separation, between generations, is due to the velocity with Internet users change from one application to another. In this sense, we can segment promotion by considering the habits.

Also, we observe that city is a feature that drives the decision to enroll in a graduate program: non-common programs (like Economic Development or Strategic Planning) attract students that live in cities far away of the city where the university that we analyze is located. In opposition, students that live close to the university choose to enroll in more popular academic programs like Management and Mechatronics. Hence, if the university wants to attract students from cities that are not close to the university, the promotion must focus on those programs that distinguish the university concerning other universities. On the contrary, to maintain local demand, the promotion needs to focus on the advantages that the university provides by enrolling in classical/typical/common graduate programs.

Although data-driven marketing appears as a solution to improve the designing of promotion strategies, we also observe some problems

in our case of analysis. Such problems are related to data collection. For example, data collection ignores variables like job status and academic antecedents, which the literature points out as features that drive the academic career of students. Also, the registration process is not watched out, which compromised the quality of the dataset. In a first experiment, we try to do a dynamical analysis, but not all datasets satisfy data mining quality standards. Hence, online marketing must ensure how the data is collected.

Appendix

A. Graduate Programs

The university offers **masters programs** in *Administración de Empresas de Servicio, Administración de Empresas de Servicio online, Agronegocios, Agronegocios online, Arquitectura, Ciencia y Gestión Política online, Derecho Corporativo y Bancario, Desarrollo Económico y Sectorial Estratégico, Dirección de Organizaciones online, Dirección y Finanzas online, Dirección y Gerencia de Sistemas de Salud, Dirección y Mercadotecnia online, Dirección y Producción Publicitaria, Diseño y Comunicación Persuasiva, Ingeniería Mecatrónica online, Logística y Dirección de la Cadena de Suministro, Logística y Dirección de la Cadena de Suministro online, Planeación Estratégica y Dirección de Tecnología, Planeación Estratégica y Dirección de Tecnología online, Psicología del Trabajo y de las Organizaciones, Sistemas Integrados de Manufactura y Estrategias de Calidad, Tecnologías de Información y Análisis de Decisiones, Tecnologías de Información y Análisis de Decisiones online, Ventas Estratégicas y Negociación, Vivienda*.

The university offers **PhD programs** in *Desarrollo Económico y Sectorial Estratégico, Desarrollo Económico y Sectorial Estratégico online, Dirección de Organizaciones, Dirección de Organizaciones online, Dirección y Finanzas, Dirección y Finanzas online, Dirección y Mercadotecnia, Dirección y Mercadotecnia online, Ingeniería Biomédica, Ingeniería de Software, Ingeniería Mecatrónica, Ingeniería Mecatrónica online, Logística y Dirección de la Cadena de Suministro, Logística y Dirección de la Cadena de Suministro online, Planeación Estratégica y Dirección de Tecnología, Planeación Estratégica y Dirección de Tecnología online, Sistemas Integrados de Manufactura y Estrategias de Calidad, Tecnologías de Información y Análisis de Decisiones, Tecnologías de Información y Análisis de Decisiones online*.

B. The CONACYT's Certification

Table V presents the centroid of the clusters that we get when we cluster Ph.D. programs by age. Fig. 13 illustrates where the centroids are located when we analyze grouped Ph.D. programs by age.

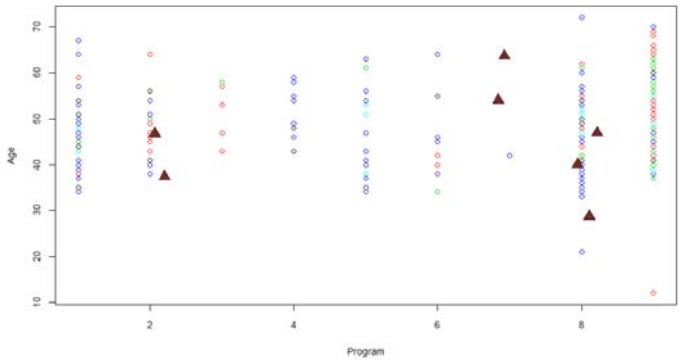


Fig. 13. Clusters of Ph.D. programs with the CONACYT's certification.

TABLE V. Centroids of Clusters When We Consider the CONACYT's Certification

Cluster	Program	Age
1	6 (Logística)	54
2	8 (Planeación)	29
3	2 (Desarrollo Económico)	47
4	2 (Desarrollo Económico)	37
5	7 (Planeación)	40
6	6 (Logística)	64
7	8 (Planeación)	47

Table VI presents the centroid of the clusters that we get when we cluster masters programs by age. Graphically, this situation is presented in Fig. 14.

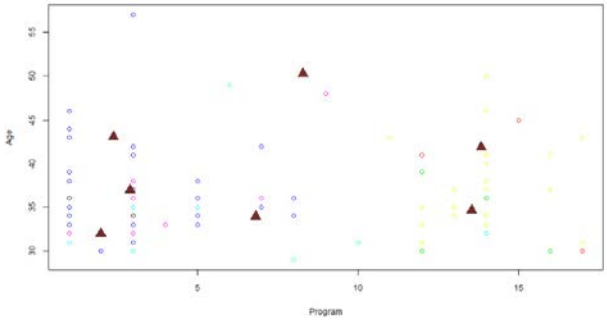


Fig. 14. Clusters (by age) of masters programs with the CONACYT's certification.

TABLE VI. Centroids of Masters Programs with CONACYT Certification

Cluster	Program	Age
1	3 (Logística)	36.9
2	13 (Planeación)	41.9
3	8 (Desarrollo Económico)	50.2
4	2 (Desarrollo Económico)	43.07
5	2 (Planeación)	31.9
6	6 (Logística)	33.9
7	13 (Planeación)	34.6

References

[1] H. H. Pham, K. Farrell, H. M. Vu, and Q. H. Vuong, "Using YouTube videos to promote universities: a content analysis," *Working Papers* CEB 17-004, Université Libre de Bruxelles, 2017.

[2] A. S. S. Tota, and M. C. U. Aguirre, "Marketing digital en universidades privadas en el estado Zulia," *Poliantea* vol. 13, no. 24, pp. 5-26, 2017.

[3] J. R. Posselt, "Inside graduate admissions," in Harvard University Press, 2016.

[4] J. Avorn, "Academic detailing: "marketing" the best evidence to clinicians," *Jama*, vol. 317, no. 4, pp. 361-362, 2017.

[5] F. Pucciarelli, and A. Kaplan, "Competition and strategy in higher education: Managing complexity and uncertainty," *Business Horizons*, vol. 59, no. 3, pp. 311-320, 2016.

[6] G. R. Johnson, C. Jubenville, and B. Goss, "Using institutional selection factors to develop recruiting profiles: Marketing small, private colleges and universities to prospective student athletes," *Journal of Marketing for Higher Education*, vol. 19, no. 1, pp. 1-25, 2009.

[7] M. Yadav, "Social media as a marketing tool: Opportunities and challenges," *Indian Journal of Marketing*, vol. 47, no. 3, pp. 16-28, 2017.

[8] C. Dimartino, and S. B. Jessen, "Selling school: The marketing of public education," Teachers College Press, 2018.

[9] M. Coccoli, A. Guercio, P. Maresca, and L. Stanganelli, "Smarter universities: A vision for the fast changing digital era," *Journal of Visual Languages & Computing*, vol. 25, no. 6, pp. 1003-1011, 2014.

[10] B. Freeman, B Kelly, L Baur, K Chapman, S Chapman, T. Gill, and L. King, "Digital junk: food and beverage marketing on Facebook," *American journal of public health*, vol. 104, no. 12, pp-56-64, 2014.

[11] D. Vergidis, and C. Panagiotakopoulos, "Student Dropout at the Hellenic Open University: Evaluation of the Graduate Program, `Studies in Education," *The International Review of Research in Open and Distributed Learning*, vol. 3, no. 2, 2002.

[12] D. Airey, and D. P. Stergiou, "Returning to education: A stressful experience?," in Lifelong Learning for Tourism, Routledge, pp. 51-69, 2017.

[13] A. Yasmin, S. Tasneem, and K. Fatema, "Effectiveness of digital marketing in the challenging age: An empirical study," International *Journal of Management Science and Business Administration*, vol. 1, no. 5, pp. 69-80, 2015.

[14] K. H. Mok, "Massification of higher education, graduate employment and social mobility in the Greater China region," *British Journal of Sociology of Education*, vol. 37, no. 1, pp. 51-71, 2016.

[15] Y. J. Lee, and M. Sabharwal, "Education–job match, salary, and job satisfaction across the public, non-profit, and for-profit sectors: Survey of recent college graduates," *Public Management Review*, vol. 18, no. 1, pp. 40-64, 2016.

[16] R. Pizarro-Milian, "What's for sale at Canadian Universities? A mixed-methods analysis of promotional strategies," *Higher Education Quarterly*, vol. 71, no. 1, pp. 53-74, 2017.

[17] M. E. M. Rosay, "La deserción en el posgrado: estudio comparativo entre maestristas de una universidad pública y privada," in Congresos CLABES, 2017.

[18] C. X. Ling and C. Li, "Data mining for direct marketing: Problems and solutions," in *Kdd*, vol. 98, pp. 73-79, Aug. 1998.

[19] E. O. Turner, "Marketing diversity: selling school districts in a racialized marketplace," *Journal of Education Policy*, vol. 33, no. 6, pp. 793-817, 2018.

[20] A. M. Kalimullin, and S. G. Dobrotvorskaya, "Higher Education Marketing Strategies Based on Factors Impacting the Enrollees' Choice of a University and an Academic Program," *International Journal of Environmental and Science Education*, vol. 11, no. 13, pp. 6025-6040, 2016.

[21] H. Jabbar, "Selling schools: Marketing and recruitment strategies in New Orleans," *Peabody Journal of Education*, vol. 91, no. 1, pp. 4-23, 2016.

[22] C. Lubienski, "Marketing schools: Consumer goods and competitive incentives for consumer information," *Education and Urban Society*, vol. 40, no. 1, pp. 118-141, 2007.

[23] C. Kanich, C. Kreibich, K. Levchenko, B. Enright, G. M. Voelker, V. Paxson, and S. Savage, "Spamalytics: An empirical analysis of spam marketing conversion," in Proceedings of the 15th ACM conference on Computer and communications security, pp. 3-14, Oct. 2008.

[24] N. Larsen, "Market segmentation: A framework for determining the right target customers," Aarhus school of business, Denmark, unpublished, 2010.

[25] D-E. Gibaja-Romero, "Interacciones en economías de plataformas," in Perspectivas de la Industria 4.0, AlfaOmega, 2019.

[26] M. Coccoli, A. Guercio, P. Maresca, and L Stanganelli, "Smarter universities: A vision for the fast changing digital era," *Journal of Visual Languages & Computing*, vol. 25, no. 6, pp. 1003-1011, 2014.

[27] E. Blanco-Livera and F. Martínez-Montes, "Las instituciones de educación superior que imparten programas en México," in Diagnóstico del Posgrado en México: Nacional, pp. 28-39, 2015.

[28] R. Flores-Lopez, and J. M. Ramon-Jeronimo, "Marketing segmentation through machine learning models: An approach based on customer relationship management and customer profitability accounting," *Social Science Computer Review*, vol. 27, no. 1, pp. 96-117, 2009.

[29] A. Zancajo, "Educational providers in the marketplace: Schools' marketing responses in Chile," *International Journal of Educational Research*, vol. 88, pp. 166-176, 2018.

[30] C. Lubienski, "Marketing schools: Consumer goods and competitive incentives for consumer information," *Education and Urban Society*, vol. 40, no. 1, pp. 118-141, 2007.

[31] L. Quinn, T. Hines, and D. Bennison, "Making sense of market segmentation: a fashion retailing case," *European Journal of Marketing*, vol. 41, no. 5/6, pp. 439-465, 2007.

[32] A. I. Canhoto, M. Clark, and P. Fennemore, "Emerging segmentation practices in the age of the social customer," *Journal of Strategic Marketing*, vol. 21, no. 5, pp. 413-428, 2013.

[33] M. N. Tuma, R. Decker, and S. W. Scholz, "A survey of the challenges and pifalls of cluster analysis application in market segmentation," *International Journal of Market Research*, vol. 53, no. 3, pp. 391-414, 2011.

[34] L. Quinn, and S. Dibb, "Evaluating market-segmentation research priorities: Targeting re-emancipation," *Journal of Marketing Management*, vol. 26, no. (13-14), pp. 1239-1255, 2010.

[35] S. Dibb, and L. Simkin, "Implementation rules to bridge the theory/practice divide in market segmentation," *Journal of marketing management*, vol. 25, no. (3-4), pp. 375-396, 2009.

[36] V. Kumar, Z. Rahman, A. Kazmi, and P. Goyal, "Evolution of sustainability as a marketing strategy: Beginning of new era," Procedia-Social and Behavioral Sciences, vol. 37, pp. 482-489, 2012.

[37] A. Ansari, and A. Riasi, "Taxonomy of marketing strategies using bank customers' clustering," *International Journal of Business and Management*, vol. 11, no. 7, pp. 106, 2016.

[38] C. Ding, and X. He, "K-means clustering via principal component analysis," in Proceedings of the twenty-first international conference on Machine learning, p. 29, Jul. 2004.

Rosa María Cantón Croda

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

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, his research interests are Game Theory and Mathematical Economics.

Fernando Reay Castillo Villar

Fernando Reay Castillo Villar, Ph.D. in Social Sciences, a Master's Degree in Political Analysis, and a Bachelor's Degree in Marketing from Tecnológico de Monterrey (ITESM). During his doctoral studies, he made a stay at the Metropolitan University of Manchester (MMU). Professor in the Graduate Programs of Management and Marketing at UPAEP-University. His areas of interest are brand development as well as the implementation of sociological concepts in Marketing. He has published research papers in international journals.

Voice Analysis Using PRAAT Software and Classification of User Emotional State

M. Magdin*, T. Sulka, J. Tomanová, M. Vozár

Constantine the Philosopher University in Nitra, Faculty of Natural Sciences, Department of Computer Science, Tr. A. Hlinku 1, 949 74 Nitra (Slovakia)

Received 31 October 2018 | Accepted 17 January 2019 | Published 15 March 2019

Abstract

During the last decades the field of IT has seen an incredible and very rapid development. This development has shown that it is important not only to shift performance and functional boundaries but also to adapt the way human-computer interaction to modern needs. One of the interaction possibilities is a voice control which nowadays can‘t be restricted only to direct commands. The goal of adaptive interaction between man and computer is the human needs understanding. The paper deals with the user's emotional state classification based on the voice track analysis, it describes its own solution - the measurement and the selection process of appropriate voice characteristics using ANOVA analysis and the use of PRAAT software for many voice aspects analysis and for the implementation of own application to classify the user's emotional state from his/her voice. In the paper are presented the results of the created application testing and the possibilities of further expansion and improvement of this solution.

Keywords

Voice Analysis, PRAAT, EMOREC, Emotional State Identification, Classification.

DOI: 10.9781/ijimai.2019.03.004

I. Introduction

SIRI (Apple), Cortana (Microsoft), Alexa (Amazon), Bixby (Samsung) - these are the names of virtual personal assistants that the user can communicate with using his/her voice. They are still far from the perfectly intelligent entities that can be seen in science fiction stories but they have a very important feature - they are able to understand human speech, to interpret it correctly and to execute user‘s commands or requests - for example to open an application, to create and to send a dictated e-mail or to find out on the internet the closest restaurants in the neighborhood and then to reply to the user in the human speech. The companies that stand behind their development do not hide the fact that they are working on another important feature implementation - the ability to recognize emotions in human speech.

Machine recognition of emotions in a human voice is another step towards real artificial intelligence and implementation of voice assistants is just one of its many potential uses. The advantages that the emotional recognition system could bring are evident in many areas, for example user-computer interaction, medicine or the business sphere. At first let’s deal with the basic terms related to the topic of the article as are emotions and their expressions which can be found and identified in a human voice.

The oldest researches of neuroscientists and psychologists have shown that human emotions play a big role in decisions making [5]. The relationship between emotions and decision is two-sided - the result of the decision affects emotions and, on the contrary, emotional state affects the decisions, this means that decision is to a certain degree predictable.

This paper is divided into several parts. The section Related work presents similar researches that have affected research in this scientific field and, of course, they have influenced this work too. The third section introduces the theoretical background of the problematics. The fourth section presents a novel approach to the emotional state recognition and classification. The fifth section deals with the results of the experiment and subsequently is presented a discussion and conclusion of the paper.

* Corresponding author.

E-mail address: mmagdin@ukf.sk

II. Related Work

Exact definitions for terms of emotion and emotional state in the explored scientific field do not exist. The term emotion is defined by psychologists, pedagogues and neuroscientists in different ways [1], [6], [13]. Depending on external or internal stimulus and on the way of perception, there are used different strategies for emotional state classification. Most commonly used classifications are the Ekman’s Classification and the Russell’s Circle Model. The defined next approaches are used depending on the assessment method, for example, emotional state categorization, so-called dimensional approach, assessment the emotional state by another qualified person (most often by psychologist) [8]:

1. Categorical approach,
2. Dimensional approach,
3. Appraisal-based approach.

All strategies for emotional states determining have, when focused on a dimensional approach, common the term valence. When is the emotional state classification considered as a psychological process, for instance, during emotion determining from speech, then valence is a variable that determines a size of change of the emotional state (e.g. from the positive level to the negative one). In case of dimensional or more precisely multidimensional approach is used for emotional state measurement and classification another variable arousal. A multidimensional approach model (Fig. 1) using these two variables is described by [2], and there is a number of works using two-dimensional models to emotional state classification too [4], [22], [10].

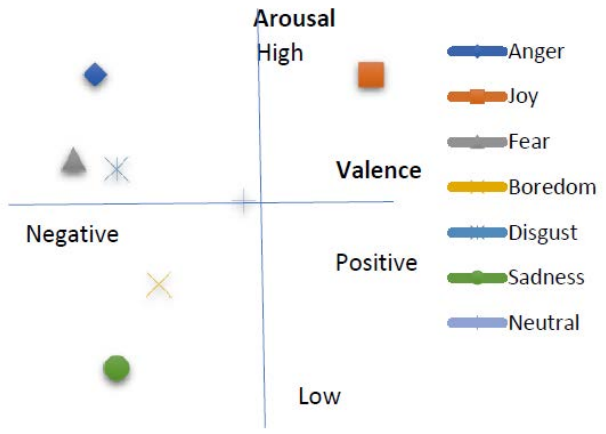


Fig. 1. Arousal-valence dimensional approach [2].

Using Arousal-valence dimensional approach model it is possible to characterize most of the emotional states that can be recognized from speech.

However, a long time ago [19], the authors have pointed to the fact that it is equally important, except valency and degree of excitement in the voice, to deal with how long this state persists. That is the reason why there is to two-dimensional model added the third dimension – time. The main reason is that according to Ekman's theory, emotion is a reaction to external stimuli, and therefore it is necessary to consider not only the emotional state that occurred, but what caused it and how long it lasted too. However, use of the approach of Scherer et al. is limited to automatic emotional state recognition only (real-time) because it causes many other research questions [21].

Most of the existing systems have been developed by using recordings of fully developed emotions but real emotions are not always obvious and can overlap each other. The recordings of different emotions are contained in speech databases. Nowadays, there is no official standard database which observe to comparing the results of different researches and methods used to recognizing emotions [12]. Emotions recognition researches are focused on recognizing only 5-6 emotions because most of databases do not cover wider spectrum of emotions [23].

Although variables valence and arousal are measurable items, emotions recognition systems can be influenced by the spoken language because values of these two variables do not match for different languages (when we read the same text in English, German or Chinese, measured values will be different). According to [11], valence can be better measured on the basis of semantic properties, while variable arousal is better measurable on the basis of acoustic functions. In ideal case, these two variables should be completely independent from the language [12]. They should be sufficient to be used in real situations too - they should recognize emotion in speech despite the noise and the din [12]. In the fourth section is presented a way how to prevent these problems by using the necessary characteristics analysis.

III. Theoretical Background

Trend of the emotional state classification from speech (from sound track) as was shown above in the Related work section will have an increasing character. However, similarly as in case of the emotional state classification e.g. from the user's facial features it is necessary to carry out a number of researches focused on the different phases of the recognition process: detection, extraction and the classification itself. The detection phase means sound analysis, the aim of which is to separate the spoken word (voice) from the other sound part of the soundtrack. The extraction phase includes the selected (necessary) soundtrack parts extracting and the analysis. The classification is the last and, essentially, the most important part of the whole process because it represents the first and quite serious problem which occurs in the detection phase right from the start. Word emotion is commonly used in many meanings and it is difficult to define objectively what the emotion actually is and when it really occurs [18]. Current psychological theories and models of emotions due to their common features can be characterized by so-called models:

1. Dimensional models: One-dimensional [7], Multidimensional [16], [17].
2. Discrete models: Circular [15], Models of Basic Emotions [9].
3. Meaning-oriented models: Lexical models, Social constructivist models.
4. Component models [14].

To sum up the previous ideas, in the models mentioned above the most frequent emotions recognition and classification studies focus only on 5-6 emotions recognition because most of databases do not cover their wider spectrum [23]. As the basic model, Ekman's Classification (happiness, surprise, anger, disgust, fear, sadness) is most commonly used in research.

Most of the existing systems listed in the Related work section dealing with the emotional state classification from the voice track were created using recordings of fully developed emotions but the actual emotions are not always obvious and can overlap. The recordings of various emotions are contained in speech databases because of the classification phase needs, similarly, for example, as in the case of the emotional state classifying from extracted parts of the face. Currently, there are available various robust databases (simulated, induced and natural) that contain photographs of people. Using the photographs of people faces it is possible to implement the classification phase using SVM, neural networks, HMM, Bayes classifier or other techniques and methods.

However, as mentioned above, in the case of emotional state classification from the voice track there is no standard database currently available to compare the results of different researches and methods that are used for emotions recognition and classification [12]. This is, however, a serious problem of the classification phase, since systems recognizing and classifying emotions can be influenced by the language of speaker. As an example, the English word "okay", depending on the situation, can express admiration, consent or disinterest. Ideally, these systems should be completely language – independent. [12]. Human voice is characterized by several features. The basic ones are:

F0 - the basic frequency of voice that contains the tonic and rhythmic characteristics of speech and is dependent on lexical content [3].

Formants F1, F2, F3 - are the amplification of certain frequencies in the spectrum and lead to the resonance of the vocal tract. They are characterized by frequency, amplitude and bandwidth. The first two formants are sufficient to identify vowels. They can also be used to emotions determining from speech [20].

Intensity - in practice, it is expressed by the acoustic pressure level and its unit is decibel. The human ear perceives intensity subjectively because it perceives the volumes of certain frequencies differently.

Jitter and shimmer - the term jitter means the cyclical changes of basic frequency. Changes are not perceived as changes of voice height but of voice quality. Shimmer indicates similar cyclical changes of energy. They require counting of the voice basic frequency during consecutive periods [20].

IV. Novel Approach of The Emotional State Recognition and Classification

In the Theoretical background section the various features that can significantly affect recognition of emotional state and then its classification are mentioned. But not all of the characteristics are sufficient or necessary to emotional state recognition and classification from the voice. From this reason, it is necessary to determine exactly the importance of these characteristics. In this case has been used the PRAAT program that is designed for scientific analysis and speech synthesis, it also allows own scripts writing and exporting the results to files. In this program were created two scripts:

1. voice_analysis.praat,
2. measure_characteristics.praat.

The flow chart in Fig. 2 is a complete description of the pipeline to process and extract relevant features from audio waveforms.

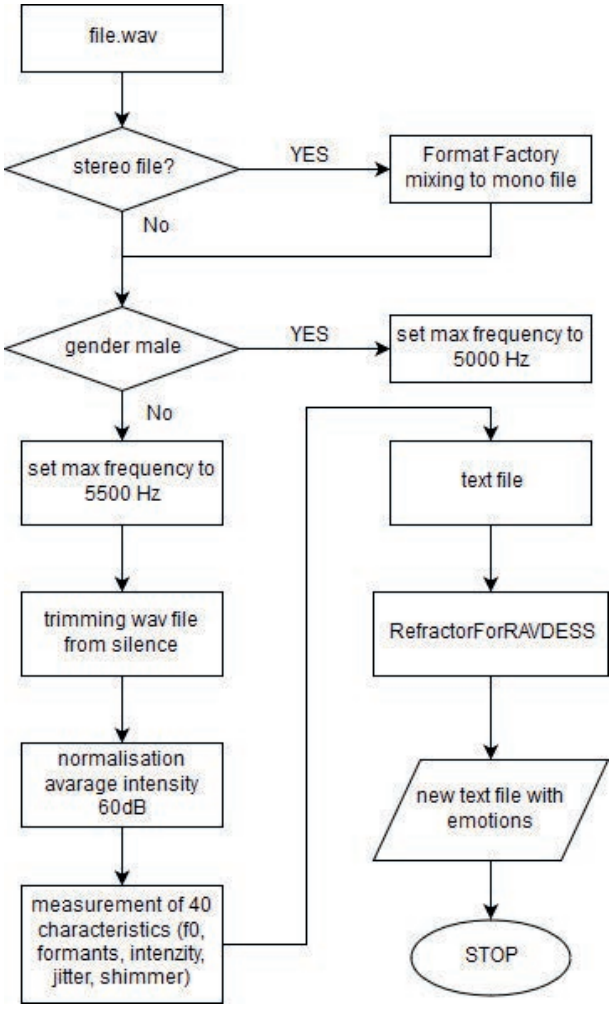


Fig. 2. The flow chart of process for extract relevant features from audio waveforms.

Inputs of the script voice_analysis.praat are selected .wav files. This script will export to output file next items: the first will be the header with the names of the measured attributes and into next lines will be written a .wav file name and then 40 measured values separated by the tabs. The script will choose a .wav file, then, before measurement, the script will edit the .wav file so that its average intensity is 60 dB and then the script will strip the silence. This is a form of normalization and the purpose is to prevent errors caused by files of varying intensity and longer passages of silence. High intensity values are typical for strong emotions such as anger but the intensity is greatly influenced, for example, by the recording technique, by settings or by the speaker distance from the microphone. In case of no correction, it could happen that the sadness expressing recording in which the speaker is close to the microphone will get the same intensity as the anger expressing recording of other speaker who is further away from the microphone. On the other hand, this correction also has an unwanted negative effect - recordings become less dynamic. Passages that contain silence without noise affect the intensity too - in some cases the intensity may reach negative values, so the silence is cropped off. We attempted to set cropping in PRAAT so that no pauses between words were cut. Created script measures 40 different characteristics (F0, formants, intensity, jitter and shimmer) for each of the input files.

The limitations could not be avoided - the script cannot be used to measure characteristics when the .wav files are mixed from men's and women's recordings; it is necessary to set the gender in the input dialogue after running the script. The reason of this limitation is that there must be used a different maximum frequency for men (5000 Hz) and for women (5500 Hz) when searching formants, otherwise the output may be incorrect. Because of the same reason, it is not possible to use the script to measure the children's voice characteristics because when searching formants it is necessary to use the frequency 8000 Hz. So, this possibility was not implemented because it was not possible to obtain child recordings for testing the application. The last limitation was caused by the fact that the function for silence cropping requires a mono track. Because of this reason if the files are recorded in two channels as stereo, the file must be pre-mixed into one mono channel before running the script. In the presented solution, was used the FormatFactory program to pre-mix stereo files.

The second script measure_characteristics.praat is a modification of the first script and it was created after the most appropriate characteristics were identified by ANOVA analysis. This script won't write a header with the names of the attributes but only the gender abbreviation - M for men and Z for women, in comparison with the first script. Then it will not measure all 40 characteristics but only the top 15 according to the analysis. There are the same limitations for both scripts.

A. RefactorForRAVDESS

When choosing the characteristics, the need to group the measured values according to emotions came out. It was a reason for simple auxiliary console application creating. It is written in C# and is called RefactorForRAVDESS. The name of this application is derived from the audiovisual database created at the Ryerson University in Toronto. It is free available database of 12 male and 12 female voice tracks of actors. The actors in these soundtracks imitate eight emotions - neutral emotion, peace, sadness, happiness, anger, fear, surprise and disgust that correlate with emotions according to the Ekman's classification. All emotions except neutral were recorded with two levels of expression - normal and strong expression. Each emotion is recorded 192 times, neutral 96 times, because it has only a normal degree of expression1. Files in the RAVDESS database are named according to key which can give us a lot of information, for instance, to identify emotion.

1 https://smartlaboratory.org/ravdess/home/

Created application RefactorForRAVDESS will read the output file from PRAAT program, in each its row will overwrite the name of the original .wav file to the emotion, and a new version of the file will be saved into the same folder as the original output file is stored. The files with characteristics prepared this way are ready to be used by ANOVA analysis.

B. Welch's Corrected ANOVA Analysis and Results

There was necessary to select from 40 measured characteristics only a few ones which can help to distinguish emotions as good as possible. Our assumption was that the characteristics tend to cluster around certain values depending on specific emotion and the purpose was to use this feature to select characteristics. The goal was to choose the characteristics that are as diverse as possible for different groups. The characteristics were chosen separately for men and women because it was assumed that key emotional expressions may be different for women and men.

There are many ways and algorithms determined for characteristics selection, one of them is to use a single-factor analysis of variance, otherwise called ANOVA. This analysis is based on unpaired t-test principle which compares the data of two groups. ANOVA is an extension of this analysis for more groups. During testing there is determined the dependence of the quantitative variable, that measures characteristics in this case, from the categorical variable, that is emotion. Emotion will be an independent variable in the analysis and the characteristics will become dependent variables, not the opposite, because the emotion is primary and the characteristics will not influence the speaker's emotional state, but they result from it. However, only the characteristics with the largest variance between groups allow the prediction of unknown emotions.

In ANOVA analysis F criterion expresses the rate of variance which is the ratio of the variance of the values between the groups and the variance of the values within the groups. An ideal case is the highest value F that means the largest variance between the groups and small variance within the groups. The use of ANOVA analysis assumes that several conditions are satisfied. Measurements results must be independent, they have to have a normal distribution and equality of the variances must be met. Meeting the first condition is relative because there is no test that would identify independence but it depends on the focus of research, on the data acquisition and measurement way. If two statistical files are created, one consisting of men and the second consisting of women, and they will be analyzed each one separately, then one set consists of twelve speakers demonstrating seven different emotions and each emotion is represented by 96 recordings, except neutral which is represented by 48 recordings. It is clear, from these numbers, that one speaker had to show one emotion more times, what indicates dependence of the samples. However, the specific case of emotion observed is a unique combination of several conditions under that it was recorded into the .wav file - the speaker, the sentence that he says, the degree of his/her expression (normal and strong expression) and the repetition (even if the speaker tries to repeat the same sentence with the same expression, the measured values of the characteristics won't be the same). It is also necessary to consider the goal of the analysis: the characteristics are measured for a small group of recordings that is available, with the aim to find any pattern (degree of variance between groups). This will be the starting point for generalization of different people in different conditions or in different languages that will be tested.

Random selection also supports independence - randomly were chosen about 73% of the recordings of each emotion - 35 recordings of neutral emotion and 70 recordings of each of other emotions. Another problem occurs when the condition of equality of variances, as shown in Fig. 3, it is not always fulfilled, as was confirmed by the Leven test. This problem can be solved by using Welch's corrected ANOVA analysis that does not require equality of variances.

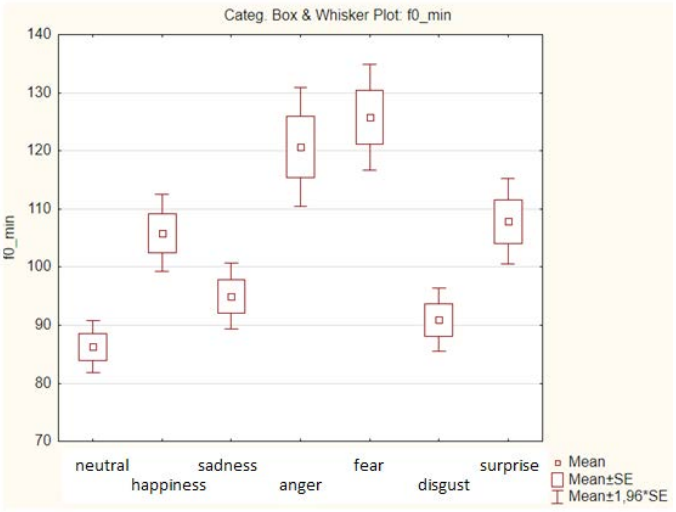


Fig. 3. Example - condition of equality of variances of characteristics f0_min is not fulfilled.

The last condition, which is also not always met, is the normal distribution of variables, but the robustness of the ANOVA test allows to bypass this condition by using a sufficiently large statistical set.

The analysis was performed using the STATISTICA tool and the characteristics were sorted according to the Welch's F from the largest to the smallest value. Of the 40 characteristics, have been selected 15 best rated men and women. It has been selected the 15 best rated features for men and women from 40 features. The features selected for men were:

f1_min, f1_max, f1_range, f1_median, f1_q1, f1_q3, f1_variance. intensity_min, intensity_max, intensity_range, intensity_median, i1_q1, i1_q3, i1_variance. jitter_ppq5, jitter_ddp, shimmer_apq5, shimmer_ddp.

The assumption that the selected features will vary depending on gender was satisfied - nine of them are identical, but on a different position in sorted lists mentioned above.

Explanation of characteristics:

F0: values are measured in Hertz.

- f0_min –the minimum F0 value of the voice recording,
- f0_max – the maximum F0 value of the voice recording,
- f0_range – differences between maximum and minimum,
- f0_median – the median of the F0 values,
- f0_q1 – the value of the first quartile of the F0 values,
- f0_q3 – the value of the third quartile of the F0 values,
- f0_variance – the standard deviation F0,
- f0_slope – absolute value of average slope F0.

Formants F1, F2, F3: values are measured in Hertz. For second and third formants the same characteristics are measured.

- f1_min – the minimum value of the first formant,
- f1_max – the maximum value of the first formant,
- f1_range – differences between maximum and minimum of the first formant,
- f1_median – the median of the values of the first formant,
- f1_q1 – the value of the first quartile of the values of the first formant,

- f1_q3 – the value of the third quartile of the values of the first formant,
- f1_variance – the standard deviation.

Intensity: values are measured in decibels.

- intensity_min – minimum value of intensity,
- intensity_max – maximum value of intensity,
- intensity_range – r differences between maximum and minimum value of intensity,
- intensity_median – the median of intensity values,
- i1_q1 – the value of the first quartile of intensity,
- i1_q3 – the value of the third quartile intensity,
- i1_variance – standard deviation of intensity.

Jitter: values are measured as a percentage.

- jitter_ppq5 – five-point PPQ (periodic perturbation quotient), the mean absolute value of the difference of the period and its average, and its four closest neighbors, divided by the average length of the period,
- jitter_ddp – the mean absolute value of the difference between successive values divided by the average period.

Shimmer: values are measured as a percentage.

- shimmer_apq5 – five-point APQ (amplitude perturbation quotient), the mean absolute value of the difference of the period of amplitude and average of this amplitude, and its four closest neighbors, divided by the average of amplitude,
- shimmer_ddp – the average absolute value of difference between successive amplitudes.

C. EmoRec

The application EmoRec was written using the programming language C# in order to further process of PRAAT program outputs and evaluate the unrecognized emotions. EmoRec is WPF application with graphical interface and was created in Microsoft Visual Studio Community 2017. The program uses .NET Framework 4.7. The application works in two modes - learning mode and recognition mode of unknown emotions.

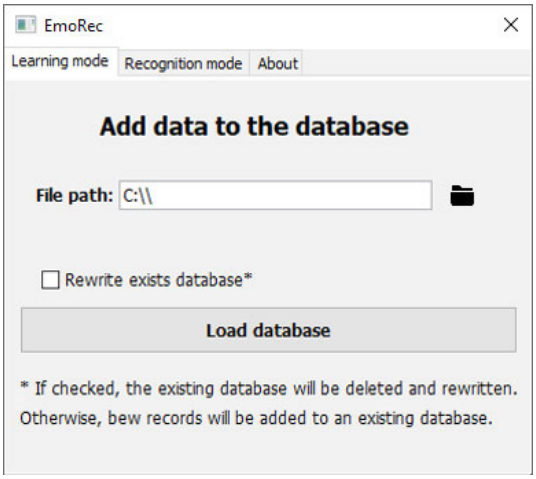


Fig.4. EmoRec – learning mode.

The learning mode (see Fig. 4) creates a database from the cases in which the values of the features have been determined for the known emotion. This mode requires a .txt input file where the first line contains a single letter - the gender abbreviation. The abbreviation ("Z" for women and "M" for men) is obligatory because EmoRec creates a database separately for men and women. So, later during emotions determining it is not necessary to find out whether the line contains a record of a woman or of a man. The remaining rows contain name of the emotion and specific numerical values of fifteen features that have reached the best F value in ANOVA analysis. When creating a database, it is possible to attach the lines into an existing database or overwrite an existing one. If the database does not exist, it will be created. When inserting a .txt file, the integrity check will be performed first and the database will be recorded only if the input file has the required structure.

The recognition mode of unknown emotions assumes as input a similar .txt file - the first line contains the gender abbreviation, the other lines contain name of the original .wav file and the measured numerical values of the attributes. Similarly, the integrity check will be performed. In this tab, it is also possible to set the type of test to be performed and the tolerance.

D. Emotion Determination Algorithm (K-NN)

EmoRec will select a male or a female database according to the first line of the input file. Then the program sequentially reads the rows from the input file and values of the attributes of unknown emotions compares by using a certain degree of tolerance with the values stored in the database. Sequentially, one row of the input file is compared with all rows in the database.

If the program finds a value in the database that matches the value of the unknown emotion within the tolerance, it increases the emotion evaluation value that is in the corresponding row of the database. At the end, the unknown emotion will be determined as the emotion with the largest evaluation value. Then the program works with the next lines of the input file. The result is exported to a .txt file with the .wav filenames and corresponding emotions.

In the principle was used the K-NN algorithm. This algorithm locates the case in the property space and compares it with the k-nearest neighbors, which are the samples of training dataset. Then the same tag is assigned to the property of the same class if it is a known neighbor. The result of ranking is decided by a majority of k-nearest neighbors.

The type of the test determines how many attributes will be compared. The 5-test means the comparison of the best five attributes, the 10-test compares the best features and the 15-test compares all the characteristics. The tolerance coefficient determines how much the value in a database can differ from the value from input file so that the result was evaluated as a match. In the program, it is possible to set the tolerance onto 1%, 3%, and 5%.

V. Experiment 1 - Results of the Tests with Different Databases

Three different databases were selected to test the application. The data for the known case database for EmoRec was obtained using the measure_characteristics.praat script and the RefactorForRAVDESS program from the same part of the RAVDESS database which was also used for the features selection. Each of the databases fulfilled its role in testing. The RAVDESS database was used to test reliability in identifying the unknown emotions of the same speakers on which the program was trained, SAVEE database tested the reliability in case of testing the files with worse audio quality, noise and unwanted sounds and EMO-DB database was used to test files that were recorded in another language.

The obtained results were put into tables for better understanding. Explanation of the labels in the tables: each label is a shortcut of the name of the emotional state. As an example, are presented the results of women testing by using the 5-test with 1% tolerance and the RAVDESS database.

TABLE I. Example of EMO-REC Program Results

		Supposed emotion							
	(%)	Ne	Ha	Sa	An	Fe	Di	Su	exchange
Detected emotion	Ne	23.07	0.00	7.69	0.00	0.00	7.69	0.00	2.56
	Ha	7.69	19.23	15.38	15.38	23.07	7.69	11.53	13.46
	Sa	15.38	11.53	15.38	7.69	3.84	23.07	11.53	12.17
	An	7.69	26.92	15.38	19.23	19.23	15.38	11.53	16.02
	Fe	0.00	3.84	7.69	23.07	34.61	3.84	0.00	6.41
	Di	46.15	23.07	26.92	19.23	0.00	30.76	3.84	19.87
	Su	0.00	15.38	11.53	15.38	19.23	11.53	61.53	12.18
	Success	29.12							

On the diagonal of the table (see Table I) it is possible to see percentages of the cases in which the emotion was determined correctly. It can be deduced from the columns percentage of the cases that emotion was exchanged by another emotion (the emotion was determined incorrectly). For example, sadness was determined to be neutral emotion in 7.69% of cases. The sum of values in each column of predicted emotions is 100%. Column Exchange shows how often in average the emotion was exchanged by another one. The disgust was incorrectly determined instead of any of another emotions averaged in 19.87% of cases. Success rate is an average of diagonal values and represents the overall average of success determinations of this test for this group.

A. RAVDESS Database

For the testing were used 27% data of the database that were not previously used to attributes selecting and creating of database of known cases for EmoRec. The smallest success rate of emotions determination was reached by neutral emotion. This emotion was almost never correctly determined, even if testing by this database achieved the best results among all three testing's. The explanation of this problem is that at such a low tolerance (5%), neutral emotional state is often changed with the emotional state - calmness. This is a problem, because then it is not possible to use Ekman's classification but the Russell model.

Testing of men reached an average success rate from 25.81% to 35.71% and testing of women reached an average success rate from 21.98% to 34.64%. For both, men and women, the best result was achieved by the 15-test with 5% tolerance (see Table II, Table III). The best-determined emotion for men in this test was anger in 61.53% of cases, followed by happiness in 53.84% of cases and disgust in 46.15% of cases. The worst-determined emotions were fear and surprise, both were correctly determined in 26.92% of cases. Neutral emotion in this test was never correctly determined, most often it was classified as sadness. Surprise and sadness were incorrectly determined as happiness, surprise in 34.61% of cases and sadness in 30.76% of cases.

Whilst for men it was the surprise classified correctly only in a few cases, the results for women were totally the opposite, the surprise was correctly classified in 88.46% of cases. The success rate for next two emotions was a bit lower. The anger was correctly classified in 53.84% of cases and the disgust in 42.30% of cases. Neutral emotion was classified as disgust in 69.23% of cases. The happiness was correctly classified only in 7.69% of cases. The disgust and the anger were exchanged between each other. The fear and the happiness were often classified as a surprise (see Table III).

TABLE II. RAVDESS – The 15-Test with 5 Percent Tolerance, Men

		Supposed emotion							
	(%)	Ne	Ha	Sa	An	Fe	Di	Su	exchange
Detected emotion	Ne	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Ha	7.69	53.84	30.76	15.38	19.23	0.00	34.61	17.95
	Sa	46.15	7.69	34.61	0.00	19.23	26.92	3.84	17.31
	An	23.07	19.23	19.23	61.53	26.92	26.92	23.07	23.07
	Fe	0.00	3.84	3.84	0.00	26.92	0.00	11.53	3.20
	Di	0.00	0.00	7.69	19.23	0.00	46.15	0.00	4.49
	Su	23.07	11.53	3.84	3.84	7.69	0.00	26.92	8.33
	Success	35.71							

TABLE III. RAVDESS – The 15-Test with 5 Percent Tolerance, Women

		Supposed emotion							
	(%)	Ne	Ha	Sa	An	Fe	Di	Su	exchange
Detected emotion	Ne	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Ha	0.00	7.69	19.23	11.53	11.53	0.00	3.84	7.69
	Sa	30.76	0.00	19.23	0.00	11.53	7.69	0.00	8.33
	An	0.00	23.07	11.53	53.84	11.53	38.46	0.00	14.10
	Fe	0.00	7.69	11.53	0.00	30.76	0.00	0.00	3.20
	Di	69.23	26.92	23.07	30.76	3.84	42.30	7.69	26.92
	Su	0.00	34.61	15.38	3.84	30.76	11.53	88.46	16.02
	Success	34.61							

B. SAVEE Database

This audio visual database was recorded in English by four male actors aged from 27 to 31 at the University of Surrey. This database contains seven emotions - neutral emotion, happiness, sadness, fear, anger, surprise and disgust. It is made up from 480 recordings, each actor has recorded 30 neutral sentences and 15 sentences for each of remaining six emotions. To check the quality of the database, subjective evaluation was performed after the database was recorded. Ten people, five men and five women participated in it. Half of them were British and the other half were people who have lived in the UK for over a year. The people were shown before testing three face images, two audio files and one video for each emotion. Then they

should play audio files, visual files and audio-visual files, and assign one emotion to each file. In the case of sound files these people were able to determine emotions correctly on average in 66.5% of cases, in visual files cases in 88.0% and in combination of sound and visual aspect case in 91.8% of cases. In case of sound recordings the disgust was most often classified instead of neutral emotion and the fear was determined instead of the sadness and the surprise. The happiness and the surprise were exchanged between each other.

The created application with this database has achieved significantly worse results than with the database RAVDESS. The average success rate ranged from 17.38% to 20.83%. The best result was achieved in the 5-test with 5% tolerance (see Table IV), the sadness was correctly determined in 51.66% of cases and the surprise in 40% of cases but disgust was correctly classified only in 1.66% of cases - this emotion was mostly incorrectly determined as the sadness and the fear. The worst result was reached for the neutral emotion which was determined as the sadness in 92.5% of cases. The happiness was classified as the surprise in 40% of the cases, while the surprise was in 21.66% of cases determined as the happiness and this is the similarity with the subjective evaluation conducted by the University of Surrey.

TABLE IV. SAVEE – Evaluation of the 5-Test with 5% Tolerance

		Supposed emotion							
	(%)	Ne	Ha	Sa	An	Fe	Di	Su	exchange
Detected emotion	Ne	0.83	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Ha	0.00	28.33	28.33	25.00	21.66	5.00	21.66	16.94
	Sa	92.50	11.66	51.66	28.33	13.33	45.00	3.33	32.36
	An	0.00	8.33	0.00	10.00	18.33	5.00	18.33	8.33
	Fe	2.50	8.33	13.33	15.00	13.33	36.66	15.00	15.14
	Di	4.16	3.33	3.33	13.33	1.66	1.66	1.66	4.58
	Su	0.00	40.00	3.33	8.33	31.66	6.66	40.00	15.00
	Success	20.83							

C. EMO-DB Database

The EMO-DB or otherwise Berlin Database of Emotional Speech was recorded in 1997 and 1999 at the Technical University of Berlin. As well as the previous two databases, this database is also included among simulated databases. It was recorded by five men and five women. The database contains neutral emotion, anger, disgust, fear, happiness, sadness and boredom.

The created application was tested with this database mostly because of the reason that the emotion recognition ideally should not be influenced by the spoken language, and this database is recorded in German language that is significantly different from English. Another special feature of this database is that it does not contain emotion surprise and contains emotion boredom. During testing, we skipped the files with the emotion boredom because the created application was not able to recognize this emotion and this emotion would always be misclassified as one of the Ekman's basic emotions.

Even this database did not match the best result reached with the RAVDESS database. As well as with the previous two databases, almost no neutral emotion was classified at all but there occurred several interesting phenomena. After evaluating the tests performed by men group it was found out that: in the 5-test with 3% tolerance the best- classified emotions were the happiness in 40.74% of cases, sadness in 48% of cases and the anger in 36.66% of cases. Disgust and sadness were exchanged between each other, anger was classified in many cases as the happiness or the surprise (see Table V).

In the 10-test and in the 15-test the success rate of the anger and the disgust determining began to increase rapidly, while other emotions were correctly classified only in very few cases and were exchanged with two mentioned emotions. This means that the anger and the disgust are very well distinguishable between each other, but the other emotions aren't at all. In the 15-test the anger was correctly determined in 86.66% of cases and the disgust was determined with 100% success rate (see Table VI).

TABLE V. EMO-DB – Evaluation of the 5-Test with 3 Percent Tolerance, Men

		Supposed emotion							
	(%)	Ne	Ha	Sa	An	Fe	Di	Su	exchange
Detected emotion	Ne	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Ha	7.69	40.74	0.00	25.00	25.00	0.00	11.54	11.54
	Sa	64.10	3.70	48.00	0.00	19.44	63.63	30.17	30.17
	An	2.56	11.11	8.00	36.66	13.88	0.00	7.11	7.11
	Fe	2.56	7.40	0.00	5.00	16.66	18.18	6.63	6.63
	Di	23.07	11.11	44.00	1.66	13.88	18.18	18.74	18.74
	Su	0.00	25.92	0.00	31.66	11.11	0.00	11.45	11.45
	Success	24.53							

TABLE VI. EMO-DB - Evaluation of the 15-Test with 5 Percent Tolerance, Men

		Supposed emotion							
	(%)	Ne	Ha	Sa	An	Fe	Di	Su	exchange
Detected emotion	Ne	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Ha	2.56	0.00	0.00	0.00	2.77	0.00	1.07	1.07
	Sa	12.82	0.00	4.00	0.00	5.55	0.00	3.67	3.67
	An	30.76	70.37	32.00	86.66	47.22	0.00	36.07	36.07
	Fe	2.56	3.70	4.00	0.00	11.11	0.00	2.05	2.05
	Di	51.28	25.92	60.00	13.33	30.55	100.00	36.22	36.22
	Su	0.00	0.00	0.00	0.00	2.77	0.00	0.46	0.46
	Success	28.89							

In the women group the EmoRec achieved less success rate than in the men group. For most of the tests an average success rate was under 20%, the best result appeared even in the 15-test with 5% tolerance where the anger was correctly determined in 65.67% of cases and the disgust in 54.28 % of cases. Other emotions are repressed again and the disgust is more often misinterpreted as the anger (see Table VII).

TABLE VII. EMO-DB – Evaluation of 15-Test with 5 Percent Tolerance, Women

		Supposed emotion							
	(%)	Ne	Ha	Sa	An	Fe	Di	Su	exchange
Detected emotion	Ne	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Ha	0.00	4.54	13.51	7.46	6.06	5.71	6.55	6.55
	Sa	10.00	2.27	2.70	0.00	0.00	5.71	3.60	3.60
	An	15.00	63.63	35.13	65.67	39.39	31.42	36.91	36.91
	Fe	2.50	4.54	2.70	5.97	15.15	0.00	3.14	3.14
	Di	72.50	13.63	29.72	5.97	30.30	54.28	30.42	30.42
	Su	0.00	11.36	16.21	14.92	9.09	2.85	9.07	9.07
	Success	21.63							

VI. Discussion of the Experiment 1

The experiments were performed as the 5-test, the 10-test and the 15-test combined with 1%, 3% and 5% tolerance with each database for both genders (except the SAVEE database that does not include women's records), together were realized 45 tests. EmoRec did not achieve good results in the tests, its average success rate of all tests was in 24.58% of cases. In all tests, neutral emotion was determined as the worst. It was most often classified as the sadness in case of RAVDESS and SAVEE databases and as the disgust in case of EMO-DB. On average, neutral emotion was correctly determined only in 2.92% of cases. Exchange with the sadness may be caused by the fact that the sadness is characterized by low intensity and monotony. On the contrary, the anger was classified as the best, in 40.45% of cases, and the disgust in 39.31% of cases. These two emotions often achieved over 50% success rate in the individual tests but in the case of EMO-DB they were incorrectly determined as the other emotions. According to the emotion determination success rate are these two emotions followed by these ones: the surprise in 34.74% of cases, the sadness in 22.64% of cases, the happiness in 18.91% of cases and the fear in 17.82% of cases.

The average success rate was negatively influenced by the low success rate of neutral emotion determination and by the low success rate of the tests with SAVEE database which is recorded in significantly lower quality. An interesting situation occurred during testing with EMO-DB in German language when only two emotions were very well distinguishable. This phenomenon could be caused by the difference of the German language from English. The best results were achieved with the same database as the program was trained for. The results could be also affected by a various number of samples, for example, EMO-DB contains only 11 recordings of the disgust emotion for men but 35 for women.

It is necessary to remark that success rate in real conditions would probably be even lower because all three databases that were tested, the EmoRec was recorded by actors, so it was simulated. The problem of real-life recordings is their gathering, the presence of background noise and the fact that people often do not show emotions openly and with full intensity - they tend to hide their real emotional state.

Considering the reached low success rate, it is requisite to classify the critical parts in the solution process that could be improved.

VII. Experiment 2 – Use of Neural Networks and Testing the Databases

From the results of Experiment 1, we can see that using the K-NN algorithm we has obtain a very low overall success rate of recognition of the emotional state. The classifier is considered successful if the overall average recognition rate is more than 80%. For this reason, we decided to change the algorithm used and use a different type of classifier. In present existing lot number of classifiers that are successfully used in recognition process (for example, in face detection and subsequent classification of the emotional state). Most often they are Hidden Markov Model (HMM), Naive Bayes classifier, Neural Nets or Support Vector Machine.

In the voice recognition and the classification of emotional state we can used all these classifiers. However in contrast to HMMs, Neural Nets create no premise about feature statistical properties and therefore where used as recognition models with high successful rate in emotion classification process for speech recognition.

For this reason, we decided to use to increase the overall success of classification phase the Artificial Neural Network. The voice flow (Fig. 5) that is representing as input data is a matrix of the features extracted from the speech inputs. The output data represent the emotional states. The percentage of input data represent the 7 various categories of Ekman classification for neural nets training. These data is validate and test chosen randomly. The training dataset find the optimal weights for each feature and the validation dataset debug the parameters of a classifier. The final model of neural net is tested using test set for estimation of the error rate.

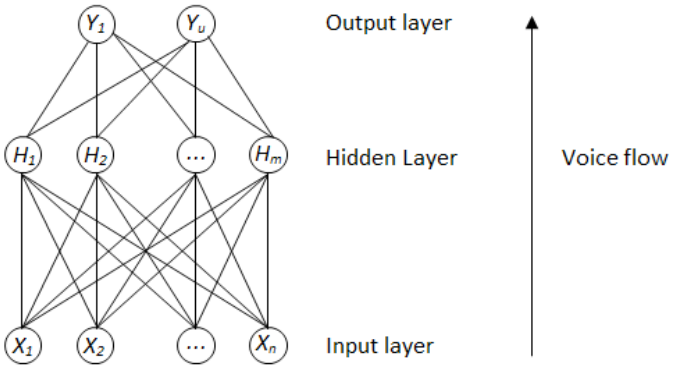


Fig. 5. Example of typical Artificial Neural Network.

To create of this network, we used a verified algorithm from Luigi Rosa (http://www.advancedsourcecode.com/neuralspeech.asp). This algorithm was written for Matlab. The test set verifies the final model and estimates the error rate. The default value for training set is 90% (the required quality) and 10 % is for error rate. We use in initially the default values but in next step the concrete count of hidden layers is chosen. The rule is: the more layers, the better the result. But to other side this rule complicated the system - we must the network trained several times.

We repeated the experiment with the same input conditions (same tests with the same tolerance). In case of the 15-Test with 5% tolerance, we achieved a total success rate of 93.26% for men. The exchange between individual emotional states was not as major when as using the K-NN algorithm (see Table VIII). Here we have achieved a total success rate of 94.66%. The exchange was in interval from 0.00-1.48.

The total success rate of emotion classification if we use the SAVEE database is 94.31% (see Table IX). The average change of emotional states is only from 0.00-2.34. We used the 5-test with 5% tolerance.

TABLE VIII. RAVDESS – Evaluation of 15-Test with 5 Percent Tolerance, Women

		Supposed emotion							
	(%)	Ne	Ha	Sa	An	Fe	Di	Su	exchange
Detected emotion (Men)	Ne	95.34	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Ha	0.12	89.84	0.89	3.21	5.32	0.12	0.26	1.65
	Sa	3.21	5.63	91.16	0.00	3.19	0.95	1.28	2.38
	An	0.56	1.27	1.62	95.41	2.34	1.12	0.14	1.16
	Fe	0.00	1.41	1.32	0.00	87.89	2.31	0.36	0.90
	Di	0.21	0.53	2.37	0.17	0.28	95.23	0.00	0.59
	Su	0.56	1.32	2.64	1.21	0.98	0.27	97.96	1.16
	Success	93.26							
Detected emotion (Women)	Ne	96.12	0	0	0	0	0	0	0.00
	Ha	0.23	93.45	0.21	2.32	2.38	0	1.6	1.12
	Sa	0.35	3.28	98.25	1.42	3.01	0.21	0.63	1.48
	An	1.27	1.29	0.91	94.23	1.28	0.62	0	0.82
	Fe	0.61	1.01	0.14	1.17	91.41	1.14	0	0.68
	Di	0.12	0.56	0.17	0.25	1.04	97.14	5.74	1.31
	Su	1.3	0.41	0.32	0.61	0.88	0.89	92.03	0.74
	Success	94.66							

TABLE IX. SAVEE – Evaluation of 15-Test with 5 Percent Tolerance, Women

		Supposed emotion							
	(%)	Ne	Ha	Sa	An	Fe	Di	Su	exchange
Detected emotion	Ne	91.25	0	0	0	0	0	0	0.00
	Ha	2.18	94.62	2.87	1.23	3.66	2.32	1.75	2.34
	Sa	1.35	1.37	93.24	0.23	1.01	0.41	0.11	0.75
	An	1.37	0.84	1.25	95.17	0.98	0.37	0.14	0.82
	Fe	1.05	1.25	0.33	1.53	92.14	0.14	0.13	0.74
	Di	1.97	1.35	1.21	0.98	1.24	96.41	0.52	1.21
	Su	0.92	0.57	1.1	0.86	0.97	0.35	97.35	0.80
	Success	94.31							

The similar results have been obtained also when we used the EMO-DB database. However, we only show results from the 15-Test with 5% tolerance (see Table X).

The K-NN algorithm in classification phase calculates on the basis at the nearest neighbour's (limit values). This algorithm can be used successfully for facial detection and subsequent classification. The algorithm recognize only a very high limit value, therefore algorithm does not work properly. However our goal was to classify emotional states that can be easily changed (for example, happiness and surprise). Therefore, the results of experiment 1 obtained very low success rate. In Experiment 2, Neural Networks have proven to be an appropriate means to increase overall success.

TABLE X. EMO-DB – The 15-Test with 5 Percent Tolerance, Men and Women

		Supposed emotion							
	(%)	Ne	Ha	Sa	An	Fe	Di	Su	exchange
Detected emotion (Men)	Ne	94.23	0	0	0	0	0	0	0.00
	Ha	0.86	91.52	0.43	0	0.72	1.07	1.12	0.70
	Sa	0.34	1.04	94.28	0	0.96	2.31	1.34	1.00
	An	1.62	2.97	1.94	97.69	1.51	1.49	0.01	1.59
	Fe	1.34	1.24	1.34	0	95.26	1.43	1.25	1.10
	Di	1.02	1.36	0.89	2.2	0.61	91.42	3.25	1.56
	Su	0.68	1.87	1.12	0.11	0.94	2.28	93.03	1.17
	Success	93.92							
Detected emotion (Women)	Ne	95.23	0	0	0	0	0	0	0.00
	Ha	0.74	94.25	0.28	1.65	0.26	0.91	1.81	0.94
	Sa	0.29	0.89	97.28	1.53	0.52	1.26	0.97	0.91
	An	0.35	1.52	0.74	91.28	1.24	0.87	1.25	1.00
	Fe	1.28	0.97	1.01	1.78	96.52	0.87	2.13	1.34
	Di	0.97	1.61	0.37	2.2	0.59	94.51	2.48	1.37
	Su	1.32	0.94	0.32	1.56	0.87	1.58	91.36	1.10
	Success	94.35							

VIII. Conclusion

The emotions determining from the user's voice is a complex issue and its solution used to be ambiguous and complicated. Different people have different voices and the creation of general rules for emotions determining is not elementary. The best results in this area are achieved by neural networks and the systems that combine the emotions determining from voice, from facial expression and other biometrics. Although the purpose of the current researches focused on emotions determining methods is to find the best one, it is clear that this area will have huge potential of being used in the future. Number of emotions models which define various emotions contribute to the complexity of this problem solution. In principle, the more emotions are trying to recognize, the harder it is to achieve high success rate of classification.

In this paper it was attempted to propose and implement solution novel approach that has been tested under various conditions. The voice recordings were analyzed, and selected the most appropriate attributes. A novel application was created that is able to determine the emotions from voice. Using neural networks, the overall success rate was about 94%, which can be considered a very good result. The greatest challenge for the future remains the creation of a system that will be able to determine the emotions reliably in real time and in real situations under not always suitable conditions for proper measurement and the evaluation of the attributes.

References

[1] Abramson, L., Marom, I., Petranker R., & Aviezer, H. "Is fear in your head? A comparison of instructed and real-life expressions of emotion in the face and body," *Emotion*, vol. 17, no. 3, pp. 557-565, 2017, doi:10.1037/emo0000252.

[2] Baron-Cohen, S. "The Eyes as Window to the Mind." *American Journal of Psychiatry*, vol. 174, no. 1, pp. 1-2, 2017, doi:10.1176/appi.ajp.2016.16101188.

[3] Busso, C., S. Parthasarathy, A. Burmania, M. Abdelwahab, N. Sadoughi, and E. M. Provost. "MSP-IMPROV: An Acted Corpus of Dyadic Interactions to Study Emotion Perception." *IEEE Transactions on Affective Computing*, vol. 8, no. 1, pp. 67-80, 2017.

[4] Cowie, R. "The enduring basis of emotional episodes: Towards a capacious overview", *2015 International Conference on Affective Computing and Intelligent Interaction, ACII 2015*, pp. 98. 2015. doi:10.1109/ACII.2015.7344557.

[5] Damasio, A. R. "*Descartes' error: emotion, reason, and the human brain*. " New York : Avon Books, 1994.

[6] Davletcharova, A., et al. "*Detection and Analysis of Emotion from Speech Signal." Procedia Computer Science*, vol. 58, pp. 91-96, 2015, doi:10.1016/j.procs.2015.08.032.

[7] Duffy, E. "An Explanation of "emotional" Phenomena without the use of the Concept emotion." *Journal of General Psychology*, vol. 25, no. 2, pp. 283-293, 1941.

[8] Ekman, P. "What Scientists Who Study Emotion Agree about." *Perspectives on Psychological Science*, vol. 11, no. 1, pp. 31-34, 2016, doi:10.1177/1745691615596992.

[9] Ekman, P. "Respect in a Pinch." *Science*, vol. 337, no. 6099, pp. 1173. 2012. doi:10.1126/science.337.6099.1173.

[10] Haq, S., T. Jan, M. Asif, A. Ali, and N. Ahmad. "Speaker-Dependent Human Emotion Recognition in Unimodal and Bimodal Scenarios." *Proceedings of the Pakistan Academy of Sciences*, vol. 52, no. 1, pp. 15-25, 2015.

[11] Karadogan S. and Larsen J. "Combining Semantic and Acoustic Features for Valence and Arousal Recognition in Speech." *in Proceedings of IEEE 3rd International Workshop on*: *Cognitive Information Processing*, Baiona, pp. 1-6, 2012.

[12] Koolagudi, S. G., Y. V. S. Murthy, and S. P. Bhaskar. "Choice of a Classifier, Based on Properties of a Dataset: Case Study-Speech Emotion Recognition." *International Journal of Speech Technology*, vol. 21, no. 1, pp. 167-183, 2018, doi: 10.1007/s10772-018-9495-8.

[13] Kragel, P. A., et al. "Decoding Spontaneous Emotional States in the Human Brain." *PLoS Biology*, vol. 14, no. 9, 2016, doi:10.1371/journal.pbio.2000106.

[14] Lövheim, H. "A New Three-Dimensional Model for Emotions and Monoamine Neurotransmitters." *Medical Hypotheses*, vol. 78, no. 2, 2012, pp. 341-348.

[15] Perlman, M. and G. Lupyan. "People can Create Iconic Vocalizations to Communicate various Meanings to Naïve Listeners." *Scientific Reports* vol. 8, no. 1, 2018, doi: 10.1038/s41598-018-20961-6.

[16] Revathi, A., N. Sasikaladevi, R. Nagakrishnan, and C. Jeyalakshmi. "Robust Emotion Recognition from Speech: Gamma Tone Features and Models." *International Journal of Speech Technology*, 2018, doi:10.1007/s10772-018-9546-1.

[17] Sharar, S. R., A. Alamdari, C. Hoffer, H. G. Hoffman, M. P. Jensen, and D. R. Patterson.. "Circumplex Model of Affect: A Measure of Pleasure and Arousal during Virtual Reality Distraction Analgesia." *Games for Health Journal*, vol. 5, no. 3, pp. 197-202, 2016, doi:10.1089/g4h.2015.0046.

[18] Shi, Y., K. A. Peng, B. Chen, Y. Gong, J. Chen, Y. Li, and Q. -J Fu. "Interaction between Speech Variations and Background Noise on Speech Intelligibility by Mandarin-Speaking Cochlear Implant Patients." *Speech Communication*, vol. 104, no. 89-94, 2018, doi:10.1016/j.specom.2018.09.007.

[19] Smith, J., A. Tsiartas, V. Wagner, E. Shriberg, and N. Bassiou. 2018. "Crowdsourcing Emotional Speech". ICASSP, IEEE international conference on acoustics, speech and signal processing – proceedings. 5139 p. 2018. doi:10.1109/ICASSP.2018.8462224.

[20] Swain, M., A. Routray, and P. Kabisatpathy. "Databases, Features and Classifiers for Speech Emotion Recognition: A Review." *International Journal of Speech Technology*, vol. 21, no. 1, pp. 93-120, 2018, doi: 10.1007/s10772-018-9491-z.

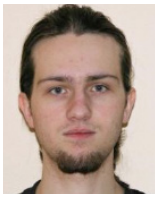
[21] Trabelsi, I., D. B. Ayed, and N. Ellouze. "Evaluation of Influence of Arousal-Valence Primitives on Speech Emotion Recognition." *International Arab Journal of Information Technology*, vol. 15, no. 4, 2018, pp. 756-762.

[22] Truesdale, D. M. and M. D. Pell. "The Sound of Passion and Indifference." *Speech Communication*, vol. 99, pp. 124-134. 2018, doi:10.1016/j.specom.2018.03.007.

[23] Uribe, A., A. Gómez, M. Bastidas, O. L. Quintero, and D. Campo. "A Novel Emotion Recognition Technique from Voiced-Speech." Paper presented at the *2017 IEEE 3rd Colombian Conference on Automatic Control, CCAC 2017 - Conference Proceedings*. 2017. doi:10.1109/CCAC.2017.8276451.

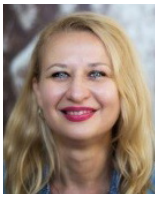
M. Magdin

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.



T. Sulka

He graduated bachelor's degree in Applied Informatics. His research area is speech analysis, identification and classification of emotional state from speech.



J. Tomanová

She works as a professor assistant at the Department of Computer Science. Her research area is the field of formal languages theory and automats, computer graphics and problems of computational complexity of algorithms.

M. Vozár

He works as a professor assistant at the Department of Computer Science. He is an expert in mathematics in computer science and programming. It deals with programming in MatLab environment (eg. neural networks).

A User-centered Smartphone Application for Wireless EEG and its Role in Epilepsy

Sofía Ahufinger[1,2]*, Paloma Balugo[3], María Mercedes González[3], Elías Pequeño[2], Henar González[4], Pilar Herrero[1,2]

[1] Ageing Lab, Centro de Tecnología Biomédica (CTB), Universidad Politécnica de Madrid (UPM) (Spain)
[2] Escuela Técnica Superior de Ingenieros Informáticos (ETSIINF), Universidad Politécnica de Madrid (UPM) (Spain)
[3] Servicio de Neurofisiología Clínica. Hospital Clínico San Carlos (HCSC) (Spain)
[4] Unidad de Innovación. Instituto de Investigación Sanitaria San Carlos (IdISSC) (Spain)

Received 18 March 2019 | Accepted 28 April 2019 | Published 22 June 2019

Abstract

Electroencephalography is well-known for its importance in the diagnosis and treatment of mental and neurological disorders and abnormalities, being especially noted in critically ill patients who suffer a variety of cerebral injuries and altered states of consciousness. However, there is an important lack of adapted equipment and applications designed to suit the clinical and research needs. Hence, patients, physicians and researchers suffer, in most cases, from a restricted mobility due to non-portable devices and wires which keep them attached to the bed, leading to an uncomfortable patient experience or difficulties during the recording. In addition, nowadays, both physicians and researchers need to access the recordings and patient information from different places such as different units or hospitals. To solve this problem, this paper presents the design and evaluation of the high-fidelity prototype of a wireless EEG smartphone application based on a user-centred design, including expert panel guidance, paper and high-fidelity prototyping and usability testing, which confirm the accuracy of the defined context of use and the validity of the prototyped application to suit the clinical and research needs. In fact, since the EEG is the most efficient and specific way to define the epileptogenic cortex, we will focus on the possible use of the presented App in epilepsy diagnosis, which is one of the main targets in the field.

Keywords

Electroencephalography, Epilepsy, Smartphone Application, User-centered Design, Wireless.

DOI: 10.9781/ijimai.2019.06.004

I. Introduction

ELECTROENCEPHALOGRAPHY (EEG) plays a central role in the diagnosis and management of several neurological disorders since its discovery in 1929 by Hans Berger [1]. Its role has been especially important for epilepsy since it is a convenient and relatively inexpensive way to detect physiological manifestations of abnormal cortical excitability [2]. The use of EEG is not only important for epilepsy diagnosis but for the location of epileptogenic zones and therapy evaluation [3]. In this case, the early diagnosis and management of the condition play an essential role when talking about overcoming seizure-related problems [4].

In fact, the use of continuous EEG monitoring is increasing in critical care units [5]. In an important number of cases, seizures are identified within the first 24 hours of monitoring [6][7]. On the other hand, within the research field, EEG is being used to study and understand a variety of neurological disorders. Nevertheless, EEG requires highly qualified and expertise neurophysiologists, plus an extensive technical infrastructure. There is an important lack of equipment and applications designed to suit the clinical and research needs [8]. Hence, patients, physicians and researchers usually suffer from restricted mobility which is reflected in the recordings since it may lead to an uncomfortable patient experience or recording difficulties.

The user-centred design aims to provide a design and development methodology which involves end-users from the very beginning until the design validation [9]. This methodology includes expert panel guidance, paper and high-fidelity prototyping and usability testing, confirming and validating the defined context of use and the prototyped application. This approach is particularly interesting and important when talking about medical applications for two main reasons which are firstly, the strengthening of personalized medicine and secondly, that all the involved users –physicians, researchers and patients– need to focus on other important aspects rather than learning how to use a complex application, reason why some developed applications are not being used.

At last but not least, citizens nowadays require more and better healthcare services, promoting personalized medicine and transforming the health system model [10]. Information should be available from different locations such as different units within the same hospital or allowing the patient's monitoring from its own house [11].

II. Epilepsy

Epilepsy is one of the most common and disabling neurologic disorders and can be defined as a heightened risk of suffering

* Corresponding author.
E-mail address: sofia.ahufinger@ctb.upm.es

recurring episodes of seizures [12]. A seizure is an abnormal and hypersynchronous discharge of a group of neurons in the brain producing an altered neuronal function, induced by defects in cerebral tissues [13]. Furthermore, it is important to consider that isolated seizures or seizures which are directly related to acute brain aggression, such as drugs or head injury, are not considered as epilepsy. In fact, differentiating an epileptic seizure from a syncope is one of the main problems of this pathology.

The definition of epileptic syndrome refers to a group of clinical features and symptoms usually occurring together (e.g., types of seizures, patient age, precipitating factors).

Epileptic seizures are mainly classified as generalized and focal seizures. At the same time, generalized seizures can be divided into two depending on the presence or absence of convulsions. Focal seizures can be divided into three types: simple seizures (consciousness is retained), complex seizures (affecting the deepest zones of the temporal lobe) and secondarily generalized seizures.

Epilepsy diagnosis is essentially clinical as it is carried out mainly by evaluation of the detailed patient history, supported in most cases by findings in EEG since other clinical tests are often normal.

A. Epilepsy in Children

Although epilepsy has a higher occurrence in both extremes of life [14], during its development, the brain has a heightened susceptibility to seizures, which leads to an elevated number of epilepsy cases during childhood [15].

Nonetheless, performing EEGs to kids is not an easy task. Babies and kids tend to be lively, always in constant movement, increased by the medical environment, leading to two main problems.

This constant movement makes hard to keep the headset in the correct position and at the same time makes hard to perform the recording.

Then again, this may cause altered recordings that may lead to an impossible or wrong diagnosis. Hence, nurses and doctors often require entertaining them by using different toys and games. For example, balloons can often be found in neurophysiology units for children entertainment, producing static electricity when they are used, that is registered as a spike-and-wave pattern in EEG, which is a typical pattern found in epileptic seizures, causing a mistaken diagnosis if this problem is not detected.

III. Electroencephalography

Electroencephalography allows measuring the electric activity of the brain using electrodes in different locations of the scalp. There are several diseases capable of mimicking seizure features, easily leading to a wrong diagnosis due to a mistaken clinical presentation [16]. Withal as we previously mentioned, EEG is the most specific noninvasive method to define the epileptogenic cortex, and hence, to diagnose epilepsy [17]. Nonetheless, its sensitivity and specificity are variable since they depend on several factors such as the procedure, montage or type of recording [18]. For this reason, both, positive and negative findings in EEG might not include or exclude the diagnosis of epilepsy respectively depending on the presence or absence of symptoms.

It is recommended to perform an EEG when the patient is suffering an apparently non-induced seizure, in order to determine wherever it is or not an epileptic seizure, type of seizure and predicting if the patient is going to suffer more episodes.

There are several types of EEG recordings, which are widely used for different purposes along the epilepsy diagnosis and monitoring. The basal EEG is performed while the patient is awake, with no need of previous preparation. The sleep-privation EEG is the same as the basal EEG but with a previous sleep-privation of 24 hours, to favour the correct patient sleep during the recording. Video-EEG provides video information synchronized with the EEG recording in order to provide information about the patient routine which can be important for the EEG interpretation. This last type of EEG is especially used in patients which are hard to diagnose or treat.

In adults with a first non-induced seizure, the first EEG shows epileptiform discharges not in all cases, but only in few of them. The probability of finding abnormalities increases when the EEG is performed during the first 48 hours since the first seizure. If the routine EEG does not detect any alteration, performing an EEG under sleep-deprived conditions increases the probability of detecting epileptiform discharges by an additional 13-35% of cases. If the diagnostic uncertainty persists, then a long-term video-EEG monitoring is considered when the epileptic events are frequent enough.

Nevertheless, the EEG technology has some limitations since the low signal amplitude (1μV-200 μV) involves an important background noise, being easy for the electrodes to record a signal that was not generated by the brain or even by the body (e.g., device interferences). Hence, it is especially important to amplify and filter these signals in different ways in order to make possible its correct interpretation. Usually, EEG devices count with a high-pass filter (HPF) to delete low frequencies (e.g., eye movement) and a low-pass filter (LPF) to delete high-frequency signals (e.g., muscle movements).

A. Montages

The EEG system is based on the comparison of signals obtained by two different electrodes. Each line or EEG channel records the potential difference between two electrodes and amplifies it (derivation). The minimum number of electrodes that should be used for a normal EEG is 10, however, in the case of epilepsy, there is a minimum number of 32 electrodes for allowing its correct diagnosis.

Montages should be as simple as possible but they must follow two rules which are: going from anterior to posterior locations and going from the left side to the right side of the scalp. Each combination of derivations is a different montage, being possible to use adjacent electrodes or distant ones, with two main types which are bipolar and monopolar montages.

In the bipolar montage, the difference between a pair of electrodes is amplified and recorded. This kind of montages can be anterior-posterior (sagittal) or transverse (coronal).

The monopolar montage, also known as referential montage, works comparing the signal obtained by a certain electrode with the one obtained by another electrode which is located in a reference zone with a lack of neural activity (e.g., the ear lobe), taking into consideration that if the reference electrode is located far from the location of the main electrode, the background noise of both electrodes would not be the same, possibly affecting the recording.

1. 10-20 System

The 10-20 system, which name refers to the electrode location in the scalp, is the most used system for electrode location. For this system, both cranial bone protuberances (nasion and inion) are taken as references. Nasion is the zone when the forehead comes together with the nose, while inion refers to the external occipital protuberance.

To locate the electrodes, the distance between inion and nasion is measured. This measure should be divided into intervals of 10% and 20%. The 10% of that distance measured from the nasion is where the fronto-polar (Fp) point should be located and the 10% measured from the inion locates the Occipital (O) point. The frontal (Fz), central (Cz) and parietal (Pz) points should be located between the Fp and O points (Fig. 1). The remaining electrodes are located following the same

methodology, being possible to use a total amount of 22 electrodes. Both pre-auricular points are frequently used as ground references for the formed electrical circuit since both of them receive the same background noise but minimal neural activity.

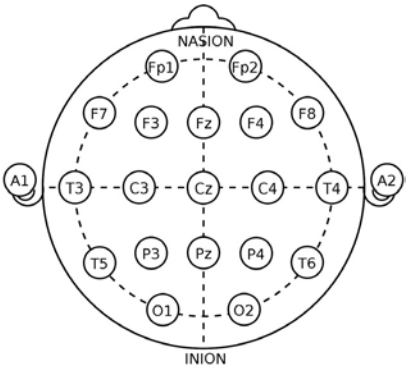


Fig. 1. International 10-20 system for electrode location in EEG procedures.

The head then is divided by the location of the electrodes, determining each one of the zones. Taking the middle line (z subindex) as the reference, even numbers should be located on the right side, while odd numbers should be located on the left.

IV. Hardware

For the development of this project, we have followed a user-centred design which focuses on giving a solution to the needs and requirements specified by the end-users of the application. This group of needs and requirements are specified in the Context of Use document, which is defined after the observation and interviews of end-users. The different hardware parts were selected taking into account this Context of Use, in order to satisfy all needs and requirements of both, medical and research staff.

On the one hand, the OpenBCI [19] platform was selected because of the possibility of performing both hardware and software modifications freely, since it is the only platform providing a wireless and open-source solution for the human-computer interaction [20].

On the other hand, as previously mentioned, the epilepsy diagnosis requires an elevated number of channels, up to 24-32. This together with the need of real-time data streaming instead of saving data in a SD card, makes the Cyton board plus the Daisy module the best option.

The Cyton board offers 8 channels, with the addition of the Daisy module, the number of channels increases to 16, which is not enough. To solve this problem, we proposed two different solutions.

The first option was modifying the Daisy module to include 3 ADS1299 chips, obtaining 32 channels when connected to the Cyton board that includes another ADS1299. This option was discarded due to the major modifications that the PCB would suffer, requiring the complete PCB redesign, unaffordable for its elevated cost in money and time and the high risk of a not working connection with the Cyton board.

The second and selected option was using two sets of Cyton board plus Daisy module, with 16 channels each of them. The main problem of this solution was synchronizing the two systems to obtain both samplings at the same time. Each Daisy module uses the clock signal of its own Cyton, and, in order to solve the previous issue, we decided to use the clock signal of one of the Cyton boards, in a way that both Cytons have the same sampling time and both Daisy modules follow the same clock signal.

The data obtained with these boards can be received by using the RFDuino included, but, this device cannot be used from a smartphone. For this reason, we decided to obtain the recorded data using a WiFi shield connected to the Cyton and Daisy boards, obtaining at the same time a bigger bandwidth.

The headset, also from OpenBCI, is 3D printed following the 10-20 system, which means that there are specific places designed to locate the electrodes. The headset counts with 32 gaps for electrode location which can be used without any restriction and is available in different sizes. The headset will have both sets of boards and a battery attached, being completely wireless as seen in Fig. 2.

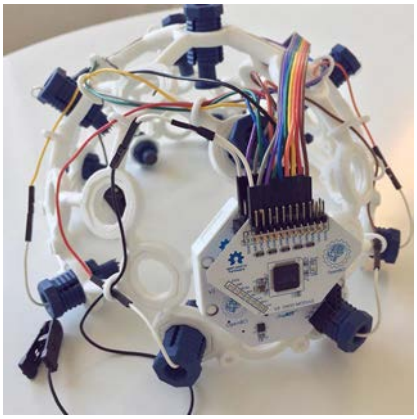


Fig. 2. OpenBCI wireless 3D-printed headset with the Cyton board and the Daisy module attached.

V. Implementation

The application interface and functionalities were designed following the Context of Use, just like the selection of hardware components. Firstly, a low-fidelity prototype was created and evaluated with different kinds of end-users, including doctors and researchers. All usability issues were found and corrected, and, all the suggested modifications and improvements were also included. Secondly, after the mentioned modifications, a high-fidelity prototype (Fig. 3) was created and evaluated by end-users, just like in the low-fidelity prototype case. After the evaluation, all the detected usability issues were again corrected, and the final features of the application were defined.

Fig. 3. Example screen obtained from the high-fidelity prototype of the application created for its evaluation.

The mobile industry is growing each and every day, with over 5 billion smartphone users in the world by the end of 2022 [21]. Almost all of these smartphones run on three well-known platforms, which are Android, iOS and Windows mobile. For many years, the development of applications for each platform required specific tools and these applications could only run on a single platform, leading to a time-consuming and expensive process since it required different specialized development teams for each platform.

During the last few years, the use of cross-platform mobile development has increased with the apparition of several frameworks like Xamarin or Flutter which allow developers to run the same code on different operating systems. However, cross-platform technologies allow for sharing 96% of the written code, requiring some platform-specific code if a native experience is wanted.

Xamarin (Microsoft) is one of the main cross-platform development technologies. This tool uses C# as programming language and native platform libraries wrapped in a .NET layer, providing support for the three main platforms.

Alternatively, Flutter (Google), also open source, uses Google's Dart as programming language and the C++ engine, highlighting its good performance, but its applications are not portable to Windows mobile and 32-bit OS devices. In any case, this is not an important drawback in this study, since most of the used smartphones nowadays run on Android or iOS and the required processing power can only be supplied by relatively new smartphones.

About the programming language, Xamarin uses a widely used language which is C#, instead of Dart like Flutter does. Despite this, although Dart is not as popular as Xamarin and its support community is still growing but quite limited at this point, it is very similar to Java and other OOP languages, making it easy to learn. For all the previously exposed reasons, the selected cross-platform framework for our project was Flutter.

A. OpenBCI Simulator

In order to develop the application without recording in real time with the headset, a board simulator was created. The WiFi shield acts just like an HTTP server, which is capable of sending the information to the smartphone or connected device. This shield responds to the Simple Service Discovery Protocol (SSDP) and hosts its unique name on local networks using multicast DNS (mDNS). To simulate this, we decided to simulate the board's REST service, creating an API web using .NET Core 2.1. For this purpose, we decided to use Swagger which is a language-agnostic specification for describing REST APIs, without any direct access to the implementation, minimizing the amount of work needed to connect disassociated services and document such service. Using the generated specification, the web-based user interface (UI) provided by Swagger offers information about the service (Fig. 4). This simulator provides the data that would be obtained by using a real OpenBCI board, offering data in two formats: JSON packets with Network Time Protocol (NTP) timestamp

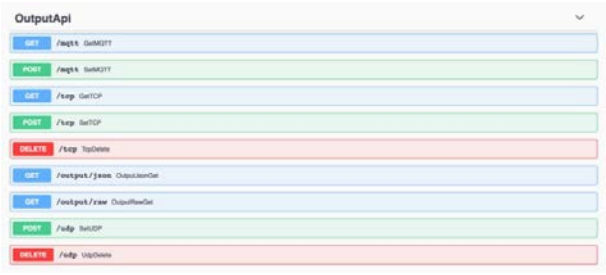


Fig. 4. OpenBCI WiFi shield REST service simulator in Swagger. The Swagger's UI provides information about each service, including the JSON and the raw data outputs.

In microseconds and arrays with channel data in nanovolts and 33-byte raw packets, which is exactly the same format as the one used by the Cyton protocol.

B. System Validation

Before the final system establishment, the user-centered design defines the need for software testing and validation, being an important phase in the life cycle. The aim of the functional testing is to verify whether if each function of the application behaves in conformance with the specified requirements or not, considering the aim for which the software was designed.

This kind of tests provide higher stability and confidence about the system performance from the end-user point of view and not from the developer point of view. This is known as "black box" testing since it is not concerned about the source code, but about how the application behaves when being used for the defined test cases. Each functionality of the system is tested by providing appropriate input and verifying the type of output and its quality by comparing it with the expected results. The different aspects that can be evaluated are the mainline functions of the application, its basic usability and accessibility (previously evaluated with the low and high-fidelity prototypes) and the error conditions. It is important to consider that the effectivity of the functional testing increases when the test conditions are defined based on the requirements defined in the Context of Use.

The results of the functional tests showed some connectivity and signal issues related to the used hardware. Firstly, connectivity problems were observed during the data acquisition, stopping the recording a few times. Secondly, the obtained signal presented a high noise level making it difficult to analyze and to use the results for an accurate patient evaluation. These two problems have been related to the used hardware and its connectivity and signal quality. Finally, we found problems in the signal synchronization between the two 16-channel groups, due to the modifications that we had to perform in the PCBs in order to use one single clock signal for both Cyton+Daisy sets, also explaining the signal noise described previously.

VI. Conclusions and Future Work

Nowadays, more than ever, it is important to manage healthcare resources as efficiently as possible and this can only be done by creating and using new systems adapted to the current environment and needs. In this paper, we proposed a user-centred design process for the development of a wireless EEG that provides a useful and actualized tool adapted for the requirements of neurophysiologists and researchers in order to understand, diagnose and treat epilepsy.

Additionally, when the patient's freedom of movement is compromised, they tend to feel anxious and, therefore, their behaviour does not correspond to their usual routine. This situation makes hard to evaluate if seizures are related to any activity or specific moment during the patient's usual routine. We solved this problem with the OpenBCI wireless headset, offering a better patient experience, and also simplifying the procedure for medical and research staff, allowing patients to move freely.

Finally, the evaluation of both prototypes showed that the initial design met all the requirements exposed in the Context of Use, obtaining the final UI design and functionalities. All the found connectivity and noise issues will be solved during the development of the final version of the application, which is being implemented using the latest technology available, Flutter. It is expected to be used not only for the epilepsy diagnosis and treatment but also for Intensive Care Units monitoring and other medical or research applications where the patient's mobility plays a key role.

Acknowledgment

This work has been supported by grants from the Community of Madrid (B2017/BMD-3760/NEUROCENTRO-CM) for research groups in Biomedicine, financed jointly by the European Social Found and European Regional Development Funds. In fact, Neurocenter (NEUROCENTRO-CM) is focused on the integrative study and treatment of neurological disorders.

References

[1] S. Siuly, Y. Li, and Y. Zhang, “Significance of EEG Signals in Medical and Health Research,” Springer, Cham, 2016, pp. 23–41.

[2] S. J. M. Smith, “EEG in the diagnosis, classification, and management of patients with epilepsy,” J. Neurol. Neurosurg. Psychiatry, vol. 76, no. suppl_2, pp. ii2-ii7, 2005.

[3] J. Engel, “A practical guide for routine EEG studies in epilepsy.,” J. Clin. Neurophysiol., vol. 1, no. 2, pp. 109–42, Apr. 1984.

[4] J. C. Das, “Electroencephalogram (EEG) in the management of epilepsy in children.,” Mymensingh Med. J., vol. 23, no. 2, pp. 406–11, Apr. 2014.

[5] N. S. Abend, D. J. Dlugos, C. D. Hahn, L. J. Hirsch, and S. T. Herman, “Use of EEG Monitoring and Management of Non-Convulsive Seizures in Critically Ill Patients: A Survey of Neurologists,” Neurocrit. Care, vol. 12, no. 3, pp. 382–389, Jun. 2010.

[6] N. Jette, J. Claassen, R. G. Emerson, and L. J. Hirsch, “Frequency and Predictors of Nonconvulsive Seizures During Continuous Electroencephalographic Monitoring in Critically Ill Children,” Arch. Neurol., vol. 63, no. 12, p. 1750, Dec. 2006.

[7] K. Williams, R. Jarrar, and J. Buchhalter, “Continuous video-EEG monitoring in pediatric intensive care units,” Epilepsia, vol. 52, no. 6, pp. 1130–1136, Jun. 2011.

[8] A. Jakab et al., “Novel wireless electroencephalography system with a minimal preparation time for use in emergencies and prehospital care,” Biomed. Eng. Online, vol. 13, no. 1, p. 60, May 2014.

[9] A. D. V. DABBS et al., “User-Centered Design and Interactive Health Technologies for Patients,” CIN Comput. Informatics, Nurs., vol. 27, no. 3, pp. 175–183, May 2009.

[10] D. Rojas and J. Carnicero, "Big Data and Public Health Systems: Issues and Opportunities," Int. J. of Interact. Multimed. Artif. Intell., vol. 4, no. 7, pp. 53-59, 2018.

[11] A. Jalal, S. Kamal, and D. Kim, "A Depth Video-based Human Detection and Activity Recognition using Multi-features and Embedded Hidden Markov Models for Health Care Monitoring Systems," Int. J. of Interact. Multimed. Artif. Intell., vol. 4, no. 4, pp. 54-62, 2017.

[12] J. P. Karis and Expert Panel on Neurologic Imaging, “Epilepsy.,” AJNR. Am. J. Neuroradiol., vol. 29, no. 6, pp. 1222–4, Jun. 2008.

[13] P. A. Dekker, “Epilepsy A manual for Medical and Clinical Officers In Africa,” 1998.

[14] V. Beletsky and S. M. Mirsattari, “Epilepsy, Mental Health Disorder, or Both?,” Epilepsy Res. Treat., vol. 2012, pp. 1–13, 2012.

[15] C. E. Stafstrom and L. Carmant, "Seizures and Epilepsy: An Overview for Neuroscientists," Cold Spring Harbor Perspectives in Medicine, vol. 5, no. 6, pp. a022426–a022426, June 2015.

[16] A. Tafakhori, V. Aghamollaii, S. Faghihi-Kashani, P. Sarraf, and L. Habibi, “Epileptic syndromes: From clinic to genetic.,” Iran. J. Neurol., vol. 14, no. 1, pp. 1–7, Jan. 2015.

[17] M. Ercegovac and I. Berisavac, “Importance of EEG in intensive care unit,” Clin. Neurophysiol., vol. 126, no. 9, pp. e178–e179, Sep. 2015.

[18] S. Noachtar and J. Rémi, “The role of EEG in epilepsy: A critical review,” Epilepsy Behav., vol. 15, no. 1, pp. 22–33, 2009.

[19] P. J. Durka, R. Kuś, J. Żygierewicz, and M. Michalska, “User-centered design of brain-computer interfaces : OpenBCI . pl and BCI Appliance,” vol. 60, no. 3, pp. 427–431, 2012.

[20] A. Vourvopoulos and S. B. I. Badia, "Usability and Cost-effectiveness in Brain-Computer Interaction: Is it User Throughput or Technology Related?," in Proceedings of the 7th Augmented Human International Conference 2016 on - AH ’16, Geneva, Switzerland, 2016, pp. 1–8.

[21] “Smartphone penetration worldwide 2014-2021 | Statista.” [Online]. Available: https://www.statista.com/statistics/203734/global-smartphone-penetration-per-capita-since-2005/. [Accessed: 26-Feb-2019].

Sofía Ahufinger

She is a researcher at Centro de Tecnología Biomédica from Universidad Politécnica de Madrid. She received her degree in Health Engineering, major in Biomedical Engineering from Universidad de Málaga in 2017 and her Master Degree in Translational Research and Personalized Medicine from Universidad de Granada in 2018. She is preparing her PhD in Software, Systems and Computing at Universidad Politécnica de Madrid. In the research field, she is currently developing user-centered applications for diagnosis and treatment of several neurological disorders and visualization and analysis of brain connectivity.

Paloma Balugo

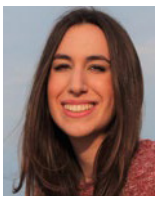
She is a Consultant in Clinical Neurophysiology. She received her MD from Universidad Complutense, Madrid, Spain and finished her specialist training in Clinical Neurophysiology from Hospital Clínico San Carlos, Madrid, Spain. She has a Master Degree in Sleep Disorders and is pursuing a Master Degree in Epilepsy. She is a member of Spanish Society for Clinical Neurophysiology. She is currently heading the Electroencefalography and Evoked Potential areas of Clinical Neurophysiology Department, Hospital Clínico San Carlos, Madrid, Spain. Her area of clinical and research interest includes electrodiagnosis of epilepsy and coma.

María Mercedes González

Dra. María de las Mercedes González Hidalgo, Graduate in Medicine and Medical Doctor by Complutense University (Madrid), Specialist Qualification in Clinical Neurophysiology in Clínica Puerta de Hierro (Madrid), Medicine Department assistant professor (Neuroloy Chair) of School of Medicine at the Complutense University (Madrid). Currently, head of Clinical Neurophysiology department of Hospital Clínico san Carlos in Madrid.

Elías Pequeño

B.S. in Computer Science from Universidad Politécnica de Madrid (UPM) with 13 years of experience in professional consulting services and implementation of Microsoft Dynamics ERP and CRM enterprise solutions. Currently studying a Ph.D. in Software, Systems and Computation at the UPM while working as CTO at Goom Spain, where he guides the design and development of customized business solutions and the architecture of systems and integrations based on Microsoft Azure cloud services.



Henar González

She is a Biomedical Engineer from Universidad Politécnica de Madrid. She is currently working at the Innovation Unit & SmartHealth Lab of the Health Research Institute of Hospital Clínico San Carlos. She works on the designing and implementation of clinical decision support systems and on the improvement of the efficiency of healthcare processes within the organization. During her MSc in Management and Innovation in Healthcare Technologies, she combined an internship in the MIT Regional Entrepreneurship Acceleration Program, which was aimed to foster Madrid entrepreneurial ecosystem, with the study of Open Innovation strategies in Life Sciences corporations. She is particularly interested in encouraging the engagement of women in technical careers, and thus has participated as a speaker in different initiatives such as WomenTechMakers.

Pilar Herrero

She is an Associate Professor at the Universidad Politécnica de Madrid (Spain). Author of more than 80 peer-reviewed international publications and editor of more than 15 international publications, European Ph.D. in Computer Science and extraordinary Ph.D. Award, in the last few years, Prof. Herrero has been involved in the organization of more than a hundred international events. She has also been involved in more than one hundred and fifty International Program Committees.

Model for Prediction of Progression in Multiple Sclerosis

Cristina Pruenza[1]*, María Teresa Solano[1], Julia Díaz[1], Rafael Arroyo[2], Guillermo Izquierdo[3]

[1] Instituto de Ingeniería del Conocimiento, Madrid (Spain)
[2] Hospital Universitario Quirónsalud Madrid y Hospital Rúber Juan Bravo, Madrid (Spain)
[3] Hospital Vithas NISA, Seville (Spain)

Received 25 January 2019 | Accepted 10 June 2019 | Published 24 June 2019

Abstract

Multiple sclerosis is an idiopathic inflammatory disease of the central nervous system and the second most common cause of disability in young adults. Choosing an effective treatment is crucial to preventing disability. However, response to treatment varies greatly between patients. Because of this, accurate and timely detection of individual response to treatment is an essential requisite of efficient personalised multiple sclerosis therapy. Nowadays, there is a lack of comprehensive predictive models of response to individual treatment.This paper arises from the clinical need to improve this situation. To achieve it, all patient's information was used to evaluate the effectiveness of demographic, clinical and paraclinical variables of individual response to fourteen disease-modifying therapies in MSBase, an international cohort. A personalized prediction model to three stages of disease, as a support tool in clinical decision making for each MS patient, was developed applying machine learning and Big Data techniques. These techniques were also used to reduce the data set and define a minimum set of characteristics for each patient. Best predictors for the response to treatment were identified to refine the predictive model. Fourteen relevant variables were selected. A web application was implemented to be used to support the specialist neurologist in real time. This tool provides a prediction of progression in EDSS from the last relapse of an individual patient, and a report for the medical expert.

Keywords

Machine learning, Big Data, Predictive Models, Multiple Sclerosis, relapses, DMT, EDSS.

DOI: 10.9781/ijimai.2019.06.005

I. Introduction

MULTIPLE sclerosis (MS), also known as demyelinating myelopathy, is a disease characterized by the appearance of demyelinating, neurodegenerative and chronic lesions of the central nervous system. MS is the second most common cause of disability in young adults and is associated with significant societal costs [1].

At this moment, no neurodegenerative or remyelinating therapies are available for clinical use and so the core of multiple sclerosis management lies in preventing episodic inflammation and relapse-related disability accrual. This fact together with the search for its causes, still unknown, is an active field of research.

Because of disease's effects on the central nervous system, it can present a series of symptoms that appear in the form of relapses or progress slowly over time, can result in reduced mobility and disability in the most severe cases. In general, five years after the appearance of the first symptoms, slightly more than 50% of patients have some kind of mild affectation, while 40% of patients show moderate affectation and less than 10% of patients die from the consequences of multiple sclerosis or its complications.

Despite the rapid development of multiple sclerosis pharmacotherapy over recent years, there is no cure for this disease, but there are several multiple sclerosis disease-modifying therapies (known as DMTs). It is interesting to evaluate patients' response to each DMT and which variables are good predictors for that response.

Previous works evaluate and describe which clinical and biochemical variables predict the progress of a patient toward multiple sclerosis [2] [3], and study the long-term influence of relapsing-remitting MS (RR) [4]. But not only do traditional clinical factors influence the prognosis of this disease, [5] shows that quality of life levels provide additional prognostic information about MS disability. This reinforces the importance of incorporating other factors into the study.

Therefore, accurate and timely detection of individual response to these DMTs is an essential requisite of efficient personalised multiple sclerosis therapy. Even though the prediction of individual disease course has now become feasible [6], [7], prediction of individual treatment response remains an area that needs further research.

In this paper the efficacy of demographic, clinical and paraclinical variables of response to DMTs is evaluated and models are explored to determine those relevant in three stages of disease (Clinically Isolated Syndrome (CIS), Relapsing-Remitting MS (RRMS) and Secondary Progressive MS (SPMS)), without considering the other two stages (Primary Progressive MS (PPMS) and progressive relapsing MS (PRMS)). The international MSBase cohort has been analysed, and Machine Learning algorithms and Big Data techniques have been used to extract knowledge about this disease.

* Corresponding author.
E-mail address: cristina.pruenza@iic.uam.es

II. Proposed Methodology

A tool has been developed that is able to predict an individual treatment response by using the patient's clinical information. This tool serves as a consultation and aid in medical decision-making. In order to achieve this work, a multidisciplinary team has been formed with clinical experts and data scientists, who have worked together to validate the data used and the results obtained. This work provides an opportunity to study the evolution of multiple sclerosis in an international sample of considerable proportions.

This section is structured as follows: First, patient data collection is discussed. Second, a descriptive analysis of the sample is carried out as per the different variables of the patients, with a study of their different behaviour depending on different factors, and a global vision of the disease is acquired. Most relevant variables are used to predict the response to a treatment. A study is presented on how to quantify this response based on available data and to implement a prediction model using the most relevant variables with Machine Learning and Big Data techniques.

A. Data Integration

1. Data Collection

The real-world data set in the *MSBase Neuro-Inmunology Registry*[1][8] was provided in plain text files directly from the international MSBase cohort. Every file contains anonymized clinical information about patients, compliant with the General Data Protection Regulation (G.D.P.R.), along with longitudinal data from their electronic medical records (EMRs).

Five databases were used as data sources. They stored 551,440 visits, 150,217 relapses, 235,902 magnetic resonances imaging (MRIs), 193,130 treatments and 48,186 patients from 64 different countries. All these records were in a structured format.

The raw data consisted of all the available data up to that moment, coming from clinical records, covering a period of time slightly longer than 50 years, from 1966 until May 2017.

A total of 185 variables were available in the raw data set. A treatment to prepare the variables was necessary due to the nature of the data and the disparate way in which doctors and countries filled in the variables. The selected variables (attributes or features) were reviewed by a team of clinical experts. As many as 92 new variables were built, and two variables extracted directly from the raw data, resulting in 94 explanatory variables to describe a patient.

2. Data Preparation and Validation

The extracted data are often incomplete, contain unnecessary or ambiguous information, suffer disruptions due to noise or pose other difficulties that affect the performance of the predictive models. Therefore, it is necessary to pre-process and validate them to avoid future inconveniences.

The process of extracting variables out of the patients' data is long and tedious and needs some collaboration from the clinical expert to validate them. The first step was the validation of each extracted variable with the medical team, thus obtaining those that are relevant and are usually filled in correctly by the medical community. In that process, we study how each variable is filled in, its degree of objectivity and the presence of absent values.

The inclusion criteria to enrol patients were: availability of the minimum dataset (i.e. *patient gender, birth date, first symptom date* and *first stage of disease*) and consistency between dates in their records and date of birth.

Visits were included, but only those without relapses (i.e. there is no relapse in at least 30 days prior to the visit). This decision was adopted due to the large number of absent values in longitudinal data relating to Extended Disability Status Scale (EDSS) score, which measures the patient's disability [9], associated with each visit and with lesions in the MRIs. The clinical expert team consider EDSS information of great relevance.

Names of each *Disease-Modifying Therapy* (DMT) were coded with the main active principles corresponding to MS treatments, and any medication not specific to the disease or lacking a start date was discarded. Fourteen DMTs were studied, namely *Aubagio Avonex, Betaferon, Extavia, Copaxone, Cyclophosphamide, Gilenya, Lemtrada, Novantrone, Plegridy, Rebif, Rituximab, Tecfidera. Tysabri. Betaferon* and *Extavia* have the same active principle (*Interferon beta-1b*) and therefore they were grouped together for the study.

This data assessment used a process of data quality and generalizability that allowed identification of any incomplete, invalid or inconsistent entry.

3. Data Exploration

Some findings originated during data exploration should be highlighted.

In the study of the evolution of DMTs in patients following a CIS-RRMS-SPMS progression, different behaviours are observed whose start date is between two periods of time: 1975-2005 (Fig. 1) and 2005-2017 (Fig. 2). The fourteen DMTs described above have been named as follows: *Aubagio (TR1) Avonex (TR2), Betaferon* and *Extavia (TR3), Copaxone (TR4), Cyclophosphamide (TR5), Gilenya (TR6), Lemtrada (TR7), Novantrone (TR8), Plegridy (TR9), Rebif (TR10), Rituximab (TR11), Tecfidera (TR12). Tysabri (TR13).* If several DMTs were supplied at the same time, the combination of these DMTs has been adopted.

Fig. 1 (study from 1975 to 2005) presents a total of 1,687 patients with a maximum of six DMT changes. If a patient does not reach those six changes, the last DMT given was considered for the following changes. Fig. 2 (study from 2005 to 2017) presents 521 patients. In this case, the maximum number of DMT changes is seven. As before, if those seven changes are not reached, the last DMT provided was taken in the following changes.

In Fig. 1, 24.42% of patients have at least one DMT change. However, in Fig. 2 the percentage increases to 71.4%.

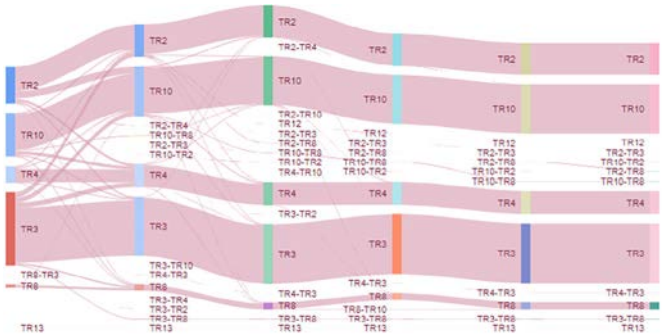


Fig. 1. Evolution of the treatment whose start date is between 1975 and 2005.

This increase in the DMT change coincides with the appearance of new treatments, like *Aubagio Gilenya, Lemtrada, Plegridy, Tecfidera* or *Tysabri* in the years of the second period.

Furthermore, with older treatments, there were more DMT changes and, when new ones appear, DMT changes are reduced. New DMTs are supplied for longer periods.

1 https://www.msbase.org/

Fig. 2. Evolution of the treatment whose start date is between 2005 and 2017. There are 521 patients who have seven changes in their DMT.

The most noticeable changes range from an old treatment to a new one. For instance, 39.6% of patients who start taking *Avonex (TR2)* and have a change, do so at a new DMT. The same occurs with *Betaferon* and *Extavia (TR3)*, increasing the percentage to 46.5% of patients.

In addition, Fig. 3 shows average solar radiation per country. It is an example of demographic variables in the patient that are included in the model.

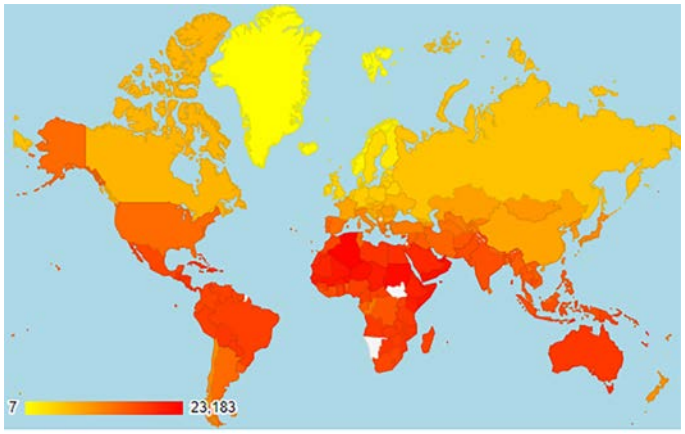


Fig. 3. Average solar radiation per country.

B. Prediction Model

A model was implemented to evaluate the response to fourteen DMTs using the demographic, clinical and paraclinical variables previously described.

MS behaviour was predicted, only patients whose course of the disease follows a CIS-RRMS-SPMS progression using the relapses stored in the clinical records. All information prior to the relapse, as well as the most recent treatment provided between the relapse and the time horizon to predict, was used. The case of not treating the patient is included as another predictable result.

As an additional requirement to those set during Data preparation described in subsection II.A, patient relapses were studied only if the following pre-relapse information exists:

- An EDSS record in the absence of relapses (previous 30 days).
- If cerebral or spinal cord type MRIs is present, the variables *number of lesions in T1 Gadolinium+* and *number of lesions in T2 fields* cannot be absent.

1. Prediction Outcomes

The response to a treatment is evaluated by the evolution of the patient's disability through EDSS. Therefore, this model forecast the progression of disability in two ranges, namely 1 and 2 years.

Progression of disability was defined as [10]:

- Increase of 1,5 or more steps in EDSS if previous EDSS was 0.
- Increase of 1 or more EDSS steps greater than or equal to 1 and less than or equal to 5.5.
- Increase of 0,5 or more steps in EDSS greater than or equal to 6.

Only the progression events confirmed over greater than or equal to 6 months (with the confirmation EDSS recorded greater than 30 days following previous relapses and irrespective of treatment status at the time of confirmation).

Progression of disability in 1 year was defined and indicated whether the patient has progressed in the EDSS in a year. Progression of the disability in 2 years indicates whether the patient has progressed in the EDSS in two years, assuming that the patient remains at the same level of EDSS during the first year.

In this case, only relapses that have an EDSS record after the first or second year were taken into account, confirming that the patient had not progressed in their level of disability. Otherwise, there was no follow-up of the patient and it could not be confirmed/discarded that the event occurred.

2. Machine Learning Techniques

The target was to predict a discrete label: *No progression* or *Progression* in EDSS. In other words, the estimation of the probability of an event either occurring or not. These algorithms provide a score of the probability of the event occurrence. The interest was also to obtain a probability score associated with the non-occurrence of the event. Moreover, a threshold was set to decide this non-occurrence (i.e. our model predicts *No progression* if the provided score is above the threshold).

a) Feature Selection

A Random Forest model was implemented to predict the progression in the level of disability in response to treatment in1 and 2 years. It is able to provide the importance of variables for classification. The 14 most relevant variables chosen are shown in Table I.

TABLE I. Most Important Variables Obtained with Random Forest

Random Forest	
gender	tab_rel_2
birth_date	num_rel
first_symptom_date	les_t1_gado
diagnosis_date	les_t2
date_last_visit	current_dmt
edss	time_ms_dmt_ini
tab_rel_1	time_current_dmt

The most relevant variables for prediction are the following:

- *Gender* is the patient gender, which is a factor type variable. Two binary variables were generated that indicate whether the gender is male or female.
- *Birth_date* is date of birth of the patient. The age of the patient is used in the model.
- *First_symptom_date* indicates the date on which the first symptom occurs. The age of the first symptom is used in the model.
- *Diagnosis_date* is the date of diagnosis of the disease. The age of diagnosis of the disease is used in the model.
- *Date_last_visit* is the date of the most recent visit.
- *Edss* is the last EDSS record in the absence of relapses. This situation occurs when there is no previous relapse in the 30 days prior to the visit.
- *Tab_rel_1* is the annualized relapses rate in last year. This variable corresponds to the number of relapses in last year.

- *Tab_rel_2* is the annualized relapses rate in last two years, which corresponds to the number of relapses in the last two years divided by two.
- *Num_rel* is the total number of relapses.
- *Les_t1_gado* is the number of lesions in the last RMI of the brain, spinal cord or cervical cord in T1 Gadolinium+ (relaxation time measured by MRI). If RMIs are not simultaneous, only the last RMI of the brain is taken into account.
- *Les_t2* is the estimated number of lesions in the last RMI of the brain, spinal cord or cervical cord in T2 (relaxation times measured by MRI). If RMIs were not simultaneous, only the last RMI of the brain is taken into account.
- *Current_dmt* is the treatment currently prescribed. It is a factor type variable, with fourteen factors corresponding to treatments described.
- *Time_ms_dmt_ini* is the time elapsed between the date of diagnosis and the start of the first prescribed DMT.
- *Time_current_dmt* is the time elapsed between the start date of currently prescribed DMT and the date of last visit.

b) Classification Algorithm

To build the prediction model, the algorithm chosen was also Random Forest.

Random Forests

Random forests are ensemble learning methods that operate by constructing a multitude of small decision trees at training time and outputting the class that is the mode of the classes of the individual trees. Random forest is considered one of the best performing algorithms, especially for problems that have many explanatory variables [11].

The prediction model was trained using 70% of the train data, with the conditions and requirements previously explained. The remaining 30% was used to evaluate the accuracy of the model (test) in predicting the evolution of the patient for the treatment provided from the outbreak and compared to the actual evolution.

III. Results and Discussion

In this section, the evaluation of the model on the set of test patients is explained, considering the accuracy and validity of the results for different thresholds.

A. Results

In order to validate the model, the Receiver Operating Characteristic (ROC) curve for both years was first performed (Fig. 4), obtaining an area under the curve (AUC) of 0.8 and 0.82 respectively.

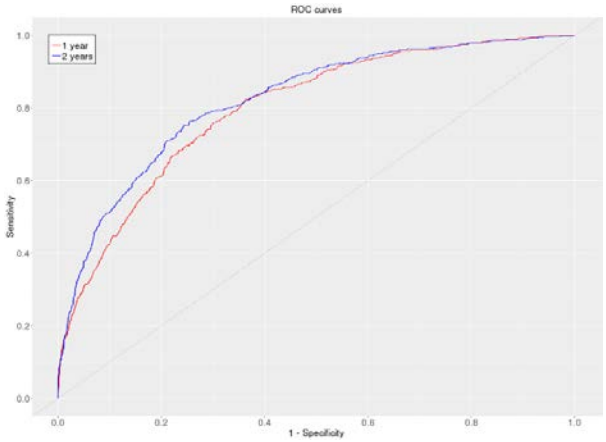


Fig. 4. ROC curve for the first and second year models.

In a ROC curve the true positive rate (Sensitivity) is plotted in function of the false positive rate (1-Specificity) for different cut-off points. Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold. A test with perfect discrimination (no overlap in the two distributions) has a ROC curve that passes through the upper left corner (100% sensitivity, 100% specificity). Therefore, the closer the ROC curve is to the upper left corner, the higher the overall accuracy of the test [12].

In addition, a measure of the error in terms of the patient's non-progression was defined. That way, the probability the event non-occurrence with a treatment.

For a given threshold, the error metric used (see eq. 1) indicates the proportion of patients who do not actually progress from the set of patients that the model predicts that would not progress.

$$Accuracy_{\text{threshold}} = \frac{N(score>\text{threshold})_{No\ progression}}{N(score>\text{threshold})} \quad (1)$$

A score above the established threshold ensures, with a probability equal to the accuracy of the model, that the patient will not progress in EDSS.

It was found that 76% of patients in the data set do not progress in the first year and in the second year that proportion is reduced to 67%.

Fig. 5 shows the accuracy of models for 1 year and 2 years, and the comparison with the option to choose always non-progression. The model for second year (in blue) presents a greater difference with respect to the systematic choice of non-progression (blue dashed line).

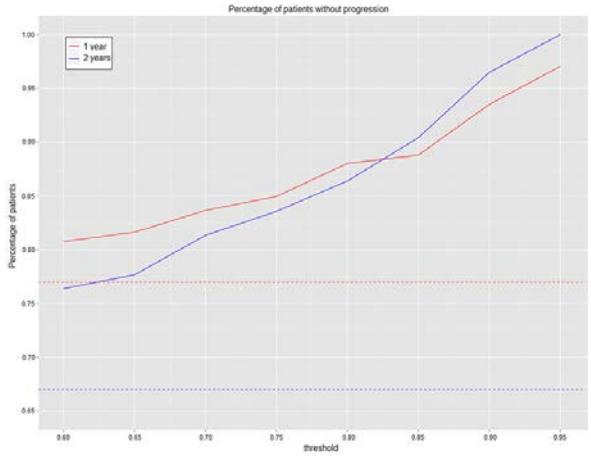


Fig. 5. Percentage of patients who do not actually progress from the set of patients that the model says do not progress for each of the established thresholds. Horizontal dashed lines represent the proportion of patients in the set who do not progress to 1 year (red dashed line) and 2 years (blue dashed line).

As the threshold increases, the accuracy with which the model predicts that a patient will not progress increases considerably.

B. Functionalities

A web application called Model MS was built to support the specialist neurologist with the results of the model. This tool is one more component in the study process of a patient to decide what treatment to prescribe. Please note, however, that it is ultimately the doctor's decision to determine the treatment based on their experience and direct contact with the patient. Model MS is just a support tool.

1. Input Data

Fig. 6 shows the main screen of the tool, where the expert can: a) consult the minimum characteristics used for each patient (and previously described in Table I), b) search patients, c) upload patient data from an Excel file extracted directly from the platform for medical

software systems ©iMed [13] or d) enter these minimum data manually. In this screen, it is also possible to change language and view a short description of the tool.

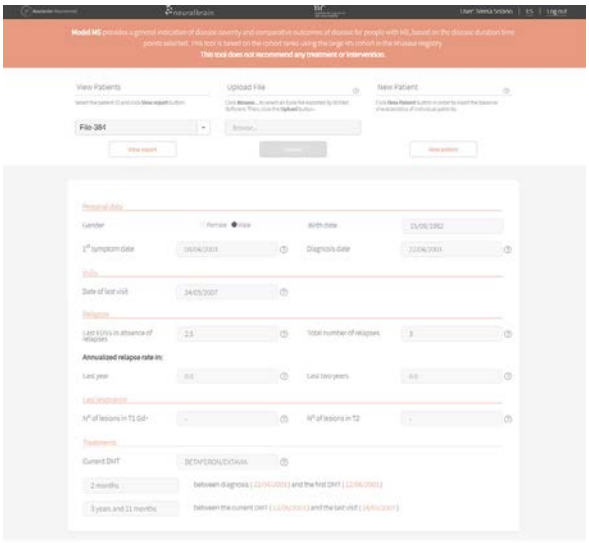


Fig. 6. Tool input data. Minimum set of fourteen characteristics used for each patient in the predictive model.

2. Output Data

Model MS presents three graphical and textual outputs, which are, respectively, a prediction of progression, evolution of the patient, and a description of the patient study that helps the expert.

a) Prediction of Progression

On the screen shown in Fig. 7, experts can view the progression forecast in EDSS from the last relapse. The predictive model uses the patient's current conditions (most relevant variables) and provides a score for each of fourteen treatments (described in subsection II.A), including the possibility of not treating the patient. The higher the score, the greater the likelihood that the patient will not progress if that treatment is prescribed.

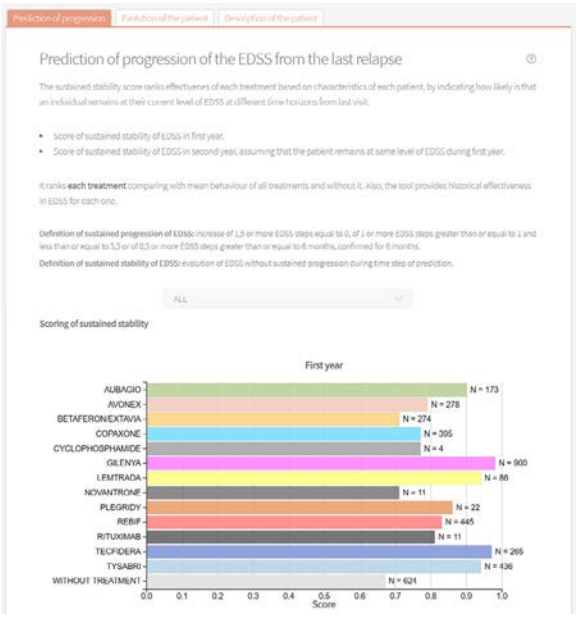


Fig. 7. Progression forecast. Score for each of the fourteen treatments that can be provided to patient. The higher the score, the greater the likelihood that the specific patient will not progress if that treatment is prescribed.

b) Evolution of the Patient

In Fig. 8 treatments prescribed to patients and evolution of course of them during MS duration are shown. The age of the patient when first symptom occurred, and MRIs are provided. The lower part shows the evolution of the patient's level of EDSS along with the visits and relapses.

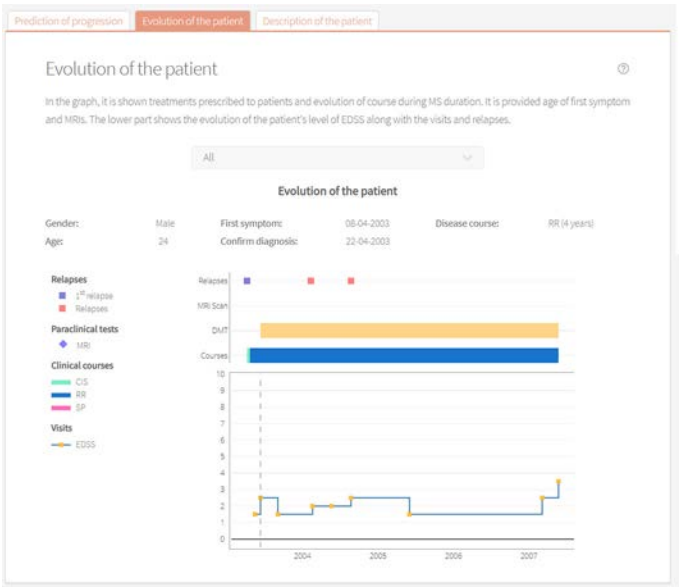


Fig. 8. Evolution of the patient. Upper part shows the first relapse (purple square) together with the rest of the relapses (red square) that the patient has suffered throughout the disease and the magnetic resonances performed (purple rhombus). Lower part shows the progression of the patient's disability (blue line) together with the visits made (yellow square).

c) Description of the Patient

In the screen shown in Fig. 9, a report is generated for the medical specialist, as a textual summary of what has been shown in the two previous screens. This report can be downloaded by the specialist.

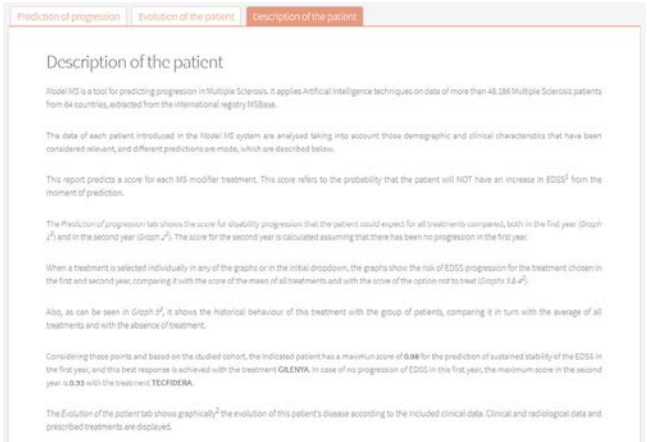


Fig. 9. Description of the patient. It is a textual summary of the two previous screens.

IV. Conclusion and Future Work

In this paper a system has been designed with a prediction model that extracts implicit knowledge out of data.

This work, and the use of the Model MS tool, has several advantages:

On one hand, it is possible to study the evolution of the disease from the different available variables (types of MS, geographical area, date of birth, gender...) and to identify several clinical factors relevant to its progression. Fourteen different DMTs and their effectiveness according to the different groups of patients analysed were studied.

On the other hand, a personalized prediction model for each MS patient, applying Machine Learning and Big Data techniques, has been developed. The resulting tool provides a probabilistic estimate of disease progression for different time horizons, and indicates the statistical conclusion of the probable impact of each treatment on the patient's evolution.

In addition, it has been verified from experimental results that not treating a patient is the worst option. The model always predicts a significantly lower score than any of the DMTs.

Finally, it has been verified that each patient presents a different evolution for each of treatments depending on their clinical state. There is no common behaviour for each DMT and it is normal that there are significant differences between the scores of the model in each one, but each patient has specific treatments that are better suited to their current clinical state. The best DMTs are those with the highest score, i.e. those in which it is most likely that the patient will not progress in EDSS.

The implemented tool has proved to be of interest and useful to neuroscientists in different countries. Model MS is being used in real time clinical practice by medical experts, to improve their results. A new line of work has been created to apply this methodology to many other branches of Medicine.

Acknowledgment

The authors are extremely grateful to *Hospital Quirónsalud and Hospital Rúber Juan Bravo in Madrid*, *Hospital Vithas NISA in Seville* and *Instituto de Ingeniería del Conocimiento, Madrid*. In addition, we would like to thank *MSBase* for their contribution to this project.

References

[1] C. Lucchinetti, M. Rodriguez, and B. G. Weinshenker, "Multiple sclerosis," *N Engl J Med*, vol. 343, pp. 938–952, 2000.

[2] G. Giovannoni, "Conversion from clinically isolated syndrome to multiple sclerosis: A large multicentre study," 2015.

[3] R. D'Alessandro et al., "Risk of multiple sclerosis following clinically isolated syndrome: a 4-year prospective study," *Journal of neurology*, vol. 260, no. 6, pp. 1583–1593, 2013.

[4] L. Kappos et al., "Factors influencing long-term outcomes in relapsing–remitting multiple sclerosis: PRISMS-15," *J Neurol Neurosurg Psychiatry*, p. jnnp–2014, 2015.

[5] K. Baumstarck et al., "Health-related quality of life as an independent predictor of long-term disability for patients with relapsing–remitting multiple sclerosis," *European Journal of Neurology*, vol. 20, no. 6, pp. 907–e79, 2013.

[6] M. Tintore et al., "Defining high, medium and low impact prognostic factors for developing multiple sclerosis," *Brain*, vol. 138, no. 7, pp. 1863–1874, 2015.

[7] T. Spelman et al., "Risk of early relapse following the switch from injectables to oral agents for multiple sclerosis," *European journal of neurology*, vol. 23, no. 4, pp. 729–736, 2016.

[8] H. Butzkueven et al., "MSBase: an international, online registry and platform for collaborative outcomes research in multiple sclerosis," *Mult Scler*, vol. 12, no. 6, pp. 769–774, Nov. 2006.

[9] C. Chruzander et al., "A 10-year follow-up of a population-based study of people with multiple sclerosis in Stockholm, Sweden: changes in disability and the value of different factors in predicting disability and mortality," *Journal of the neurological sciences*, vol. 332, no. 1–2, pp. 121–127, 2013.

[10] T. Kalincik et al., "Towards personalized therapy for multiple sclerosis: prediction of individual treatment response," *Brain*, vol. 140, no. 9, pp. 2426–2443, 2017.

[11] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[12] F. Schoonjans, "ROC curve analysis with MedCalc," *MedCalc*. [Online]. Available: https://www.medcalc.org/manual/roc-curves.php.

[13] "iMed Software." [Online]. Available: http://www.imed.org/content/healthcare/biopharma/iMed/en/home.html.

Cristina Pruenza García-Hinojosa

She has a Degree in Mathematics and Computer Engineering from Universidad Autónoma de Madrid (2009-2014) and she has got a Master's Degree in Computer Engineering and a Master's Degree in ICT Research and Innovation in Computational Intelligence at the Polytechnic School, Universidad Autónoma de Madrid (2014-2016). She is currently Data Scientist and Technical Leader at the Instituto de Ingeniería del Conocimiento (IIC-UAM) in the Health and Energy Predictive Analytics area, where she is currently working in the following innovation lines: Big Data, modelling, data analysis, evaluation of wind production forecasting, e-Health among others. Pruenza García-Hinojosa has participated in several projects related to e-Health and in the ADNI project research on Alzheimer's and some possible tools to treat MRIs.

María Teresa Solano Albares

María Teresa Solano Albares is studying a Degree in Mathematics and Computer Engineering from Universidad Rey Juan Carlos de Madrid. Currently, she is Data Scientist in the Health and Energy Predictive Analytics area at the Instituto de Ingeniería del Conocimiento (IIC-UAM) with experience in projects related to solar prediction, study of the evolution of degenerative diseases, among others. Previously, she has also collaborated with the Social Business Analytics team improving a tool for monitoring and analysis of social networks in real time.

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.

Rafael Arroyo González

Rafael Arroyo González obtained his MD cum laude from Universidad Complutense de Madrid. He is currently the head of the Neurology department at Hospital Universitario Quirónsalud Madrid and Hospital Ruber Juan Bravo, where for over 10 years he has been leading all clinical, teaching and research activities. He is a senior professor in Neurology accredited by ANECA at Universidad Europea de Madrid and possesses a Master's degree in Medical direction and clinical management from UNED. He has published 134 papers in renowned international journals; as a principal investigator, he has participated in well over 90 clinical trials in multiple sclerosis, Alzheimer's disease and other neurodegenerative diseases.

Guillermo Izquierdo Ayuso

Guillermo Izquierdo Ayuso obtained his MD from Universidad Complutense de Madrid and his PhD from Universidad de Málaga. He is currently the head of the MS department at Hospital Vithas Nisa Sevilla. International MS expert, for over 30 years he has been leading clinical, teaching and research activities. In the last five years, e has published 167 papers in renowned international journals; as a principal investigator, he has participated in well over 90 clinical trials in multiple sclerosis and other neurological diseases.

MOBEEZE. Natural Interaction Technologies, Virtual Reality and Artificial Intelligence for Gait Disorders Analysis and Rehabilitation in Patients with Parkinson's Disease

Lombardo, J.M.[1]*, Lopez, M.A.[1], Lopez, M.[1], León, M.[1], Miron, F.[1], Arranbarri, J.[2], Alvarez, D.[1]

[1] FIDESOL (Free Software R&D Foundation) (Spain)
[2] Virtualware Group (Spain)

Received 23 April 2019 | Accepted 30 April 2019 | Published 9 July 2019

Abstract

Parkinson's Disease (PD) is the most common degenerative disorder after Alzheimer's disease. Generally affecting elderly groups, it has a strong limiting effect on physical functioning and performance of roles, vitality and general perception of health. Since the disease is progressive, the patient knows he's going to get worse. The deterioration is significant not only in mobility but also in pain, social isolation, and emotional reactions. Freezing is a phenomenon associated with this disease and it is characterized by a motor disorder that leaves the patient literally stuck to the ground. Mobeeze is designed with the main objective of providing health personnel with a tool to analyse, evaluate and monitor the progress of patients' disorders as well as the personalization and adaptation of rehabilitation sessions in patients with Parkinson's disease. Based on the characteristics measured in real time which will allow the strengthening effects of rehabilitation and help to assimilate them in the long term. The creation of Mobeeze allows the constitution of a system of analysis and evaluation of march disorders in real time, through natural interaction, virtual reality and artificial intelligence. In this project, we will analyse if these non-invasive technologies reduce the stress induced to the patient when he is feeling evaluated.

Keywords

Parkinson, Freezing, Natural Interaction, Virtual Reality, Artificial Intelligence, Cognitive Computing.

DOI: 10.9781/ijimai.2019.07.003

I. Introduction

Available data concerning Parkinson Disease (PD) indicates prevalence over other neurodegenerative diseases with relation to ageing and treatment complexity. The disease is estimated to increase in the next few years on a global level making it the second most prevalent neuro-digestive disease right after Alzheimer's disease.

Parkinson's Foundation in the United States, where more than 60 thousand cases are diagnosed every year, estimates that 1 million people are affected by Parkinson only in this country and more than 10 million are affected worldwide.

The European Parkinson's Disease Association (EPDA) gave a lower estimate, pointing out that there are 6.3 million persons affected by PD worldwide; nevertheless, the number is equally disturbing. In Spain alone, the numbers are estimated between 100 to 150 cases of persons suffering from Parkinson's Disease. According to the Spanish Brain Foundation like EPDA, diagnosing today 10 thousand new cases each year, the number that is intended to double in the next 20 years and triple in the year 2050. This confirms the evaluation realized by the Spanish Society of Neurology.

More than one third of the patients with PD (mostly the ones in advanced stages of the disease) suffer from a well-known phenomenon known as "freezing". This is a temporary motor disorder that lasts seconds and is characterized by the sensation of "feet glued to the floor". It is mostly felt when the patient tries to start or resume his or her walk, turning to change direction or pace of the walk. It can also appear by just passing through a door or through narrow spaces like corridors, also specially appearing in stress episodes or with distractions. This may cause uncontrolled falls.

In the last few years giant leaps have been taken to better the standard of living of the patients of neurological diseases and related disorders such as PD, but science goes on. Concerning patient's rehabilitation, therapies are centred on minimizing and/or delaying the effects of these disorders. These therapies normally derive in optimal results though not all patients can have access to them.

A. Parkinson Patient Treatment Actual Methods

As of today, PD doesn't have any cure. However, its symptomatology can be controlled in an adequate manner [1]. Researchers are presently working on the improvement of the symptoms associated with the neural loss and neural death from several points of view:

- **Pharmacological**: There's a whole set of medications that favour the dopamine production or delay its deterioration. It's administrated according to the gravity of symptoms. Less powerful drugs are administrated in the first stages followed by a progressive dosage increase.
- **Surgical**: Procedures centred on the intervention of the damaged part of the brain are only recommended in a very low percentage of patients, not more than 5%. Patients benefited by this are

* Corresponding author.
E-mail address: jmlombardo@fidesol.org

only the ones with a very advanced PD case or null response to pharmacological treatment.

- **Subthalamotomy**: Elimination of the affected brain area and installation of a local pacemaker. This technique is also performed in cases of very advance disease or null response to pharmacological treatment.
- **Physical Rehabilitation**: A key aspect of living with the disease is the maintenance of muscle tone and motor functions, so daily physical activity supplemented by exercises to help maintain mobility in affected members is essential.

As regards to physical rehabilitation, there is not always a possibility for patients to carry out this therapy, so scenarios based on information technology is suggested as a possible solution.

B. ICT Applied to the Study and Treatment of PD by MOBEEZE

In recent years, the Information and Communications Technology (ICT) have contributed to solutions aimed at improving the healthcare quality which allows one to address the pathology through specialised systems in order to ensure a more focused attention on the patient.

It is suggested that the use of ICT for the treatment of PD could help to rehabilitate motor abilities, although further research to establish its clinical effectiveness and safety is necessary. In order to develop Mobeeze, there were taken into consideration various complementary technologies to address patient evaluation: (1) Virtual reality system (VR), in this case, VR goggles for training in immersive environments, (2) Natural interaction devices for analysing all variables related to patients' motion and manual dexterity. A three-dimensional virtual world recreation environment compatible with such devices is used. Information generated through training with these devices provides data such as the step length, speed, angle of both knees and hips, etc. Thanks to techniques of artificial intelligence (AI), there is new knowledge of medical research regarding the suitability of the application of these technologies.

The key elements for applying the VR neurorehabilitation are **repetition**, **feedback** and **patient's motivation**. Repetition is important for motor learning and neuroplastic changes to occur. These changes maintain long-term lessons learned but must be linked to a sensory feedback in the outcome of each of the embodiments. On the other hand, in order to perform time and again the activities required for neurorehabilitation, the subject motivation is fundamental. This is achieved by focusing the different activities as a video game so that the treatment sessions are much more enjoyable and attractive [2]. The use of VR as a supplemental to therapy includes other important advantages such as the ability to control each session accurately and repeatedly along with the ability to adapt the interfaces to the users' motor limitations and the recreation of secure virtual environments for practicing skills that would otherwise imply potential risk in the real world.

Despite all these advantages, there are some doubts about applying VR to neurorehabilitation, especially in regard to the equivalence between the movements in real and virtual environments. That is: the resulting functional capacity. Several studies have reached to the conclusion of relative equivalence between environments, although small deviations were accounted for due to different spatial perception [3]-[5].

In recent years, studies have been published on various neurological diseases in which the VR is applied: cerebrovascular disease [6], multiple sclerosis [7], in patients with sequelae after traumatic brain injury [8] and PD, which is the object of the current project.

Additionally, natural interaction devices (NI) can further enhance the immersion experience, allowing the gap with reality to become even lower. Some studies suggest that neurorehabilitation VR / NI could induce cortical reorganization which plays a key role in the recovery of motor [9] capacity and proves the transfer of acquired skills to the real world [10], which makes the VR itself a tool of great potential in neurorehabilitation.

II. Materials

A. System Virtual Immersion by Oculus Rift

Oculus Rift (OR) is a VR device developed and marketed by Oculus VR company, a subsidiary of Facebook Inc., which launched to the market on late March 2016.

It uses an OLED panel for each eye, and technical specifications for the projection image are perfectly designed to create a three-dimensional world where an immersive user experience is complete. It relies on two hardware devices, controllers to interact with objects in a three-dimensional environment and satellites that pinpoint the physical position of the user in order to interact with objects in three-dimensional environment.

1. Virtual Immersion in the eHealth Sector.

In recent years, interest in the sector has increased due to the inclusion of this technology in therapies and interventions [11]. OR is growing popular in health centres and even private homes, while simplifying the development of its applications. Thanks to this, it can be considered to replace traditional therapies with novel and immersive ones, even both ambulatory and domiciliary treatments.

However, an important challenge for VR treatments is the need for more evidence to show that the benefits obtained from such treatments translate into real benefits.

B. Natural Interaction System by Leap Motion

Leap Motion (LP) is a sensor that draws a virtual image of hands and joints from the wrists to the distal phalanges of the fingers and also represents the image of hands on a computer screen recording these movements. LP is used in the project for patient manual ability evolution accounting for the freezing phenomenon.

C. Monitoring System Activity by Kinect

Although no longer marketed in 2017, following its launch in 2010, the Kinect device was a breakthrough in the use and application of technologies for natural interaction in the field of medicine and study procedures, monitoring and rehabilitation was conducted with patients, as is the case of Parkinson's patients. The introduction of Kinect in this area gave these therapies a low cost device that, thanks to its processing capacity of actions and obtaining information about them, along with the inclusion of a VR system, made possible the Immersion binding a Natural Virtual-Interface, a combination that would improve the patient's immersive experience.

Thus, different studies have sought to demonstrate the optimization of results in tests performed in a controlled healthcare environment for which the experience was designed [12] using the Kinect device. The aim of Mobeeze was to determine the therapeutic effectiveness of VR in patients with PD using Kinect as an instrument for the monitoring of the freezing phenomenon.

Various research groups have also oriented part of their research to test the efficiency of Kinect and the application of VR technology to monitor and rehabilitate of PD patients. In the case of Brunel University in London, where a positive application of the Kinect device was accounted for, orienting in the same manner their research for monitoring purposes and helping people affected by this phenomenon of motion disturbance, getting satisfactory results and conclusions [13] [14]. Likewise, the Cochrane Movement Disorders Group of

the University of Lisbon conducted an independent research on VR exercise interventions for rehabilitation in PD [15].

In the case of the Mobeeze project, in which one of the goals is to identify the characteristic signs of PD motor deterioration relative to the patient's stage, Kinect was used as an affordable, reliable and validated tool for motion analysis.

D. Interactive Motor of Cognitive Inference

Cognitive computing can be considered, although not all researchers agree [16], emerged as a specific branch of AI, focusing on providing the machine a component of "understanding" that emulates the workings of the human mind. It could be defined based on a set of biologically inspired technologies, as well as those that mimic (and overcome) intelligent animal functions and easing interaction between humans and machines. IBM Watson [17] is one of the earlier examples.

Cognitive computing difficult to discern from AI itself, but rather a new way of naming it. It is not easy to differentiate between a good intelligent software and AI. John McCarthy, founder of AI, around 1955 said, "Once a program does something that previously only humans did, we stop calling it artificial intelligence." Speech recognition is a good example for this, as well as the process of written text conversion, automatic translation or the ability to dictate to a computer. Five years ago, it was an unsolvable problem. Laboratories around the world were investigating it for decades. It is present today in our cell phones and nobody is surprised nor calling it artificial intelligence despite being a colossal achievement.

We tend to think of AI as a gigantic being that performs multiple human functions. However, cognitive computing is quite different: it is small and it's everywhere. Its power lies in its ubiquity. Along with these functionalities: image and face recognition. But there is much more, for example personality traits detection through social networks or message tone analysis.

Simplification or humanization of machine interaction machines is challenging. It's difficult to accept that in order to deal with a machine human have to understand their primitive system commands. The aim is that the machines understand humans and not vice versa.

In recent years, AI is gaining ground in terms of diagnostic procedures, monitoring, analysis as well as evaluation of patients. Concerning PD, for the Michael J. Fox Foundation (named after and created by the famous actor, whose mission is to find a cure for the disease and develop improved therapies), AI strategies are priority in finding treatments and a cure for PD. Thus, in 2013, the foundation launched a challenge data in order to generate a brainstorming in the development of solutions for monitoring and treatment of PD [18]. As part of this challenge, the proposal winner planted an approach to machine learning to address the problem, demonstrating the viability and value of collecting and analysing cell phone data to monitor Parkinson. This was obtained through cell phones incorporation and low-cost sensors.

On the other hand, and through some funds granted by the same foundation, IBM developed a prototype consisting in a small sensor nail for obtaining data, which could allow specialists to track, monitor and diagnose the movement disorders in people with PD. The extracted data are studied by AI algorithms that are taught a series of patterns [19].

The Parkinson and Movement Disorders Unit at the Polytechnic University of Catalonia together with Telefónica I+D, developed a system for monitoring patients with Parkinson called Rempark, which uses AI models to remotely control the patient, monitoring in order to help and manage their own mobility. Among its components, it has sensors to identify, in real time, freezing of gait and a guidance system of travel that provides auditory cues. The information obtained from the patient through the sensors and devices with embedded enables the specialist to receive accurate and relevant information from the patient in order to make decisions or treatment information.

Research on the design of AI models that provide relevant data and information on patients with PD have not only been aimed at monitoring and/or rehabilitation patients, but also for early detection and improved diagnoses of neurodegenerative diseases, especially Parkinson's disease. In order to improve the accuracy of interpretations and evaluations made by specialists, Mobeeze is using inference engines data, designed to provide the researcher with a knowledge basis by training patients in virtual reality environments and natural interaction, providing a new perspective and new data study.

E. Environments Recreation of Virtual Worlds

For the recreation of scenarios contemplated in the project, it is necessary to have a platform for creating three-dimensional virtual worlds that can also be compatible with the use of virtual reality goggles and natural interaction elements. This article is not intended to make a study of the possibilities, especially free software, since it said that the analysis was performed on another project with similar technical characteristics, Practica [20], but oriented to create a platform for training specialty in immersive environments to improve productivity. Therefore, the selected tool was Unity 3D development environment applications and video games widely accepted by the community 2D/3D environments. This framework supports a lot of peripherals, in particular those used in Mobeeze (Oculus Rift, Leap Motion and Kinect). Although Unity 3D is not a free software license, it allows free access to the community and is enriched with a repository of extensions provided by users of this technology, where you can find all kinds of resources (free or paid) high quality. It also supports the export of software to multiple platforms such as Android, Windows, iOS and even consoles like PlayStation.

III. Methodology

Mobeeze arises to experiment and analyse the viability of different technologies applied to carry out tracking people with PD, so it allows medical personnel to have a system that tracks and determines the degree of progress of the phenomenon of freezing. It also aims to support the research by incorporating a trained cognitive engine, taking as reference the historical training conducted by Mobeeze. Using this tool, we intend to conduct a study and careful monitoring of the patient so that training can be adjusted depending on the progress of this disorder in the patient making the evolution of the disease as slow as possible and therefore offering the patient a better quality of life. To this end, it has developed a tool for recreating situations as close to reality as possible using not only visual technologies based on VR, but also acoustic tech, which is tested on the patient in certain "stress" situations. The reactions are collected and stored so that, through a series of sessions, one can analyse the evolution of the patient. The configuration of the situations in which you can find a patient has been determined to be opposite to them in the closest form to real situations, increasing or decreasing the level of stress to which the patient undergoes. Thanks to the recreation of virtual reality environments, exposure to situations of high stress level that may not be viable in a real environment is allowed, as is the case in situations where a car crosses in front of the patient.

The training of these tests based on virtual reality technologies and natural interaction generates valuable information to track both during different sessions to further analysis. Based on this, the own patient's history and the history of workouts that were previously performed in other patients, healthcare workers can access the inference by a

cognitive motor supplied with such information, including the data of the patient. To this end, it is designed in a similar way to a virtual assistant who, through a query interface-based chat, provides valuable information to adjust researcher training and determine the evolution of the disease component. It should be borne in mind, while incorporating this component at the end of the project course, it has not been possible to validate the results of using this cognitive inference engine and query interface by medical researchers participating in the project. However, we are currently working on determining the effectiveness of this component.

Therefore, Mobeeze consists of several phases: (1) acquisition of patient data and determining the training; (2) creation of virtual reality system and natural interaction to measure the performance of patients in the environment created by the system, (3) cognitive inference engine together with (4) a query interface.

A. Acquisition of Patient Data

This project created on one hand, an observational study to analyse the population characteristics and motor signs and not dependent engines disease was crated and on the other hand, two subsequent experimental studies in order to evaluate the effectiveness of the rehabilitation program based virtual reality to optimize the way in study population study:

1. Observational Study

The evaluation protocol for the observational study consisted of three parts, dedicating five to ten minutes per patient:

a) **Initial interview** through which data and pertinent information on the characteristics of the participants who took part in the study were collected.

b) **Evaluation of motor signs**, which had the following appraisals: gait, manual dexterity, gait freezing, manual freezing evaluation, freezing in a favourable virtual immersion environment, freezing in a freezing provoking virtual immersion environment.

c) **Evaluation of non-motor signs** applying a battery of scales and questionnaires to evaluate: pain intensity, pain frequency, standard of living, fear of falling, anxiety and depression.

Data acquisition of the patient involved the hardware components and also the development of a special purpose software programmed to collect data from the distinct sensors and devices for analysis.

2. First Immersive Virtual Environment-Based Study

a) **Primary:** evaluation of the effectiveness of a rehabilitation program based on virtual immersion over the patients' gait with PD on a short term.

b) **Secondary:** evaluation of the effectiveness of a rehabilitation program based on virtual immersion over functionality about patients with PD on a short term.

3. Second Immersive Virtual Environment-Based Study

a) **Primary:** evaluation of the effectiveness of a rehabilitation program based on virtual immersion over short term PD patients' gait.

b) **Secondary:** To evaluate the effectiveness of a rehabilitation program based on virtual immersion on functionality about patients with PD on a short term.

Table I describes the variables used in the above described studies.

TABLE I. Variables Studied in Both Traditional Observational Study, and Virtual Immersion

Variables	Study 1	Study 2	Study 3
Motor variables			
Gait	٧	٧	٧
Stride length	٧	٧	٧
Stride height	٧	٧	٧
Stride width	٧	٧	٧
Stride rhythm	٧	٧	٧
Stride frequency	٧	٧	٧
Arm swing	٧	٧	٧
Functionality	٧	٧	٧
Freezing	٧	٧	٧
Gait Freezing	٧	٧	٧
Manual freezing	٧	٧	٧
Freezing in a virtual favourable environment	٧	٧	٧
Freezing in a virtual provocative environment	٧	٧	٧
Balance	٧	٧	٧
Non-motor variables			
Pain intensity	٧	٧	٧
Pain Frequency	٧	٧	٧
Standard of living	٧	٧	٧
Fear of falling	٧	٧	٧
Anxiety	٧	٧	٧
Depression	٧	٧	٧
Sleep disturbance	٧		

Experimental studies with virtual immersion took about five weeks with over 15 sessions, whose protocol is as follows:

- Week 1-2: Oculus Rift plus in a sitting exercise.
- Week 3: Oculus Rift plus standing exercise.
- Week 4: Oculus Rift, standing exercise plus gait exercise.
- Week 5: Oculus Rift, standing exercise plus gait exercise with obstacles.

Without an in depth analysis on the protocol applied by the medical research personnel [21], a number of comparative analysis of the studies over two different population groups were performed, one applying the observational study and other implementing virtual immersion techniques, both in a favourable immersive environment (with both visual and acoustic stimuli, for example: an open space free of obstacles) and the other being a challenging environment (with both visual and acoustic stimuli that may cause stress in the patient).

B. VR/NI

To evaluate the freezing phenomenon through patient training using VR/NI systems, special software components, akin to these technologies have been developed. In order to achieve this goal, the features of the environment should meet at least one of the following requirements:

- **Easy to control:** intuitive control with simple movements or gestures. The system returns a response in accordance with the programmed activity, once it detects the movements thus creates a visual feedback for the environment management.
- **Attractive/Motivating:** development of a training application to be installed on a personal computer and whose handling does not require advanced computer skills.

Hence, three software components have been developed, directly related to the training and monitoring of patients: (1) editor, used by

researchers for editor layout; (2) environment launcher, wherein the stage is configured to increase or decrease the stress provocation and (3) a monitoring module, for further analysis of the information generated during training.

The scenarios built with the environment editor are intended to evaluate the variables listed above applied to gait, manual dexterity and freezing:

- **Gait evaluation:** the VR, Oculus Rift goggles, NI device and Kinect were used for that purpose, to accomplish analysis of all the variables related to the gait.

  Kinect includes an infrared light grid and a video camera capable of generating a 3D map of the patient's physiognomy, albeit with limited accuracy. However, the depth sensor is capable of capturing the patient's position and submitting the information to the database, delivering professionals a wide variety of information regarding postural control, thus endorsing the potential of this tool to evaluate posture in a clinical environment.
- **Manual dexterity:** Finger pinch exercise is used to evaluate this. The participant is to try catch with his dominant hand a certain object between his index and thumb fingers, performing variations of speed and pressure. This exercise can be parameterized with multiple levels of speed. For this, a developed component using the Leap Motion device was used, which controls the movements of the hand with sufficient precision to determine the movement variations.
- **Freezing:** for the evaluation of freezing, Oculus Rift, Kinect and Leap Motion were used. To do this, they have carried out the tasks called "Stepping in place task", "Tapping task" and balance training.
  - **Stepping in place task:** is a task developed by Nantel et al. [22] to provoke and evaluate freezing and the progress of patients with PD with a 93% specificity and 87% sensitivity respectively for freezing detection. The task involves the patient, in standing position, raises his knees alternately remaining in the same spot. The task begins with the patient standing and after ten seconds, an audio signal is used to indicate the patient to start raising knees alternately for a hundred seconds.
  - **Tapping Task:** has been used with various motifs throughout history. In this research study this task is used to evaluate the manual freezing according to Ziv et al. [23]. The task consists on the participant having to move only the index finger flexion and extension of the metacarpophalangeal joint as quickly as possible for fifteen seconds.
  - **Balance:** Rated by Tinetti Scale [24]. It is a hetero-administered questionnaire consisting on 9 aspects related to equilibrium and 7 related to gait. The score is rates 0 for subjects who fail to or just maintain balance, one for those performing the movement with compensation and 2 for those acquiring the correct position without difficulty. The maximum score is 16 for balance and 12 for gait, the combined of both is a total score of 28. It's considered that between 19-24 risk for fall is minimal, whereas <19 implies a high risk of fall.

Medical researchers were in charge of patients' personal data management as part of their studies which has been done according to compliance with the Level 3 of LOPD (Spanish Organic Law of Data Protection): patient information is for exclusive use of the research team and anyone outside the project can't access such information, except the Ministry of Economy and Competitiveness, founders of the study and Ethics Committee for Clinical Research of HULP, or any other competent authority in the matter and that requires at any time.

Patient's information concerning the interviews (patients' assigned code, gender, age, and years of evolution of PD medication since the last time) has also been stored. Several questions regarding the criteria for inclusion and exclusion and Mini-mental State Examination were also included. Scores were collected from the following evaluation scales of non-motor signs: EVA and frequency of pain. All remaining results of evaluation of non-motor signs were recorded in the questionnaire completed by the patient himself. Measurements of motor signs, except for the valuation of balance, were all digitally recorded in the database. The balance evaluation was recorded in the own Tinetti scale.

The information used for cognitive inference has been made anonymous, impeding any information disclosure through informatic data leakage or any other manner, according to LOPD.

C. Cognitive Motor

In a final phase of the project it was included as an additional component of special interest: a cognitive inference engine, so that data taken during the course of the exercises with patients stored could be classified, easier to use and distribute through scientific community for analysis. The goal of integrating a cognitive engine is to provide the platform for a standardised basis for measures to allow comparison of the variables taken as well as the different factors that impact exercises and allow the system to:

- Obtain correlations that could have been overlooked by the medical research personnel.
- Allow communication and information exchange between entities, allowing different hospitals and organisations the location and analysis of the data obtained.
- Extrapolate the results corresponding to different pathologies, extending the analysis and diversifying the platform and original study.
- Write reports based upon any existing measure on the system.

As a basis for the analysis of information collected in the training of patients, the results of the application of descriptive statistical methods on socio-demographic, motor and non-motor variables were used. For confirmation of normal distribution of the data the Kolmogorov-Smirnoff test was applied.

The analysis of variance (ANOVA) approach is used to analyse each stage group factor as PD. The normal distribution of the data was calculated using the Kolmogorov-Smirnov test: most of the variables correlated to a Gaussian distribution. The descriptive results of the study were represented by statistical frequency and percentages for qualitative variables and media with its standard deviation (SD) for the quantitative variables.

Both records of the variables studied during training and statistical information were introduced for creating rules and evaluation and perform the inference of knowledge, in order to:

- Analyse and interpret data and help medical research to decision-making regarding the evaluation of the evolution of PD, and especially the progress of the phenomenon of freezing.
- Using their own experience, based on the results obtained by researchers using machine learning (ML).
- Interaction with researchers through natural language, for increasing the understanding between the researcher and the system.

1. Analysis of Existing Solutions

Cognitive computing allows application to process regarding pattern recognition, data mining and machine learning, applied in this project to extract new knowledge continuously from the information gathered by investigators in the phases of training patients. This was done using interaction processes in natural language to ease interaction with the researches, i.e., the so-called virtual assistants.

Over the course of the project, different proposals were analysed,

these came from systems applying cognitive computing to create virtual assistants. Among the possibilities that can be found:

- **IBM Watson** suggests services focused on language processing. Watson Conversation (WCS) allows automation of the interactions between users and systems. Using WCS, users can define aspects such as NLP intentions and entities and simulate complete conversations. WCS is used with other NPL Watson services such as Watson Natural Language Classifier and Watson Discovery.
- **Microsoft LUIS** (Language Understanding Intelligence Service) is a component of Microsoft Cognitive (MCS) services used to create and process natural language models. LUIS provides a sophisticated toolkit that enables developers to train the new conversational platform models. LUIS can also be used with other word processing API in MCS, for example, Linguistic Analysis and Text Analytics.
- **Google Natural Language API** as a part of Google Cloud focuses on NLP and NLU. NL API provides the ability to detect intentions and entities, feeling analysis, classification of content and relationship graphs. The NL API also includes sophisticated tools for training and new NL authoring models.
- **Facebook Wit.ai** is the platform behind the NLP/NLU capabilities of the Facebook Messenger platform. One of the best qualities of Wit.ai is the toolkit that can be used to train new models and monitoring interactions between users and the platform.
- **Amazon Lex**/Amazon Alexa Skills Kit enables the definition of intent and relevant entities in conversational interactions. One of the great advantages of Alexa is its integration with other web services such as Amazon offers, such as AWS Lambda or AWS Cognito.
- **Google DialogFlow** (known as api.ai) is a platform that allows developers to design and implement conversational interfaces that can be integrated into external applications such as chatbots. DialogFlow enables speech recognition and NLU fulfilment and a robust management toolkit.
- **Rasa** is the only cognitive computation system based on free software and is not offered as a cloud service. Rasa provides natural language recognition (rasa NLU) and a way to carry out complicated dialogues with ML (rasa Core).

D. Conversational Interface Queries

Conversational interfaces or chatbots are gaining prominence day by day thanks to the ease with which they fit into online solutions, and in some cases, without software programming skills. For this, natural language processing utilities (NLU) are mostly used.

When talking about chatbots it is important to account for the following concepts:

- **NLP (**natural language processing). general concept that refers to the ability of computer systems to understand the meaning of a text written in natural language.
- **NLU** (natural language understanding). It refers to the processes capable of decomposing unstructured text data structures that can be classified.
- **NLG** (natural language generation). It refers to the process followed by the chatbot to provide its response based on what has been asked.

In Table II, a comparison of the services listed in the above [25] (Watson, LUIS, API.ai, Wit.ai, Lex and RASA) for NLU module functionality that they implement, to visualise what Rasa can accomplish, offering the same functionality as the other solutions except that is not offered as a cloud service, but a downloadable application of local installation, which in turn is an advantage if what is intended is to use it within private networks.

TABLE II. Comparative of Functionality based on the NLU Services Offered by Cognitive Computing Main Solutions

Service	Intents	Entities	Batch import
LUIS	+	+	+
Watson	+	+	+
API.ai	+	+	+
wit.ai	+	+	O
Lex	+	O	-
RASA	+	+	+

For Mobeeze, Rasa was chosen for the platform configuration, since its adaptability to the needs of the current project was more important than the capabilities of the other solutions.

Rasa provides software for creating virtual assistants or chatbots that allow users to perform queries through natural language. Is consists of a set of open source ML tools. The main ones are NLU Rasa and Rasa Core:

- **NLU Rasa:** NLP tool for intention classification and entity extractions. On the one hand, Rasa NLU applies a classification of intent, i.e., interprets the meaning of the question based on a previously conducted training that determines through a percentage the probability in intention by which the question is asked. On the other hand, data analysis is carried out using the entity extraction modulus.
- **Rasa Core** (or Rasa DM - dialogue management) is the module responsible of guiding the conversations, using the history and the external context of the dialogue. This solution is trained using supervised ML and interactive way in order to create sophisticated dialogs.
- **NLG Rasa:** This is the process that builds Rasa response to a query.

Rasa uses ML techniques for the training of its decision engine. Thus, instead of thousands of rules, Rasa learns real conversational patterns. It must also be remarked that Rasa is distributed through Apache 2.0 open source license.

Fig. 1 shows the process followed by Rasa for analysing a sample formulated query, search information and make corresponding response [26].

As an information extracting tool, spaCy [27] was used [28]. It's a ML-based algorithm library for extracting information and a training platform.

SpaCy is a Python developed library, also offering pre-trained models in different languages, dependency analysis, segmentation phrases syntax, tokenisation, etc.

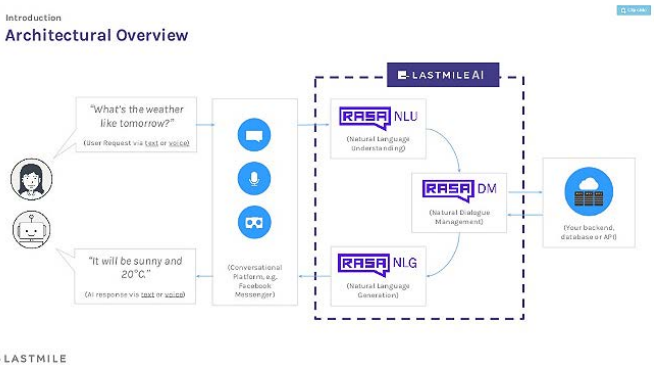


Fig. 1. Details Component involved in Rasa.

Regarding the data used for training, it must be kept in mind that in order check for correct functioning, both system capability data and evaluation data are required. Therefore, the existence of a considerable set of data is necessary.

For the current test, the training data set used was a corpus based on Wikipedia articles. This dataset is named Question-Answer Dataset [29]. Use of this corpus allowed having a set of data that served as a base to maintain a dialogue between the researcher and the chatbot.

Subsequently, a dataset was incorporated, this one consisting in specific information of the tests performed on the patients during Mobeeze virtual immersion training and also data from the results report submitted by the researchers.

IV. Results and Discussion

A. Immersion Virtual Patients with Parkinson

Throughout the Mobeeze project, samples of patients diagnosed with PD have been collected. Those participants in the control group (inhibitor freezing environment) that had a homogeneous gait with respect to walking variables (angles of both hips, knees and shoulders, stride height and speed). However, those subjects who were randomized to intervention group (favourable environment freezing) made a non-constant motion regarding the stride height and speed.

In this sense, those participants registered an increase of nearly twice the stride speed right at the moment that they came across other people or cars in the virtual environment. Furthermore, stride height is also clearly decreased prior approach to the above-mentioned virtual entities, which could be classified as a gait freezing.

In contrast, in the two immediate steps after the crossing, stride height increases significantly even in relation to the average during real world gait.

These results justify the use of the combination VR/NI to work through the freezing, coping with people or objects crossing the street in a non-invasive environment.

B. Motor Cognitive Inference

The objective pursued by the inclusion of a cognitive computing engine in the Mobeeze project is to analyse the possibility that the AI relates elements of the virtual immersion training designed as therapy for gait disturbance in PD patients.

To this end, it has been used as the basis for training the results obtained in the training stage of patients with PD in virtual immersion, added to the study conducted by the medical staff on the state of the art of the effectiveness of virtual immersion this same disease [30].

To do this, the classifications made and the results obtained for VR training groups and control groups have been used, with traditional training, based searches on scientific papers on the use of virtual immersion as therapeutic training. The methodology used followed the process indicated in Fig. 2.

The collection of information gathered by the medical team, is currently working on creating a knowledge base that provides findings (not currently validated from the medical point of view) on the effectiveness of the results of treatment in patient's PD in virtual immersion.

In Fig. 3. The results of the first tests performed using the information collected so far seen.

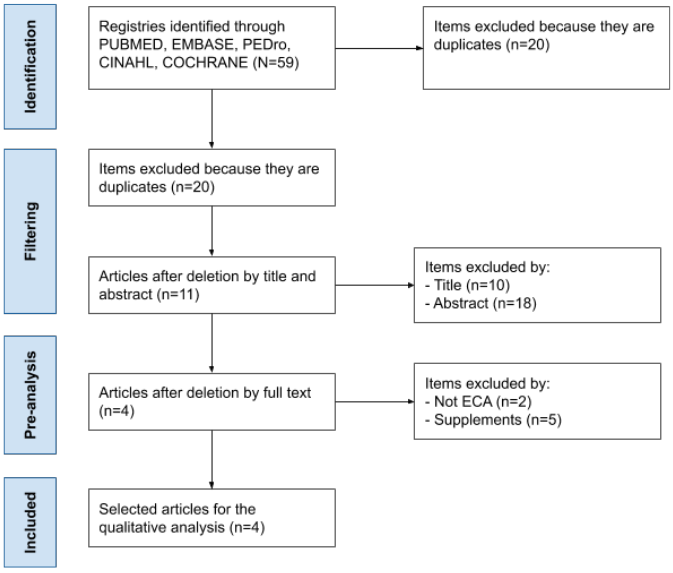


Fig. 2. Method of articles selection.

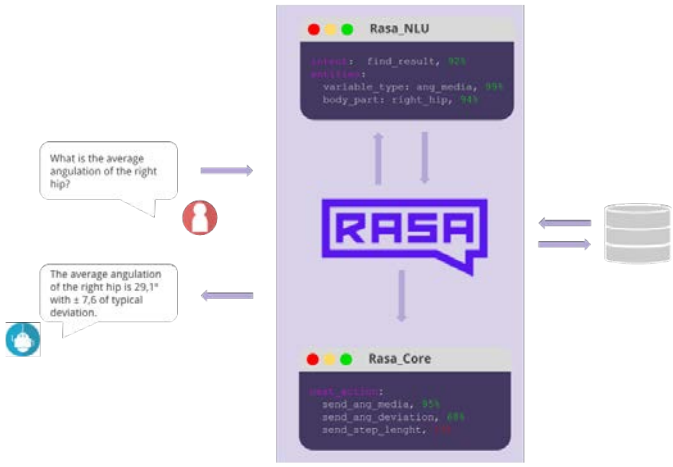


Fig. 3. Conversation example between an investigator and Rasa about the results through the training with patients.

As of today, the knowledge base is very limited, however it is expected that the incorporation of new data from training with patients and the use by researchers will help refine the answers.

V. Conclusions and Future Lines

A multidisciplinary project, called Mobeeze, has been developed, specially designed for people with Parkinson's disease, and that affects them with problems regarding mobility, specifically the phenomenon of freezing. In order to solve this, technologies like Oculus Rift, Microsoft Kinect or Leap Motion for recreation immersive virtual scenarios have been used. The purpose of this article is to transmit the evaluation made by the research team doctor who participated in the project on the applicability of ICT in the therapeutic treatment and tracking freezing. The results report developed by the medical team research the following conclusions are thrown:

1. The study conducted intends to describe the changes that occur in gait parameters for PD patients to undergo in two VR environments: inhibitor and freezing provocative.
2. VR therapy appears to modify the gait variables regardless of the environment in which the subject with PD is immersed.

3. VR therapy could be an alternative treatment to improve gait in patients with PD.

Furthermore, there has been an experimental development for extracting knowledge training conducted with patients using an AI system based on RASA, an ML system which provides an information search based conversational interface.

In order to achieve efficient machine learning, information quality and efficient focus on the problem is essential. It is the case of an employee of IBM, Jonathan Rezek, who works in the computer business of IBM promoting Watson [31] [32], who it turned out suffered from PD. To reduce the symptoms of PD, the main symptomatic therapy is L-DOPA, however, it can eventually cause dyskinesia [33]. Knowing the possibilities offered by cognitive engines, in this case Watson, contacted the medical team who treated him to carry out drug discovery that could be applied to treat dyskinesia, although originally were indicated for other diseases. The medical team gathered information training Watson and the results were encouraging as five drugs that preliminary tests showed positive results against dyskinesia, which had not been previously considered in patients with PD were found.

The first results obtained using Mobeeze in those affected with this disorder; suggest that virtual immersion when making therapy for this disorder may be beneficial. However, the project is still at a very early stage of training and requires additional tests on patients with PD, therefore it is too early to confirm its effectiveness.

A. Limitations

Introduction of the cognitive inference engine as an additional component to the project Mobeeze has led to the lack of initial data. This has resulted in a very early prototype calibration, which rendered the project results inconclusive.

B. Future Lines

As a continuation of the project, the intention is to continue the same line in a way that could enable the acquisition of new knowledge, on one hand, through training realized by the medical team that could bring a volume of information for patients with more population, group, major diversification, or the inclusion of other variables that enquire the base of knowledge of the interference motor, like the embedment of a new knowledge for the corpus data obtained by the state of art based on publications realized by other investigation teams.

The future lines in this work are directed to the study of the use of this platform through new technologies based on the speech recognition (like Alexa, Google Assistant etc.) and how to adapt it in order to carry out the therapeutic training through the virtual immersion for this disorder, in a way that could communicate with the patient during this treatment. This new panorama presented with this technology makes it easy to treat the patient by adding the virtual environment new acoustic stimuli and new study variables for a defined therapy that will be closer to reality.

Acknowledgement

This project has been funded by the Ministry of Economy and Competitiveness. It has been a collaborative project between Gestor. ex, Virtualware, Fidesol and IDIPAZ. Project code: RTC-2015-3870-1.

References

[1] Emamzadeh FN, Surguchov A. Parkinson's Disease: Biomarkers, Treatment, and Risk Factors. Front Neurosci. 2018; 12:612. 2018.

[2] Holden MK. Virtual environments for motor rehabilitation: review. CyberpsycholBehav 2005; 8: 187-211.

[3] Rizzo A, Kim GJ. A SWOT analysis of the field of virtual reality rehabilitation and therapy. Presence 2005; 14: 119-46.

[4] Viau A, Feldman AG, McFadyen BJ, Levin MF. Reaching in reality and virtual reality: a comparison of movement kinematics in healthy subjects and in adults with hemiparesis. J NeuroengRehabil 2004; 1: 11.

[5] Subramanian S, Knaut LA, Beaudoin C, McFayden BJ, Feldman AG, Levin MF. Virtual reality environments for post-stroke arm rehabilitation. J NeuroengRehabil 2007; 4: 20.

[6] Kuttuva M, Boian R, Merians A, Burdea G, Bouzit M, Lewis J, et al. The Rutgers Arm, a rehabilitation system in virtual reality: a pilot study. CyberpsycholBehav 2006; 9: 148-51.

[7] Fulk GD. Locomotor training and virtual reality-based balance training for an individual with multiple sclerosis: a case report. J Neurol Phys Ther 2005; 29: 34-42.

[8] Holden M, Dettwiler A, Dyar T, Niemann G, Bizzi E. Retraining movement in patients with acquired brain injury using a virtual environment. In Westwood JD, Hoffman HM, Mogel GT, Stredney D, eds. Medicine meets virtual reality. Newport Beach: IOS Press; 2009. p. 192-8.

[9] You SH, Jang SH, Kim Y, Hallett M, Ahn SH, Kwon Y, et al. Virtual reality-induced cortical reorganization and associated locomotor recovery in chronic stroke. An experimental- blind randomized study. Stroke 2005; 36: 1166-71.

[10] Fidopiastis CM, Stapleton CB, Whiteside JD, Hughes CE, Fiore SM, Martin GA, et al. Human experience modeler: context driven cognitive retraining to facilitate transfer of learning. CyberpsycholBehav 2006; 9: 183-7.

[11] Van Gemert-Pinjen L, Kelders S, Kip H, Sandderman R. eHealth research, theory and development: a multi-disciplinary approach. Routledge; 2018.

[12] Moya, P. Kinect para ayudar a pacientes de Parkinson con bloqueo de la marcha. Omicromo. https://omicrono.elespanol.com/2015/05/kinect-para-pacientes-de-parkinson/ (10 May 2015)

[13] Pilgrim, Tim. Novel Kinect system helps keep Parkinson's patients moving. Phys.org. https://phys.org/news/2018-05-kinect-parkinson-patients.html (24 May 2018)

[14] Alvarez M, Grogan P. Connecting with Kinect® To Improve Motor and Gait Function in Parkinson Disease. Neurology Center for Neurological Care and Research San Antonio TX. April 2012.

[15] Dockx K, Bekkers EMJ, Van den Bergh V, Ginis P, Rochester L, Hausdorff JM, Mirelman A, Nieuwboer A. Virtual reality for rehabilitation in Parkinson's disease. Cochrane Database of Systematic Reviews 2016, Issue 12. Art. No.: CD010760.

[16] Evans, D. (2017). The fundamental differences between automation and AI. VentureBeat, 4 October. https://venturebeat.com/2017/10/04/the-fundamental-differences-between-automa-tion-and-ai/. Accessed 5 January 2018.

[17] IBM Watson: la tecnología cognitiva que abre una nueva era de la computación. Ver en: https://www-03.ibm.com/press/es/es/attachment/49310.wss?fileId=ATTACH_FILE1&fileName=Dossier%20IBM%20Watson.pdf

[18] Krishna, Chico. Harnessing AI for Parkinson's Research. https://www.michaeljfox.org/foundation/news-detail.php?harnessing-ai-for-parkinson-research (14 January 2019).

[19] IBM presenta pequeño sensor de uñas con inteligencia artificial y anuncia una alianza con la Fundación Michael J. Fox para comprender mejor y predecir la progresión de la enfermedad de Parkinson (8 January 2019). Retrieved from: https://www-03.ibm.com/press/mx/es/pressrelease/54644.wss

[20] Lombardo, J. M., M. A. Lopez, V. M. García, M. López, R. Cañadas, S. Velasco, and M. León. 2018. PRACTICA. A Virtual Reality Platform for Specialized Training Oriented to Improve the Productivity. International Journal of Interactive Multimedia and Artificial Intelligence (IJIMAI). Special Issue on Artificial Intelligence Applications. Volume 5, num. 4, pags. 94-101. ISSN: 1989-1660

[21] Morales-Gómez S, Elizagaray García I, Yepes-Rojas O, De la Puente-Ranea L, Gil-Martínez A. Efectividad de los programas de inmersión virtual en los pacientes con enfermedad de Parkinson. Revisiónsistemática. Rev Neurol 2018; 66: 69-80.

[22] Nantel J, de Solages C, Bronte-Stewart H. Repetitive stepping in place identifies and measures freezing episodes in subjects with Parkinson's disease. Gait Posture 2011; 34: 329–333.

[23] Ziv I, Avraham M, Dabby R, Zoldan J, Djaldetti R, Melamed E. Early-

occurrence of manual motor blocks in Parkinson's disease: a quantitative evaluation. Acta Neurol Scand. 1999 feb;99(2):106–11.

[24] Rodríguez Guevara C, Lugo LH. Validez y confiabilidad de la Escala de Tinetti para población colombiana. Rev ColombReumatol. 19(4):218–33.

[25] Braun, D., Hernandez Mendez, A., Matthes, F., Langen, M. Evaluating Natural Language Understanding Services for Conversational Question Answering Systems. 2017. Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue. Association for Computational Linguistics. Pages 174--185. DOI: 10.18653/v1/W17-5522

[26] Rasa AI: Building clever chatbots. Slideshare. Available at: https://www.slideshare.net/TomBocklisch/rasa-ai-building-clever-chatbots

[27] Davis E. King. Dlib-ml: A Machine Learning Toolkit. Journal of Machine Learning Research 10, pp. 1755-1758, 2009.

[28] Paramveer Dhillon, Dean Foster and Lyle Ungar, Eigenwords: Spectral Word Embeddings, Journal of Machine Learning Research (JMLR), 16, 2015.

[29] Noah A. Smith, Michael Heilman, and Rebecca Hwa. Question Generation as a Competitive Undergraduate Course Project in Proceedings of the NSF Workshop on the Question Generation Shared Task and Evaluation Challenge, Arlington, VA, September 2008. Available at: http://www.cs.cmu.edu/~nasmith/papers/smith+heilman+hwa.nsf08.pdf

[30] Morales-Gómez S, ElizagarayGarcía I, Yepes-Rojas O, De la Puente-Ranea L, Gil-Martínez A. Efectividad de los programas de inmersión virtual en los pacientes con enfermedad de Parkinson. Rev Neurol 2018; 66: 69-80.

[31] Rezek, J. May 2018. Revolutionizing drug research for Parkinson's with AI. AI/Watson | Healthcare. IBM. Available at: https://www.ibm.com/blogs/client-voices/revolutionizing-drug-research-parkinsons-ai/

[32] Rezek, J. How one Parkinson's patient drove a groundbreaking research effort. Passion Projects. IBM. Available at: https://www.ibm.com/thought-leadership/passion-projects/parkinsons-watson-drug-discovery/

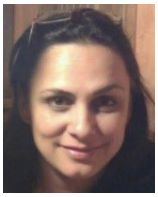
[33] Lee CS. Levodopa-induced dyskinesia: Mechanisms and management. Issue: BCMJ, Vol. 43, No. 4, May 2001, page(s) 206-209 Articles. Available at: https://www.bcmj.org/articles/levodopa-induced-dyskinesia-mechanisms-and-management

Juan Manuel Lombardo

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

Miguel Ángel López, has a degree in Engineering in Technical Engineering in Computer Systems from the University of Almería, graduates in Computer Engineering and Master in Softcomputing 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.



Mabel López

Mabel López 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.

Mónica León

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.

Felipe Miron

Felipe Miron is the European Projects Coordinator of the KTO at FIDESOL, where he manages the foresight and technological surveillance actions of the entity, both state and European level, as well as the project proposals. He is a Computer Engineer from the University of Granada. Before joining FIDESOL, he worked for the Government of the Region of Murcia leading a development team for the transformation of the File Processing system of the Labor Counseling and Social Policy. He has been part of development teams for national projects (Banesto, Telefonica, Onda Cero) and international (Credit Suisse, IBM). He has collaborated in several publications related to disciplines such as Artificial Intelligence, Mobility and Big Data.

Jon Arambarri Basanez

Jon Arambarri Basanez. Dr. Telecommunication Engineer and MBA. Responsible for Innovation at www.virtualwaregroup.com oriented to the generation of new product for eHealth and Industry4.0. My research interests combine the regulation and efficient management of Telecommunications Infrastructures (Telecommunication Networks, Visualization and Multimedia and Internet of Things - IoT) together with the optimization of ICT services (Artificial Intelligence and Predictive Systems, Cyber Security) the development of innovative business solutions, mainly for eHealth and industry4.0.

David Alvarez

David Alvarez receibed his Ph.D in Computer Science at the University of León, Spain in 2015, during which he worked on an Optical Caracter Recognition of handwritten documents. He then joined the Research Group on Numerical Simulation and Scientific Calculus (SINUMCC) where he worked on Forest Fire Spread Simulation. He is now a member of the Research Group of Fundación del Software Libre (FIDESOL). His research interes include in general computer vision, machine learning, image recognition and pattern recognition.

Do Women and Men Perceive User Experience Differently?

Kristina Aufderhaar[1], Martin Schrepp[2], Jörg Thomaschewski[1]*

[1] University of Applied Sciences Emden/Leer (Germany)
[2] SAP AG (Germany)

Received 13 November 2018 | Accepted 14 March 2019 | Published 22 March 2019

unir
LA UNIVERSIDAD EN INTERNET

Abstract

We study three web sites to see whether there are systematic differences between women and men in their rating of the user experience of the sites. One of the sites addresses especially the target group of women, another the target group of men, whereas the third site is neutral in this respect. The selection of the sites was safeguarded with gender screening. The participants in the study rated the three chosen websites with the questionnaires UEQ and VISAWI-S. The results indicate that there are no substantial differences in the perception of the UX between men and women. Personal attitudes and preferences seem to have a substantially greater influence than sex.

Keywords

Usability, User Experience, Gender Differences.

DOI: 10.9781/ijimai.2019.03.005

I. Introduction

User experience is a complex and very subjectively perceived product characteristic [1].

The perception of the user experience of an interactive product is based on the perception of several distinct aspects, for example, the efficiency, the extent to which a product can be used intuitively, the fun of use, the impression that the product is innovative or leading edge, the attractiveness of the user interface, trust in the security of a product, the extent to which the user feels in control of the product, etc. Thus, a huge number of separate and often highly subjective perceptions are responsible for the overall judgement concerning the user experience of a product [2, 3].

In addition, if we ask several users concerning their impression of the user experience of the same product, we may see a high variation in their judgements. We may find some quite satisfied and happy users as well as some frustrated users in the same investigation. This is due to the fact that different persons have a different history and experience concerning the use of interactive products. If a user is already familiar with a similar product he or she will find a new product quite intuitive to use. Another user may have the impression that the same product is extremely hard to understand, simply because he or she has no experience with similar products. Personal preferences concerning organization of information on the user interface or concerning personal working styles also play an important role for the impression of a subject concerning the user experience of a product.

Another source of variation is personal preferences or personal taste concerning design elements. A visual design perceived as attractive and beautiful by a group of users may be perceived as boring and ugly by another group. In addition, different users have different opinions about the relative importance of UX aspects for certain types of products [4].

* Corresponding author.
E-mail address: joerg.thomaschewski@hs-emden-leer.de

One interesting question is the extent to which demographic characteristics, e.g. the sex or age of a person, have an influence upon the perception of the user experience of web sites or generally interactive products. This is especially of interest for designers of pages that have a primarily female or male target group. Currently, there is only limited information concerning this question available in UX research literature.

Potential differences between males and females concerning the perception of UX can result from different strategies of information processing. An often cited paper [5] found, for example, that women process information in a more holistic way, while men use a more selective information processing strategy (similar results are reported in [6, 7, 8]).

Concerning the perception of web sites [9] found that men are in general more satisfied with the displayed information than women. This was confirmed also in a study by [10]. In addition, there are some studies [11, 12] concerning differences in the perception of the quality of the visual design, for example different preferences concerning colours used in a web site. In an experiment concerning web sites for children [13] results indicate gender specific differences in the personal preferences of boys and girls concerning visual complexity. In this study boys preferred in average a higher level of visual complexity than girls.

Other results indicate that women are more critical about the aesthetics or visual design of a website than men [10, 14].

A study [15] found a positive relationship between web knowledge of subjects and their perception of web usability. This effect is moderated by sex and website design experience in the sense that subjects with website design experience place more value on the usefulness of web sites and this moderation effect is stronger for females than for males.

Besides web-sites there is also some work concerning gender differences in software used for working or problem-solving tasks. In [16] a systematic method is described (the approach is based on persons and works out five facets of gender differences that are used in a gender-specialized Cognitive Walkthrough method) that allows to

find gender-inclusiveness issues in software, so that practitioners can design and produce software that is more usable by everyone.

A study [17] investigated a group of users who have to use a new software product in the context of the technology acceptance model [18]. They found that the perception of usefulness had a higher impact on the usage intention of men than on the usage intention of women. In contrast, the perception of ease of use had a stronger impact on the usage intention of women than on the usage intention of men.

Of course, not all websites appeal to women and men in equal measure. Many sites are designed for a predominantly female or predominantly male target group. Typical examples are web sites of online journals that cover a topic mainly targeted to males or females or web shops with an offering targeted to a specific group of persons in which one gender role is clearly dominant.

Does this design of the content for a male or female target group also influence the perception of typical user experience criteria, for example efficiency, controllability, fun of use, aesthetic appeal or stimulation?

We will investigate this question by a study with three popular German web sites. One of these websites is intended to appeal specially to the target group women, another to the target group men and the third to both groups.

The perception of the UX aspects will be measured by two established and widely used UX questionnaires, the UEQ [19, 20] and the VISAWI [21, 22], that capture together a wide range of UX aspects.

The UEQ measures UX on the following 6 scales:

- *Attractiveness*: Overall impression of the product. Do users like or dislike the product?
- *Efficiency*: Can users solve their tasks without unnecessary effort?
- *Perspicuity*: Is it easy to get familiar with the product? Is it easy to learn how to use the product?
- *Dependability*: Does the user feel in control of the interaction?
- *Stimulation*: Is it exciting and motivating to use the product?
- *Novelty*: Is the product innovative and creative? Does the product catch the interest of users?

The VISAWI measures visual aesthetics of web sites or, more generally, of the user interface of interactive products.

II. Research Hypothesis

The study by Simon [9] showed that men are usually more satisfied with the presentation of the information on a website than women. Thus, in a web site that is not particularly designed for men or woman we should be able to see this effect in the UX scales measured by the UEQ and VISAWI questionnaires.

This leads to our first hypothesis.

H1: In the case of a sex-neutral website, men are generally more satisfied with all UX aspects than women.

The quality of the content of a site is of course the most important point in the rating of the site by its users [23]. If the users find the content appealing and interesting, they will visit the site frequently and rate its UX quality positively.

This is said to apply especially to the rating of hedonic qualities, however, and not so much to the rating of the pragmatic qualities [2]. For example, how easy it is to understand the page structure and how efficient the navigation on the web site is, is not much influenced by the fact that the content is optimized for the target group. But UX aspects like *Stimulation* (fun of use) or the perception of the site as original and novel will be of course influenced by the content.

This will be also true for the aesthetic impression, which is for web sites mainly determined by the quality of the pictures in the content and the layout.

This leads to the following hypotheses:

H2: Women and men find a website specially designed for them more aesthetically pleasing than one designed for the other sex.

H3: Women and men award higher attractiveness scores (valence) for a website specially designed for them and higher values for the hedonic qualities of stimulation and originality than for a site that is designed for the other sex or neutral.

H4: The target group of a site has no substantial influence on the rating of the pragmatic qualities of efficiency, clarity and controllability.

These hypotheses were investigated in an online study with three websites.

III. Study

One of the websites is intended to appeal specially to women, another to men and the third to both groups equally. Three popular and quite common web sites were selected based on knowledge concerning their mayor target group.

In order to justify this selection based on objective criteria, gender screening [24] was carried out. In this method, first of all, the women's and men's names in the imprint and/or the contact pages are counted, in order to find out whether predominantly women or men were involved in the producing of the site. In the second step, the number of women and of men pictured in photos is counted. A qualitative analysis is then carried out to ascertain the extent to which women or men are being addressed in the text.

The web site "Die Zeit" is the online channel of a quite popular German print magazine focussed on topics from politics, history, economy, education and society. This print magazine and the web site are not addressing readers of a specific gender.

The web site of "Brigitte" is the web channel of the most popular German Woman's magazine. It is focussed on topics like fashion, beauty, health, love and general practical hints for the organization of daily live. Thus, the main target group is obviously female.

The web site "GQ" is clearly focussed on men. Typical topics are men's fashion, entertainment, cars, technology and health.

The entry pages (date 22.10.2017) of these three web sites are shown in Fig. 1 – 3 to give some impression about the design of these sites. These pages were also used as part of the study described below.

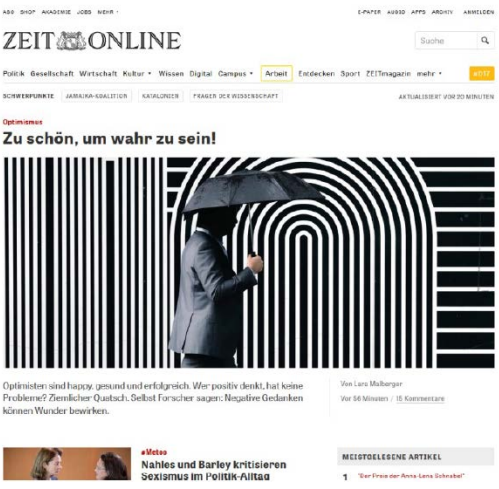


Fig. 1. Entry page of www.zeit.de (date 22.10.2017).

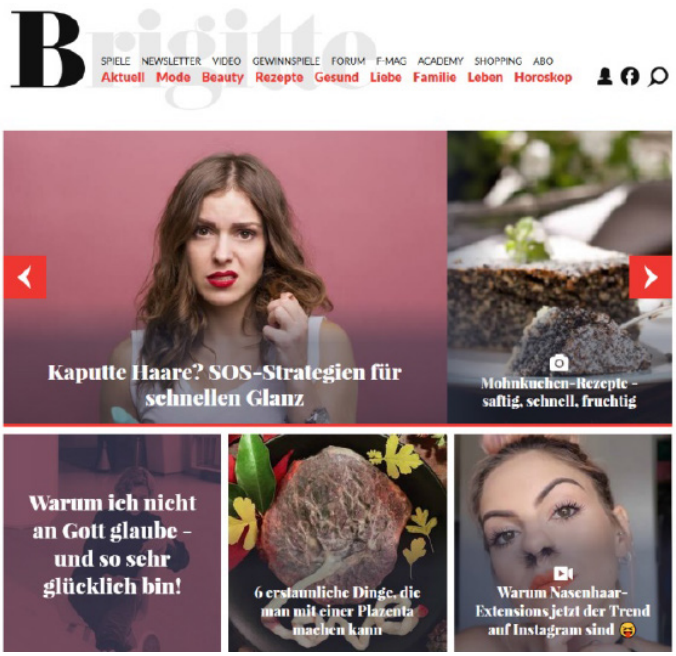


Fig. 2. Entry page of Brigitte.de (date 22.10.2017).

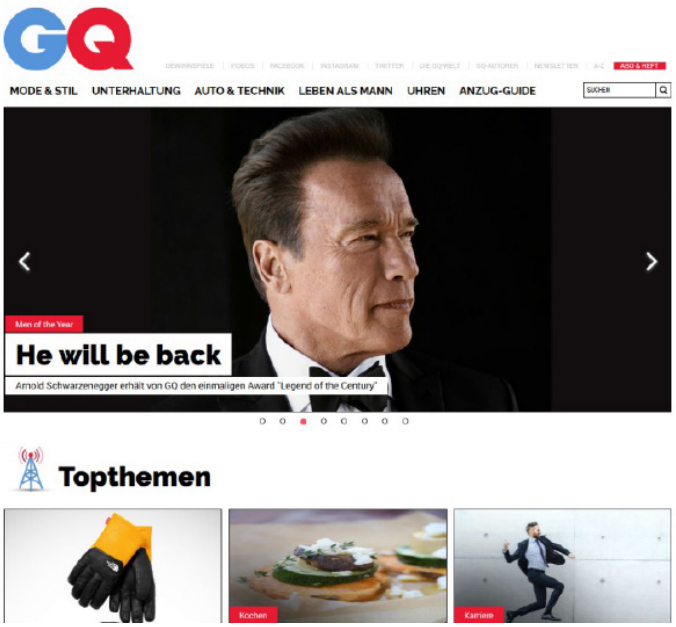


Fig. 3. Entry page of gq-magazin.de (date 22.10.2017).

The results of the gender screening of these web sites are shown in Table I. The results confirm our assumptions concerning the main target groups for these sites.

The study was advertised as an online survey via e-mail distribution lists at colleges and universities. As an incentive, several Amazon vouchers were raffled among all participants. When the link to the study was clicked upon, first a site appeared with a short briefing, a picture of the website to be rated and a link to that site. The participants were instructed to navigate onto the web site, to look at it thoroughly, and to read an article on the site. Afterwards the site was rated with the short version of the VISAWI known as VISAWI-S [21] and the UEQ [19, 20].

TABLE I. Results of the Gender Screening of the Three Investigated Web Sites

Criteria	Die Zeit	Brigitte	GQ
Quantitative Gender Screening			
Male Names	85	4	28
Photos Men	15	2	43
Females Names	77	23	15
Photos Woman	12	22	1
Qualitative Gender Screening			
Texts	Objective, no gender preferred, politics, business, society, job, news, culture, education	Much direct speech, women's names, a lot of slang, advice, recipes, ornaments, fashion, beauty	Many men's names, guidebooks, partly colloquial language, cars, fashion, occupation, technology
Pictures	Not much eye contact, mostly illustration of situations, no preference for a gender	Many photos with eye contact, many beautiful, happy motivated women	Little eye contact, mostly illustration of situations, many well-groomed, cool-looking men

The survey was started by the following number of people:

- *Die Zeit*: 184 (121 female, 63 male)
- *Brigitte*: 116 (65 female, 51 male)
- *GQ*: 149 (106 female, 43 male)

Obviously, there is a much higher number of females in the target group reached by the e-mail distribution lists.

Participants for whom there was less than 4 minutes between the start of the survey and the sending of the results or who had too many inconsistent answers in the UEQ [25] were excluded, as one can assume that the task was not performed as intended in these cases. After this sorting, the following numbers of usable data sets remained: *Die Zeit* (22 male, 52 female), *Brigitte* (22 male, 29 female) und *GQ* (22 male, 49 female).

Thus, the percentage of usable responses per web site and gender category is:

- *Die Zeit*: 40% (43% female, 35% male)
- *Brigitte*: 43% (44% female, 43% male)
- *GQ*: 47% (46% female, 51% male)

There is not much difference between the different web sites concerning the number of responses that could not be used for the data analysis. When we look in detail to the gender of the participants that quit the survey or were removed from the data analysis because of poor quality of their response data (see reasons described above), we see that web site and gender does not have a big influence on the number of removed responses.

IV. Results

Fig. 4 shows the measured values for the aesthetics of the websites and their 5% confidence intervals, sorted into men and women.

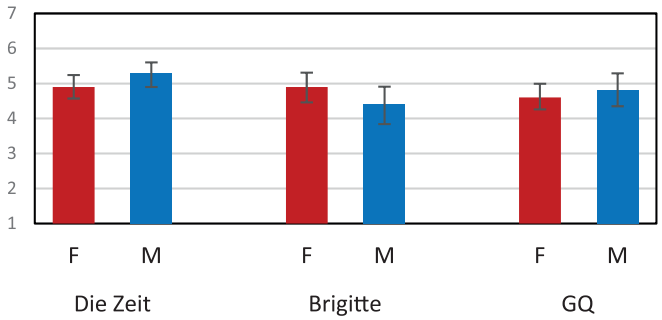


Fig. 4. Aesthetic values of the short version of the VISAWI for the three examined sites (F = female, M = male).

Women find the website *Brigitte* more aesthetically appealing than men, whereas the reverse of this is the case for *GQ*. The differences are very small, however, and in no case statistically significant (t-Test, .05).

Fig. 5-7 shows the results of the UEQ (average ratings and 5% confidence intervals for the three examined sites, sorted into men and women). The following abbreviations are used in Fig. 5-7: ATT Attractiveness; EFF Efficiency; PER Perspicuity; DEP Dependability; STI Stimulation; NOV Novelty. The scale structure is well described in [26].

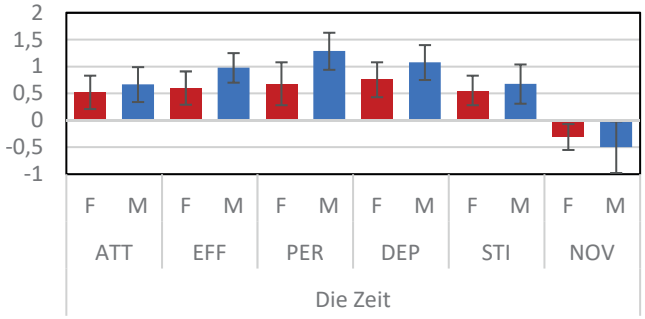


Fig. 5. Values of the UEQ scales and confidence intervals for the web page "Die Zeit".

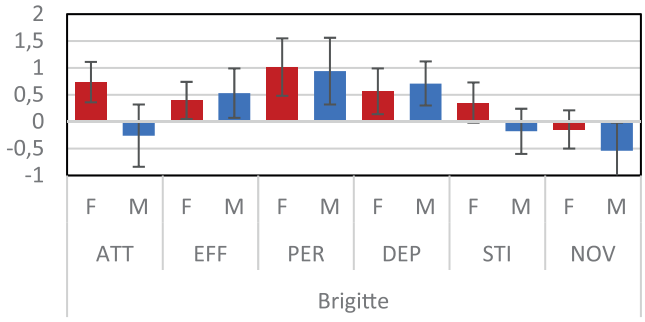


Fig. 6. Values of the UEQ scales and confidence intervals for the web page "Brigitte".

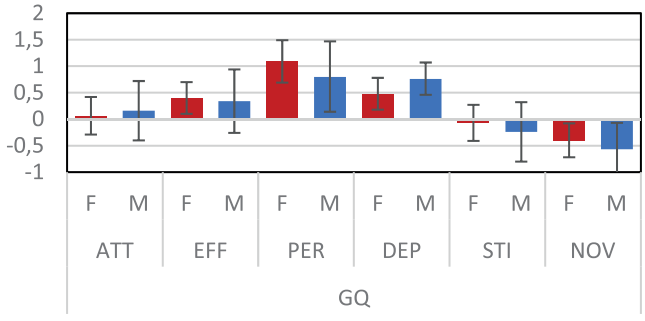


Fig. 7. Values of the UEQ scales and confidence intervals for the web page "GQ".

The results of the UEQ for *Die Zeit* tend to be indicative of hypothesis ***H1*** (with the exception of novelty, see Fig. 5), but the differences are not significant (t-Test, .05). This may, however, be due to the small sample of men.

As Fig. 4 shows, the differences regarding the perceived visual aesthetics are very slight. Hypothesis ***H2*** can thus be rejected.

The website *Brigitte* was rated significantly better by women than by men as regards attractiveness (see Fig. 6; t-Test, .05). This also applies as regards stimulation and originality, although the differences here were not significant either.

The expected effect cannot be observed with the GQ website (see Fig. 7), however. Here, the rating by both sexes is almost identical. This may, of course, be because this website propagates a relatively extreme role model, i.e. one that only actually appeals to a small proportion of the target group men. Thus, our results do not support ***H3***, but some tendency in the direction of this hypothesis can be seen with the web site Brigitte. More research seems to be required to get a clearer picture here.

The ratings of the practical qualities for the websites Brigitte and GQ are very similar for both sexes. This tends to be indicative of hypothesis ***H4***.

An analysis of variance (ANOVA) was done to find out how much of the variance in the responses could be explained by differences in gender and how much is due to other not controlled factors, which are in our case interindividual differences in the taste concerning visual design or in the opinions concerning the importance of the UX aspects measured by the UEQ. The results showed only for the web site Brigitte and the UEQ scale *Attractiveness* a non-neglectable percentage of variance that is explained by the gender difference (.17). For all other combinations of web site and UX aspect, the corresponding values range from .07 to 0. Thus, compared to interindividual differences the gender differences did not have much impact.

V. Conclusions

The results indicate that sex has no really considerable influence as regards the perception of typical user experience factors, such as those measured in e.g. the VISAWI-S or UEQ.

It must be mentioned, as a limitation of this study, that the number of male participants was quite small, which may of course have caused existing differences to not become significant. A closer look at the data shows, however, that there were very strong differences within the groups. Personal role models and attitudes obviously account for a greater proportion of the UX rating than biological sex.

However, the study just creates some first results and has obviously some limitations. The participants spend only a quite short time on the investigated web sites. Thus, the UEQ ratings concerning the pragmatic UX aspects *Efficiency*, *Perspicuity* and *Dependability* may be influenced by this limited usage and navigation experience. The hedonic UX aspects of *Attractiveness*, *Stimulation* and *Novelty* may also be influenced by this, but to a much smaller degree. It is well-known that the visual impression of a web-site forms quite fast [27], so there should be no impact of the short usage time on the results measured by the VISAWI.

In addition, there are other factors like cultural differences, age, special interests or beliefs that may have an impact on the subjective impression concerning user experience and that were not controlled in this study. Thus, further research is required to get a clearer picture here. Our participants were students, thus form a more or less homogenous group concerning age. It will be quite interesting to replicate the study with older participants, to see if the obviously existing differences in the understanding of gender roles between different generations have an impact.

References

[1] Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S. & Carey, T. (1994). Human-Computer Interaction. Workinghal, Addison-Wesley.

[2] Hassenzahl, M. (2001). The effect of perceived hedonic quality on product appealingness. International Journal of Human-Computer Interaction, vol. 13, no.4 , pp. 481–499.

[3] Schrepp, M. (2018). User Experience mit Fragebögen messen [Measure user experience with questionnaires]. Amazon Kindle Direct Publishing, ISBN: 9781986843768.

[4] Winter, D., Hinderks, A., Schrepp, M. & Thomaschewski, J., (2017). Welche UX Faktoren sind für mein Produkt wichtig? [Which UX factors are important for my product?] In: Hess, S. & Fischer, H. (Hrsg.), Mensch und Computer 2017 - Usability Professionals. Regensburg: Gesellschaft für Informatik e.V., pp. 191 – 200.

[5] Meyers-Levy, J. (1989). Gender differences in information processing: A selectivity interpretation. In Coffenator, P. & Tybout, A. (Eds.), Cognitive and affective responses to advertising. Lexington books/DC Heath & Con, Lexington, MA.

[6] Meyers-Levy, J. & Maheswaran, D. (1991). Exploring differences in males and females processing strategy. Journal of Consumer Research, vol. 18, no. 1, pp. 63-70.

[7] Meyers-Levy, J. & Sternthal, D. (1991). Gender differences in the use of message cues and judgements. Journal of Marketing Research, vol. 28, no. 1, pp. 84-96.

[8] Graham, J., Myers, J. & Graham, M. (2002). Gender differences in investment strategies: An information processing perspective. International Journal of Bank Marketing, vol. 20, no. 1, pp. 17-26.

[9] Simon, S. J. (2001). The impact of culture and gender on websites: An empirical study. The database for advances in information systems, vol. 32, no. 1, pp. 18-37.

[10] Cyr, D. & Bonnani, C. (2005). Gender and website design in E-Business. International Journal of Electronic Business, vol. 3, no. 6, pp. 565-582.

[11] Moss, G., Gunn, R. & Heller, J. (2006). Some men like it black, some woman like it pink: Consumer implications of differences in male and female website design. Journal of Consumer Behavior, vol. 5, no. 4, pp. 328-341.

[12] Djamasbi, S., Tullis, T., Hsu, J. Mazuera, E., Osberg, K. & Bosch, J. (2007). Gender preferences in web design: Usability testing through eye tracking. AMCIS 2007 Proceedings.

[13] Hsiu-Feng, W. (2014). Picture perfect: Girls' and boys' preferences towards visual complexity in children's websites. Computers in Human Behavior, vol. 31, pp. 551-557.

[14] Arcand, M. & Nantel, J. (2005). Gender differences in processing information: Implications for online search patterns and task performance. In Cunningham, P. & Cunningham, M. (Eds.), Innovation and Knowledge Economy: Issues, Applications, Case Studies – Part 2, pp. 1334-1341. Amsterdam, NT.

[15] Page, K. L., Robson, M. J. & Uncles, M. D. (2012). Perceptions of web knowledge and usability: When sex and experience matter. International Journal of Human-Computer Studies, vol. 70, no. 12, pp. 907-919.

[16] Burnett, M., Stumpf, S., Macbeth, J., Makri, S., Beckwith, L., Kwan, I., Peters, A. & Jernigan, W. (2016). GenderMag: A Method for Evaluating Software's Gender Inclusiveness. Interacting with Computers, vol. 28, no. 6.

[17] Venkatesh, V. & Morris, M. G. (2000). Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. MIS quarterly, pp. 115-139.

[18] Davis, F., Bagozzi, P. & Warshaw, P. (1989), 'User acceptance of computer technology - a comparison of two theoretical models', Management Science, vol. 35, no. 8, pp. 982–1003.

[19] Laugwitz, B., Held, T. & Schrepp, M. (2008). Construction and Evaluation of a User Experience Questionnaire. Lecture Notes in Computer Science. (5298), pp. 63–76.

[20] Rauschenberger, M., Schrepp, M., Cota, M.P., Olschner, S. & Thomaschewski, J. (2013). Efficient measurement of the user experience of interactive products - How to use the User Experience Questionnaire (UEQ). Example: Spanish Language Version. International Journal of Interactive Multimedia and Artificial Intelligence, vol. 2, no. 1, pp. 39- 45. DOI: 10.9781/ijimai.2013.215.

[21] Moshagen, M. & Thielsch, M. T. (2013). A short version of the visual aesthetics of websites inventory. Behaviour & Information Technology, vol. 32, no. 12, pp. 1305-1311.

[22] Moshagen, M. & Thielsch, M. T. (2010). Facets of visual aesthetics. International journal of human-computer studies, vol. 68, no. 10, pp. 689-709.

[23] Thielsch, M. T., Blotenberg, I. & Jaron, R. (2014). User evaluation of websites: From first impression to recommendation. Interacting with Computers, vol. 26, no. 1, pp. 89-102.

[24] Knoll, B., Inhof, C. & Posch, P. (2013): Bericht (Kurzfassung): Gender Screening – Männer an Maschinen und Frauen im Labormantel [Report (short version): Gender Screening - Men on Machines and Women in Laboratory Coats]. Studie zur Repräsentation von Frauen und Männern in ausgewählten Online- und Printmedien von technisch-ingenieurwissenschaftlichen Universitäten, Fachhochschulen und Forschungseinrichtungen in Österreich.

[25] Schrepp, M. (2016). Datenqualität bei Online-Fragebögen sicherstellen [Ensuring data quality for online questionnaires.]. S. Hess & H. Fischer (Hrsg.): Mensch und Computer 2016 – Usability Professionals.

[26] Schrepp, M., Hinderks, A. & Thomaschewski, J. (2017): Construction of a Benchmark for the User Experience Questionnaire (UEQ). International Journal of Interactive Multimedia and Artificial Intelligence, vol. 4, no. 4, pp. 39- 45. DOI: 10.9781/ijimai.2017.445.

[27] Lindgaard, G., Fernandes, G., Dudek, C. & Brown, J. (2006). Attention web designers: You have 50 milliseconds to make a good first impression!. Behaviour & information technology, vol. 25, no. 2, pp. 115-126.

Kristina Aufderhaar

Kristina Aufderhaar received a Bachelor of Science in Computer Science and Digital Media from the University of Applied Sciences Emden/Leer (Germany) in 2018. Since 2011 she works as a media and web designer. Her research interests are human-computer interaction and gender studies to improve her work.

Martin Schrepp

Martin Schrepp has been working as a user interface designer for SAP AG since 1994. He finished his diploma in mathematics in 1990 at the University of Heidelberg (Germany). In 1993 he received a PhD in Psychology (also from the University of Heidelberg). His research interests are the application of psychological theories to improve the design of software interfaces, the application of *Design for All* principles to increase accessibility of business software, measurement of usability and user experience, and the development of general data analysis methods. He has published several papers concerning these research fields.

Jörg Thomaschewski

Jörg Thomaschewski received a PhD in physics from the University of Bremen (Germany) in 1996. He became Full Professor at the University of Applied Sciences Emden/Leer (Germany) in September 2000. His research interests are human-computer interaction, e-learning, and software engineering. Dr. Thomaschewski is the head of the research group "Agile Software Development and User Experience".

LA UNIVERSIDAD
EN INTERNET

Rectorado
Avenida de la Paz, 137
26006 Logroño (La Rioja)
t (+34) 941 21 02 11

www.unir.net