Board of Directors' Profile: A Case for Deep Learning as a Valid Methodology to Finance Research

César Vaca, Fernando Tejerina, Benjamín Sahelices *

University of Valladolid, Valladolid (Spain)

Received 22 April 2022 | Accepted 29 June 2022 | Early Access 19 September 2022

Abstract

This paper presents a Deep Learning (DL) model for natural language processing of unstructured CVs to generate a six-dimensional profile of the professional experience of the Spanish companies' board of directors. We show the complete process starting with open data extraction and cleaning, the generation of a labeled dataset for supervised learning, the development, training and validation of a DL model capable of accurately analyzing the dataset, and, finally, a data analysis work based on the automated generation of the professional profiles of more than 6,000 directors of Spanish listed companies between 2003 and 2020. An RNN-LSTM neural network has been trained in three phases starting from a random initial state, (1) learning of basic structures of the Spanish language, (2) fine tuning for scientific texts in the field of economics and finance, and (3) regression modeling to generate a six-dimensional profile based on a generalization of sentiment classification systems. The complete training has been carried out with very low computational requirements, having a total duration of 120 hours of processing in a low-end GPU. The results obtained in the validation of the DL model show great accuracy, obtaining a value for the standard deviation of the mean error between 0.015 and 0.033. As a result, we have been able to outline with a high degree of reliability the profile of the listed Spanish companies' board of directors. We found that the predominant profile is that of directors with experience in executive or consultancy positions, followed by the financial profile. The results achieved show the potential of DL in social science research, particularly in Finance.

Keywords

Artificial Intelligence, Board Diversity, Deep Learning, Finance, Long Short-Term Memory, Recurrent Neural Networks.

DOI: 10.9781/ijimai.2022.09.005

I. Introduction

The amount of available open data from multiple areas has grown exponentially in recent years, which represents an enormous potential for researchers. However, due to the characteristics of much of the data generated, the processing of all this volume of information is also a huge challenge. Much of the data is characterized by its large size, high dimensionality and complex structure, making traditional econometric methods not entirely helpful, leaving a large amount of useful information behind. In this sense, the growing development of Machine Learning techniques provides researchers with effective tools for processing this type of data.

The generalization in the use of convolutional neural networks as well as multiple variants such as recurrent networks and long short-term memory, together with the use of normalization, regularization and optimization techniques, has allowed the creation of networks with many learning layers that can be trained effectively (References [1-4]). These deep neural network architectures (DL), together with the significant increase in available computing capacity thanks to the use of greatly improved GPUs, have allowed their successful application in new fields of knowledge in which they were previously not capable of generating positive results. Specifically, the progress achieved in the treatment of large series of data, of the type of text in the form of sentences and paragraphs, has given rise to sentiment classification applications that have shown great accuracy in different applications [4-8].

Our work represents a further step in this type of techniques, developing a model that generalizes sentiment classifier neural networks to achieve a regression of professional profiles in six dimensions and showing a specific application of it to the field of Economics and Finance, specifically, to the study of boards of directors. These governing structures have been the object of study in recent decades from different areas of knowledge, especially Finance, due to their important supervisory and advisory role in the company. However, the empirical studies carried out are mainly based on structured data, traditionally leaving aside a large amount of unstructured information due to the difficulties involved in its analysis.

Using data extraction and cleaning techniques, it has been possible to get unstructured data from the Annual Corporate Governance Reports of listed companies published on the website of the National Securities Market Commission (CNMV). These data are freely accessible to the public and describe the professional profile of the independent directors of the boards of directors of listed Spanish companies, i.e. those directors who are not contractually linked to the company nor are they shareholders of the company, so they are appointed on the basis of their professional background.

* Corresponding author.

E-mail addresses: cvaca@infor.uva.es (C. Vaca), ftejerina@efc.uva.es (F. Tejerina), benjamin.sahelices@uva.es (B. Sahelices).

The main contributions of the work are the following. In the first place, a methodology for extracting, cleaning and normalizing public data has been implemented that is repeatable and applicable to other conceptual environments. Secondly, starting from a state-of-the-art RNN-LSTM model for sentiment classification, it has been modified to convert it into a regression architecture so that a classification of professional profiles can be carried out in six dimensions. Thirdly, a manual labeling of a very significant set of directors has been carried out, which has allowed training the neural network and obtaining a very exact model with good generalization capacity. Fourth, a complete validation of the developed DL model has been carried out, showing its accuracy and validity for the study to be carried out. Finally, all this has allowed us to outline for the first time the professional profile of the boards of directors of Spanish listed companies. In general terms, the results show that the attribute most valued by companies when appointing independent directors is experience as an executive or consultant, followed by a purely financial profile.

The structure of the document is as follows: section II presents the state of the art of ML techniques and their possible applications to the field of Finance and Economics in general; section III describes the dataset and the model developed; section IV justifies and shows the results of the application of the model to the study of the profile of independent directors, while section V presents the conclusions.

II. Related Work

The Machine Learning (ML) model based on computational neural networks (NN), proposed many years ago, has experienced a great advance in the last 10 years due to the drastic improvement in processing speed by GPUs and new models, normalization, regularization, data augmentation and optimization techniques, among others. In the area of natural language processing, very significant advances have been made through the use of recurrent neural network (RNN) models that can handle arbitrarily large context sequences ([9], [10]). One of the most popular RNN models are long short-term memory networks (LSTMs), which have a better learning rate thanks to dropout, activation regularization, and temporary activation regularization [11].

Natural language learning is based on numerical vector representations of the text that allow capturing precise syntactic and semantic relationships. In [6], [7] the Skip-gram model is presented, which manages to establish syntactic and semantic relationships within the text with great efficiency and quality. In this way, it is possible to represent words and phrases, allowing generalization to represent and analyze documents efficiently. Document modeling is the base for a wide set of possible applications, starting with automatic text classification that allows extracting semantic information from text ([1]-[4], [12]). The most popular model used for the analysis and extraction of semantic information from a text are recurrent neural networks (RNN), and specifically the long short-term memory networks (LSTM) model, sometimes working together with convolutional networks (CNN) [11], [1]. There are multiple applications of text document modeling, including the classification based on feelings, the latent topic representation and the generation of information on the corporate culture of organizations ([13], [5], [8], [14]).

The word embedding model is used in conjunction with NN models such as RNN and CNN to perform high-level semantic analysis. In [14] corporate culture information is extracted in five dimensions, obtaining a correlation between its main corporate culture characteristics and business results and other significant corporate events. This study allows an analysis of a large number of corporations and the generation of accurate corporate information, which could not be done with traditional techniques. In the same vein, Hansen et al. (2018) [15] study how transparency affects the deliberation of monetary policy makers on the Federal Open Market Committee by analyzing its meeting records. The paper makes a methodological contribution by introducing latent Dirichlet allocation (LDA), a machine learning algorithm for textual analysis, to economic research. In asset pricing, Chen et al. (2014) [16] conduct textual analysis of articles and commentaries that were posted by investors on popular social media platforms and find that views expressed in both of them predict future stock returns. Likewise, Boudoukh et al. (2019) [17] analyze firm-specific news to predict volatility. Textual analysis is also actively used in Bellstam et al. (2021) [18], which uses the LDA method to examine analyst reports to specify a firm's level of innovation. A comprehensive review of the main applications of ML in finance can be found at Aziz et al. (2021) [19]. In general, major ML applications are in algorithmic trading, risk management and process automation.

III. Dataset and DL Model

The main objective of this work is the application of ML techniques to the analysis of the professional profile of the independent directors of companies listed in the Madrid Stock Exchange. CVs in the period 2003-2020 are used in free and unstructured format represented as simple sequences of text. The Deep Learning (DL) RNN architecture [9] is the most suitable for identifying dependencies in input text sequences. This type of network suffers from the vanishing gradient problem, which consists in the fact that the first levels of the network have a much lower learning rate than that of the last levels. The Long Short-Term Memory (LSTM) variant in the RNN architecture [1], [11] enables the propagation of the gradient to the first levels, thus improving its learning. The combination of the RNN architecture together with LSTM is perfectly adapted to the problem proposed.

The training of the DL model has been carried out in three phases. The first phase consists of creating a self-supervising neural network trained to recognize the structures of sentences and paragraphs written in Spanish, since this is the language used in the CVs. Using this network as a starting point, a second phase of training is applied to polish the understanding of sentences and paragraphs specifically in the financial, management, economic and business sectors. The objective of the second phase is to achieve a fine tuning of the neural network that allows for better results for the language and vocabulary structures used in the CVs.

The third training phase consists of using the encoder generated in phase 2 to carry out supervised training with a dataset labeled by a human expert to generate a regression model that characterizes each person using six profiles. The encoder contains the learning information of the grammatical structures of complete sentences in Spanish with an organization and vocabulary specific to the financial and business world. Using a supervised learning strategy based on manual labeling of a representative set of data, a regression neural network capable of linking free-form written language structures with six parameters that classify the manager in six profiles is generated.

A. Boards of Directors Dataset and Labeling

In order to ensure that the results of this work are consistent and repeatable, the decision was taken to train all the networks from scratch using datasets generated by us and to give up the transfer learning of networks trained with unknown data and parameters. For the first phase, we start from a network initialized with random parameters which we train to acquire knowledge of sentences and paragraphs in Spanish. The basic dataset used has been Wikipedia in Spanish [20] since the variability of its data, thematic diversity, writing styles and the breadth of vocabulary are guaranteed. Training with the total data set is not possible since the total volume of texts in Spanish

Wikipedia is very high. For this reason, a random selection of 10% of the total texts has been made, thus allowing training with viable computational requirements.

For the second phase of training, eight books have been selected from the thematic fields of finance, economics, business management, auditing, accounting and consulting. The goal is for the network to generate learning structures for sentences and paragraphs in these areas, as well as the corresponding vocabulary. The purpose of this dataset is to make a fine adjustment to our neural network, so that it can learn the differences between writing Wikipedia texts and scientific texts in the financial and managerial areas. It is expected to achieve a very efficient model in the prediction of words and phrases that constitutes a solid support for the third training phase [9].

Our proposal is related to the sentiment analysis and classification works [5], but we carry out a more detailed analysis using a regression model with six variables. The objective is to identify the profile of a director using the text of his CV provided in the company's public information. Based on sentiment classification systems, which have demonstrated their validity in different fields [5], their generalization is proposed to achieve a more detailed profile of directors. In phase three, six main activity profiles are identified: Financial (F), Executive/Consultant (E/C), Audit/Tax/Accountant (A/T/A), Legal (L), Political (P) and Academic (Ac). Based on these profiles, a regression-based model capable of assigning a numerical value to each of the profiles for each director is proposed.

The dataset consists of 137 firms listed on the Continuous Market of the Madrid Stock Exchange. Those belonging to the financial sector have been excluded, since their specific activity and regulations affect many characteristics of their governance, including the composition of their boards. The study covers the period 2003-2020. A total of 6561 director's profiles were analyzed. Most of the companies did not remain for the entire period but they were delisted at a given time or were successively incorporated.

The description of the directors' profiles (CVs) were taken from the Annual Corporate Governance Reports of the listed companies, which can be consulted on the website of the Comisión Nacional del Mercado de Valores (CNMV) in pdf format. Over the years, it can be seen that the description of the profiles has been expanding. In the early years, the descriptions were very brief or simply non-existent. Likewise, there are certain differences among companies in the degree of detail of such descriptions.

Of the 6561 director profiles, 1042 have been selected to obtain a training set on which manual labeling is performed, which is necessary for the supervised training stage of the regression model. The criterion for selecting the training set is a uniform distribution between the different types of profiles. Of these 1042 profiles, 100 have been separated, chosen randomly, to obtain a validation set that does not participate in the training and that allows us to perform a validation of the results. Additionally, the training set is internally divided into two groups to perform a standard training with a validation phase at the end of each epoch. This is used solely to measure the quality of the network's learning during its training and to make decisions about hyperparameters.

The human expert has labeled each director profile of the training dataset using the information in its public CV. For each of the six profiles, a number between 0 and 1 is assigned using a single decimal digit, that is, a standard assessment between 0 and 10 is made that estimates the weight of the corresponding profile within the CV. These profiles have not been considered mutually exclusive so maximum values (1.0) can be obtained in multiple and even in all profiles. For this reason, the criterion used in the labeling consists of assessing the CV data in a balanced way between the different directors and not making a percentage allocation. This allows a more objective and comparable assessment between the different directors. Given that the source of the data is a CV in the form of unstructured free text and in which the expression of similar merits can come from written expressions with very different structures and vocabulary, the labeling must be interpreted as an indicative fact of the professional profile and not as an exact numerical value.

B. DL Model and Training

The target of our DL proposal is to develop a statistical model of the language to estimate the distribution probabilities of the variations of linguistic units such as words, sentences, paragraphs and sets of paragraphs. This statistical model is built in two stages, the first for the training of the Spanish language and the second for the fine tuning of specific terminology. Taking this statistical model, we develop a second regression model that will allow us to carry out the multivariable profiling of the directors. To build the statistical model, a Recurrent Neural Network (RNN) with Long Shor-Term Memory (LSTM) [9] has been used. The main limitation of RNN models is that their learning capacity is small, which generates very long training times. To reduce this problem and increase the size of the analyzed context, the Long Short-Term Memory (LSTM) architecture is applied [1], [11]. This architecture uses memory units capable of storing information from multiple training sequences and, therefore, they are able to relate the information extracted from elements that are far apart from each other within the text sequence. This variant improves training times and greatly increases the length of the parsed context.

In the first phase, an RNN-LSTM network has been used to train the Wikipedia dataset and generate a base model with knowledge of linguistic structures in Spanish. In the second phase, this model has been refined using multiple books in the field of economics and finance. The encoder of this model has been used as a basic element for the training of a new RNN that has six regression outputs, each one to describe the profile of the directors, as explained in section 3.1.

As can be seen in the Figs. 1- 3 that describe the training in phases one (Wikipedia), two (Economics) and three (Regression), the network converges in a stabilized way. In phase one, no overfitting was observed, with the accuracy and perplexity values being 0.37 and 20, respectively, which shows a good relationship between the learning level and the training time. In phase two, overfitting appears, which was even more pronounced in previous training sessions and was reduced by applying a 30% dropout. Finally, in phase 3, a stable training with a reasonable value of RMSE was observed. Using a low-end GPU, the total computing time required to perform all three training phases was 120 hours.

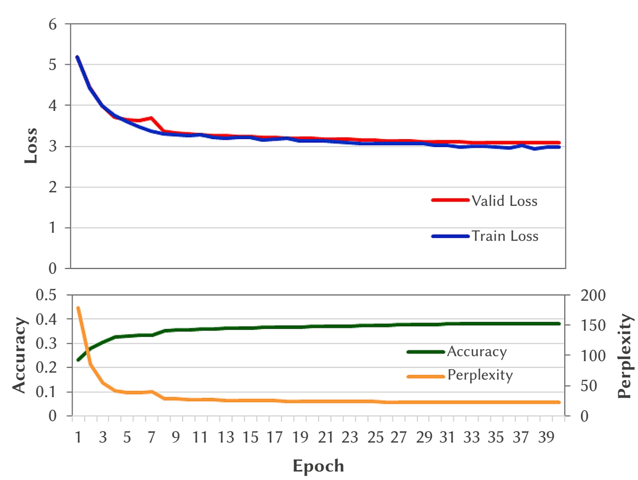

Fig. 1. Wikipedia Base Training.

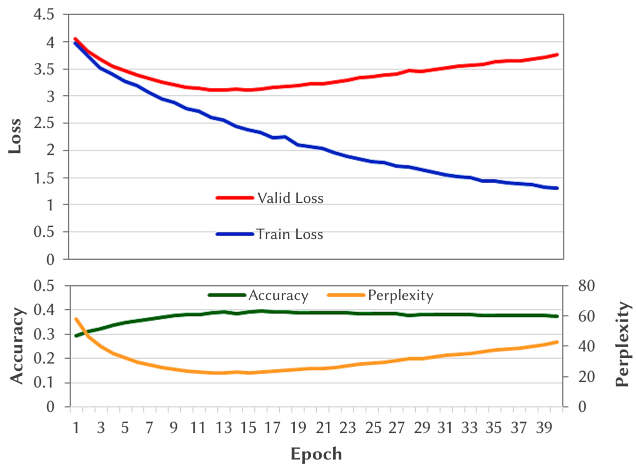

Fig. 2. Economy Base Training.

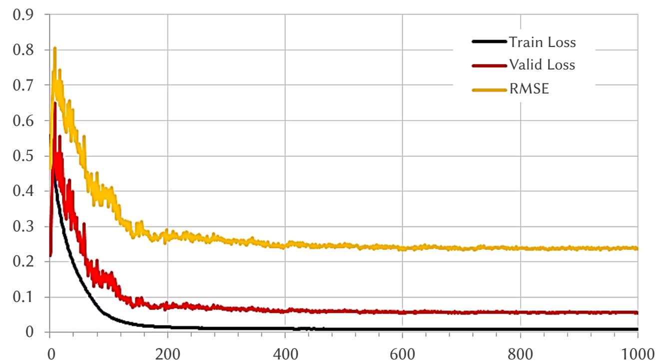

Fig. 3. Training for Regression.

C. Model Validation

To create a DL model that can be used to infer the total number of professional CVs available in the 2003-2020 period, it is necessary to carry out a prior validation phase that allows its behavior to be parameterized for correctly interpreting its results. In this sense, it is necessary to remember that the dataset used for the training, and therefore also in the inference, is an unstructured free text that describes the CV of the directors. Therefore, it is information of an imprecise nature which makes it difficult to generate results. The labeling of the dataset carried out by the human expert is also approximate and although numerical values are generated, these must be interpreted approximately as the quantification of the merits of a person referring to a specific profile.

We propose a validation process in two steps. In the first one, a numerical modeling of the model is carried out using labeled data which the network has not used for training. In this way it is possible to measure how this model mimics the behavior of the human expert. In the second step, a quantification scenario is designed to assign each of the six profiles to a category to measure the degree of accuracy with which our DL model identifies the profile of a director.

To carry out this validation stage, 100 professional profiles were randomly selected from the 1,042 that were labeled by the human expert. These 100 labeled CVs have not been used neither in training nor in the test carried out in each epoch. In other words, they have not participated in any way in the training of our DL model, nor have they affected decision-making about the hyperparameters. Therefore, this dataset allows us to carry out a clean and objective validation of our model.

The numerical behavior model of our neural network can be seen in Tables I and II which show the correlation analysis obtained by using the training and validation data respectively. As expected, our validation data has a higher error rate than the training data and provides us with a reliable indication of how well our network is able to generalize the trained knowledge. Our neural network model performs trend analysis in which the results are expressed as the average of large amounts of data corresponding to multiple profiles. As can be seen in Table II, the value of the standard error of the mean is small enough (between 0.015 and 0.033) to validate that the inference errors will in no case mask the results obtained. When performing the analysis with the 6561 profiles that make up our dataset, the value of the standard error of the mean is expected to be reduced to half of the values shown in Table II.

TABLE I. Correlation Analysis – Training Dataset

N = 942	*Professional Profile*					
	F	**E/C**	**A/T/A**	**L**	**P**	**Ac**
μ (error)	0.004	0.014	-0.008	-0.009	0.015	-0.003
σ (error)	0.097	0.129	0.083	0.099	0.046	0.092
corr. coef. (R)	0.971	0.962	0.957	0.961	0.986	0.962

TABLE II. Correlation Analysis – Validation Dataset

N = 100	*Professional Profile*					
	F	**E/C**	**A/T/A**	**L**	**P**	**Ac**
μ (error)	0.025	0.012	-0.037	-0.011	0.025	0.016
σ (error)	0.213	0.325	0.190	0.177	0.147	0.190
corr. coef. (R)	0.859	0.743	0.680	0.872	0.825	0.845
std. error of μ	0.021	0.033	0.019	0.018	0.015	0.019

Given the imprecise nature of the data used to train the network and given that the labeling performed by the human expert is, for the same reason, approximate, obtaining exact data from our model cannot be considered as the main target of our model. However, it is possible to assign a category to each of the six profiles to approximate their evaluation using a smaller number of intervals. Thus, it is possible to establish the profile using categories for each of the six values obtained in the regression. The discrete set used is made up of four categories that correspond to the intervals [0, 0.3), [0.3, 0.5), [0.5, 0.7) and [0.7,1.0] that are indexed by the integers 0, 1, 2 and 3. This reduction in the number of states is very useful for a profile analysis in which the data is imprecise, that is, it has a strong subjective component. Table III shows the success rate of the network for the training and validation datasets in which it can be seen that average success for validation data is 82.8%.

TABLE III. Quantized Categories Hit Rate

	Professional Profile					
	F	**E/C**	**A/T/A**	**L**	**P**	**Ac**
Training Set	93.8%	93.0%	96.9%	96.3%	97.1%	95.3%
Validation Set	77.0%	66.0%	92.0%	95.0%	83.0%	84.0%

Fig. 4 shows the confusion matrices for the validation dataset. It can be seen that the central categories (1 and 2) are much less represented than the limit categories (0 and 3), since in general the evaluation of a profile tends to values of the all/nothing type and the intermediate values are relatively infrequent. The results shown in this analysis validate our DL model for making scientific analysis of the profiling of boards of directors.

F profile

Infered \ Labelled	0	1	2	3
0	53	0	2	0
1	5	2	4	2
2	0	0	5	6
3	1	1	2	17

E/C profile

Infered \ Labelled	0	1	2	3
0	31	0	1	3
1	9	0	1	5
2	3	0	0	6
3	3	0	3	35

A/T/A profile

Infered \ Labelled	0	1	2	3
0	91	1	1	3
1	0	0	1	1
2	0	0	0	1
3	0	0	0	1

L profile

Infered \ Labelled	0	1	2	3
0	82	0	0	2
1	1	0	0	0
2	0	0	1	1
3	1	0	0	12

P profile

Infered \ Labelled	0	1	2	3
0	78	3	1	1
1	5	1	1	0
2	2	0	0	3
3	0	0	1	4

Ac profile

Infered \ Labelled	0	1	2	3
0	73	3	0	1
1	3	2	1	2
2	0	0	2	4
3	1	0	1	7

Fig. 4. Confusion Matrix for each profile (validation dataset).

IV. Profiling of Boards of Directors

A. Why Directors' Profile?

Sustainable finance refers to the process of taking environmental, social and governance (ESG) criteria into account when making investment or business decisions. In fact, corporate governance (represented by the G in the previous acronym) has been the subject of attention, both from the legislative perspective with the proliferation of codes of good governance and other types of soft-regulation, and from the academic field, for several decades. One of the features most discussed in this area in the financial literature is the independence of boards of directors, since it is considered essential to carry out their control function, as well as their influence on different company variables related to performance and value creation. In this sense, variables such as the percentage of independent directors, the size of the board, the separation between the figures of the chairman and the CEO or the number of meetings are analyzed (Hermalin and Weisbach, 2003 [21]). All of these would fall within the framework of what has been called structural diversity of the board.

However, in recent years boards are criticized for their excessive complacency and inability to prevent corporate crises, which has contributed to a widening of the perspectives from which they are analyzed. Thus, increasing attention has been paid in both academic and regulatory circles to board characteristics that can influence the effectiveness of the decision-making process. Such characteristics include, among others, the age, education, gender or nationality of the directors, grouped under the heading of demographic diversity of the board.

However, until now mainly those aspects of diversity that can be measured with quantitative and structured data (age, gender, years on the board, etc.) have been analyzed, leaving aside, due to the difficulty of collecting and structuring the information, aspects of great importance such as the professional profile of the board members. As Kim (2021) [22] points out, the analysis of unstructured data is one of the main fields of application of ML in finance research, although so far it is a largely unexplored field. It is only in recent years that attempts to apply ML to finance research have begun to emerge.

B. Directors' Professional Profiles

The analysis conducted allows us to understand the profile of the boards of directors according to the professional experience of each board member. As we pointed out in a previous section, we defined up to six professional profiles that, to a greater or lesser extent, have been considered in previous literature. They are as follows:

- Financial (F): Refers to those directors with experience in the financial sector, whether in banking institutions, any type of investment companies or the stock market in general. Güner et al. (2008) [23] and Booth and Deli (1999) [24] document that board members with financial expertise significantly influence firm financing and investment decisions.
- Executive/Consultant (E/C): Directors who have held or are currently holding different types of management positions in other companies or have carried out outstanding advisory tasks. These directors may have experience in different business sectors and management positions. According to the literature, independent directors help enhance firm value with their industry experiences (Drobetz et al., 2018 [25]; Faleye et al., 2018 [26]).
- Audit/Tax/Accountant (A/T/A): In this case, these are directors with specific expertise in auditing, tax or accounting. For instance, firms with accounting experts sitting on their audit committees show stronger accounting conservatism (Qiao, 2018 [15]).
- Legal (L): Lawyers and legal experts are classified in this profile. Their importance is highlighted by works such as that of Krishnan et al. (2011) [27], which shows that the presence of directors with legal backgrounds is associated with higher financial reporting quality.
- Political (P): Refers to directors who have held or are holding public offices of various kinds, especially political posts. The political profile can contribute to extending the company's relationships for its own benefit, although its presence has also been linked to sub-optimal decision-making. Houston et al. (2014) [28] find that the cost of bank loans is significantly lower for companies with board members with political ties, while Azofra and Santamaría (2004) [29] warned of the harmful effect of the presence of politicians on the efficiency of Spanish savings banks.
- Academic (Ac): Finally, this refers to those directors with academic experience. In this regard, some papers highlight that academic directors play an important governance role through their advisory and supervisory functions, leading to increased R&D performance and investment (Francis et al., 2015 [30]; Xie et al., 2021 [31]).

When appointing independent directors, companies will look for those profiles that best suit their needs. In general, independent directors perform the dual function of supervising and advising the board. The company will appoint them taking into account their ability to perform such functions on the basis of their professional profile. We assign a value between 0 and 1 to each profile for each of the independent directors analyzed. To the best of our knowledge, this is the first time that an overview of boards of directors based on the defined specific profiles has been presented.

C. Temporal Analysis

First, we examine the evolution of these profiles over time. The average number of independent directors per company fluctuates between 3 and 4, slightly exceeding this value in the last three years. Moreover, the number of companies in the sample tends to increase over the years. As a result, in absolute terms, an increase is observed in all the profiles considered, although from 2011 the "Political" and

"Legal" ones stagnated. The "Executive/Consultant" profile, on the other hand, is the one that has experienced the most marked growth in absolute terms.

Fig. 5 shows the relative importance of each profile over time. For each year, the sum of the different profiles is 1 (data have been normalized). The predominant profile in Spanish boards is by far the "Executive/Consultant" profile, with values starting at 0.35 and reaching 0.46 in 2020. The second most prevalent profile is "Financial", which is experiencing a slight but steady decline. The importance of the other four profiles is significantly lower.

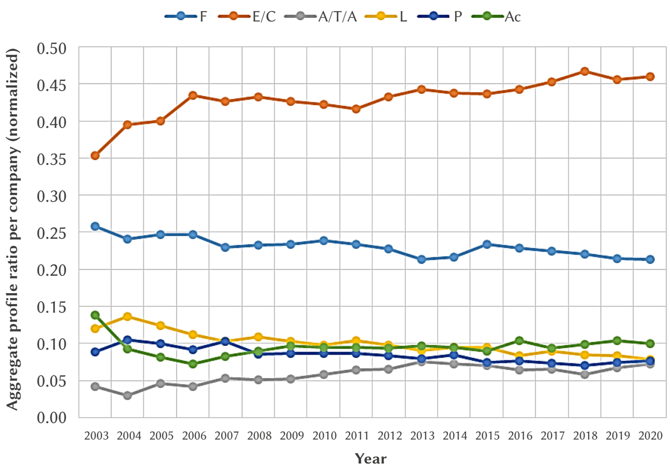

Fig.5 Aggregate Profile Ratios by Company per Year.

Regarding the evolution of each profile:

- "Executive/Consultant": after a pronounced increase, it experiences a slight decrease until 2011, from which a clearly upward trend is established.
- "Financial": the relative importance of this profile remains stable throughout the period, although with a downward trend. From an initial value of 0.26 to a final value of 0.21.
- "Legal" and "Political": of lesser relevance than the previous ones, they also experience a slight loss of importance throughout the period.
- "Academic": after a steep decline, it recovers from 2007 onwards with a stable trend.
- "Auditor/Tax/Accountant": this is the profile that has experienced the greatest increase over the years, although it is the profile with the lowest relative importance.

It should be noted that during the period in which the economic crisis was most pronounced (2007-2012), there was a certain stability in the "Financial" profile, with a slight upward trend. In turn, the "Audit/Tax/Accountant" profile experienced an increase in its relative importance.

D. Analysis by Sectors

To undertake this phase of the analysis we take the sectoral classification of the Madrid Stock Exchange, consisting of 7 sectors of activity (for the reasons given above, the financial sector, sector number 5, is not considered).

We reproduce the above analysis for each of the sectors considered (see Figs. 6-11). In all of them, the "E/C" profile remains the main one, followed by "F", although the relative importance of the latter varies for each sector.

TABLE IV. Sectors of the Madrid Stock Exchange

Sector 1	Oil and Energy
Sector 2	Basic Materials, Industry and Construction
Sector 3	Consumer Goods
Sector 4	Consumer Services
Sector 5	Financial Services
Sector 6	Technology and Telecommunications
Sector 7	Real Estate Services

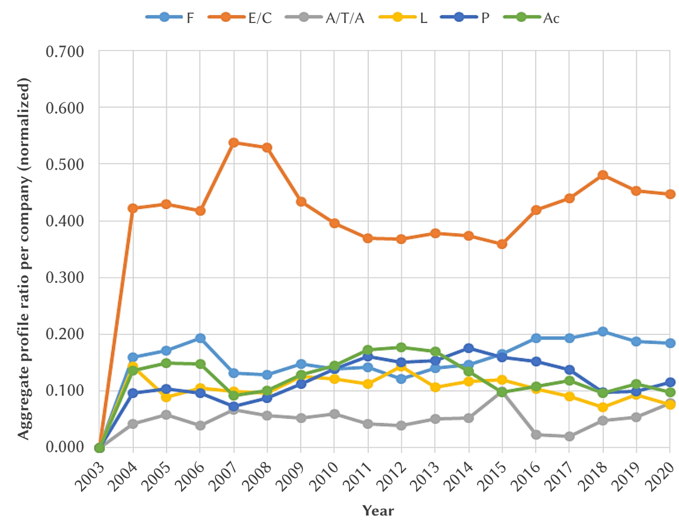

Fig. 6. Ratio of Aggregated Profiles by Company in Sector 1: The "E/C" profile is clearly the most important, although it declines sharply in the crisis period. It is noteworthy that the "F" profile is less important, especially during the crisis period, when it is overtaken by the "Ac" and "P" profiles, which grew considerably during this period. From 2012 onwards, the "F" profile became more important again, regaining second place in importance from 2015 onwards.

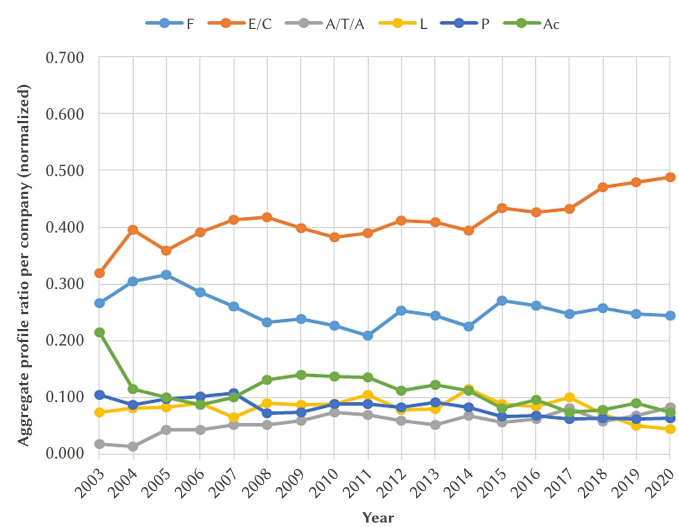

Fig. 7. Ratio of Aggregated Profiles by Company in Sector 2: Very similar to the general trend. There is a decrease in "F" during the crisis period and a slight recovery thereafter.

E. Discussion

The analysis carried out has allowed us to determine the profile of the independent directors with an acceptable degree of reliability. The preponderance of the "Executive/Consultant" profile seems to

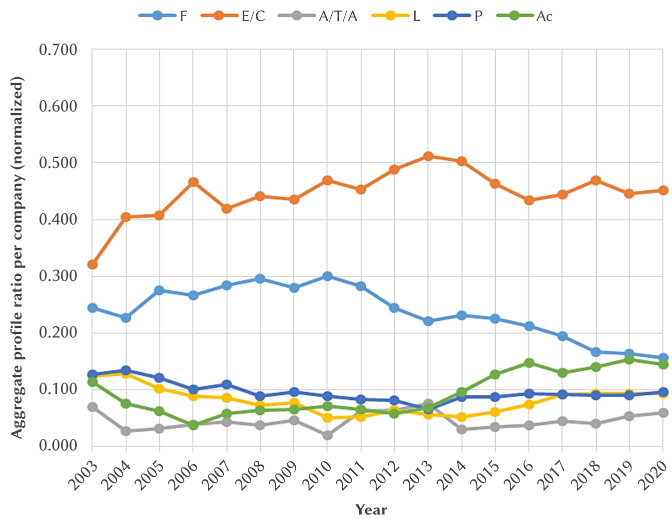

Fig. 8. Ratio of Aggregated Profiles by Company in Sector 3: Noteworthy is the sharp decline of "F" in the second part of the period in favor of the "Ac" profile, which from 2018 onwards practically equals the previous one.

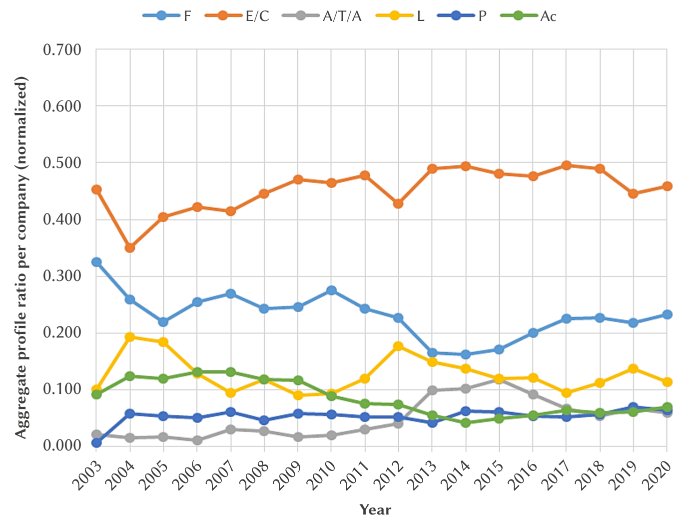

Fig. 9. Ratio of Aggregated Profiles by Company in Sector 4: The "E/C" profile continues to prevail. On the other hand, "F" has experienced a significant increase since 2015. In this sector we can highlight the relative importance of profile "L", whose relevance is significantly higher than for the total number of companies.

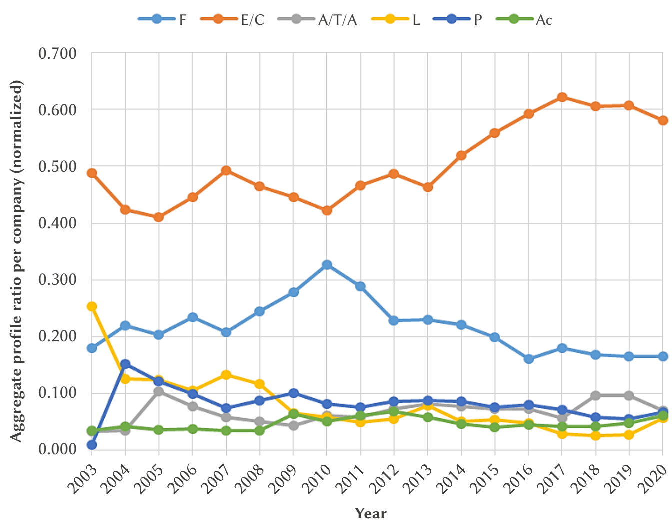

Fig. 10. Ratio of Aggregated Profiles by Company in Sector 6: After successive ups and downs, from 2010 onwards the "E/C" profile grows at the expense of "F". It is worth noting the initial importance of "L", which gradually decreases.

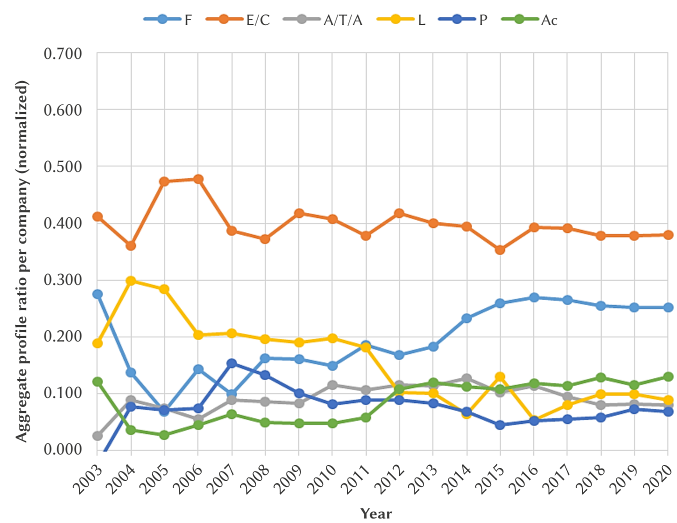

Fig. 11. Ratio of Aggregated Profiles by Company in Sector 7: From 2007 onwards, "F" increased, while "L" remained the second most important profile, ahead of "F" until 2011.

suggest that when appointing new directors, priority is given to their skills as advisors, giving less importance to their supervisory role. On the other hand, it would be expected that the "Financial" profile would increase during the years of greater economic crisis, given the financial problems it caused in a large number of companies. However, it can be seen that the relative importance of this profile remained stable and even showed a downward trend in the 2007-2012 period. Nevertheless, there was a strong increase in Sector 6 during the worst years of the crisis.

In this regard, the growing importance of the "Political" profile in Sector 1 during the crisis years is noteworthy. As this is a strategic and heavily regulated sector, the growing uncertainty caused by the crisis has encouraged the appointment of directors with this profile. Likewise, we highlight the role of "Academic" profile directors in Sector 3 since 2013. This is a profile that traditionally has not had much relevance, but which is recognized as being able to effectively perform the dual advisory and supervisory role required of boards of directors. Finally, the "Legal" profile stands out for its importance in different periods in sectors 4, 6 and 7, to some extent subject to legislative changes in the period under consideration. In this respect, it would be very interesting to look more deeply into the added value of these kinds of directors.

V. Conclusions

Advances in AI and, specifically, in DL allow for the development of learning models capable of analyzing information sequences of much greater length than previous methods were capable of successfully performing. This makes it possible to deal with complex problems in multiple areas of knowledge, and specifically, in the modeling of economic and financial data. Our work contributes to the progress of knowledge by presenting a model that in successive phases has been able to learn syntactic structures of natural language in Spanish from the area of economics and finance, and use said knowledge to generate a regression model that accurately models professional profiles of boards of directors in six dimensions through a technique that can be considered a generalization of previous work on sentiment classification.

The results obtained present a more than reasonable degree of reliability, being an example of the capacity of ML techniques in the treatment of unstructured economic-financial information. In this field the application of this methodology is still at an incipient stage. However, given that a large part of the economic data is automated and growing exponentially, the application of these techniques will expand the frontiers of research in Economics and Finance.

Future lines of research are emerging from this work. In the area of extracting information from sequential data of the unstructured natural text type, it would be very interesting to delve into other hyperparameters and new models to increase the accuracy of the regression allowing to address new more complex problems. In the field of Finance, it would be useful to deepen and broaden the profiles considered in this work, as well as to analyze the influence of the different profiles on the firms' value creation process.

Acknowledgment

We would like to thank Manuel Astorgano Antón for his help in processing the data. This work was supported by the Spanish Ministry of Science and Innovation (AEI), Grants ID. PID2019-105660RB-C21 and PID2020-114797GB-I00. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

[1] C. Zhou, C. Sun, Z. Liu and F.C.M. Lau, "A C-LSTM Neural Network for Text Classification," *The Journal of Machine. Learning Research*, vol. 12, pp. 2493-2537, 2015, doi: 10.48550/arXiv.1511.08630

[2] D. Yogatama, C. Dyer, W. Ling and P. Blunsom, "Generative and Discriminative Text Classification with Recurrent Neural Networks," 2017, arXiv 2017, 10.48550/arXiv.1703.01898

[3] A. Conneau, H. Schwenk, Y.L. Cun and L. Barrault, "Very deep convolutional networks for text classification," *15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017 - Proceedings of Conference*, Valencia, Spain, vol. 2, 2017, doi: 10.48550/arXiv.1606.01781

[4] S. Lyu and J. Liu, "Convolutional Recurrent Neural Networks for Text Classification," *Journal of Database Management*, vol. 32, no. 4, pp. 65-82, 2021, doi: 10.4018/JDM.2021100105

[5] Y. Kim, "Convolutional neural networks for sentence classification," *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, Doha, Qatar, 2014, doi: 10.3115/v1/d14-1181

[6] T. Mikolov, K. Chen, G. Corrado and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality," *NIPS'13: Proceedings of the 26th International Conference on Neural Information Processing Systems*, Tahoe, NV, United States, vol. 2, pp. 3111-3119, 2013, doi: 10.48550/arXiv.1310.4546

[7] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," *31st International Conference on Machine Learning, ICML 2014*, Beijing, China, vol. 4, 2014, 10.48550/arXiv.1405.4053

[8] B. Jiang, Z. Li, H. Chen and A.G. Cohn, "Latent Topic Text Representation Learning on Statistical Manifolds," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 11, pp. 5643-5654, 2018, doi: https://doi.org/10.1109/TNNLS.2018.2808332

[9] W. De Mulder, S. Bethard and M. F. Moens, "A survey on the application of recurrent neural networks to statistical language modeling," *Computer Speech and Language*, vol. 30, no. 1, pp. 61-98, 2015, doi: 10.1016/j.csl.2014.09.005

[10] M.G. Huddar, S.S. Sannakki and V.S. Rajpurohit. "Attention-based multi-modal sentiment analysis and emotion detection in conversation using RNN," *International Journal of Interactive Multimedia and Artificial Intelligence*. 2021;6(6):112-121. doi: 10.9781/ijimai.2020.07.004

[11] S. Merity, N.S. Keskar and R. Socher, "Regularizing and optimizing LSTM language models," *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, Vancouver, Canada, 2018, doi: 10.1109/TNNLS.2018.2808332

[12] P. Giudici, "Fintech Risk Management: A Research Challenge for Artificial Intelligence in Finance," *Frontiers in Artificial Intelligence*, vol. 1, 2018, doi: 10.3389/frai.2018.00001

[13] K. Li, F. Mai, R. Shen and X. Yan, "Measuring Corporate Culture Using Machine Learning," *The Review of Financial Studies*, vol. 34, no. 7, pp. 3265-3315, 2021, 10.1093/rfs/hhaa079

[14] S. Hansen, M. McMahon and A. Prat. "Transparency and Deliberation within the FOMC: A Computational Linguistics Approach," *The Quarterly Journal of Economics*, vol. 133, no. 2, pp. 801-870, 2018, doi: 10.1093/qje/qjx045.

[15] Z. Qiao, K.Y. Chen and S. Hung. "Professionals inside the board room: accounting expertise of directors and dividend policy," *Applied Economics*, vol. 50, no. 56, pp. 6100-6111, 2018, doi: 10.1080/00036846.2018.1489501.

[16] J. Boudoukh, R. Feldman, S. Kogan and M. Richardson. "Information trading and volatility: Evidence from firm-specific news," *Review of Financial Studies*, vol. 32, no. 3, pp. 992-1033, 2019, doi: 10.1093/rfs/hhy083.

[17] G. Bellstam, S. Bhagat and J.A. Cookson. "A text-based analysis of corporate innovation," *Management Science*, vol. 6, no. 7, pp. 4004-4031, 2021, doi: 10.1287/mnsc.2020.3682.

[18] S. Aziz, M. Dowling, H. Hammami and A. Piepenbrink. "Machine Learning in finance: A topic modelling approach," *European Financial Management*, pp. 1-27, 2021, doi: 10.1111/eufm.12326.

[19] https://es.wikipedia.org/wiki/Wikipedia

[20] J. Howard and S. Ruder, "Universal Language Model Fine-tuning for Text Classification," *ACL. Association for Computational Linguistics* 2018, doi: 10.48550/arXiv.1801.06146

[21] B. E. Hermalin and M.S. Weisbach. "Boards of directors as an endogenously determined institution: A survey of the economic literature," *FRNBY Policy Review*, vol. 9, no. 1, pp. 7-26, 2003. doi: 10.2139/ssrn.233111

[22] H. Kim. "Machine Learning Applications in Finance Research," in *Fintech with Artificial Intelligence, Big Data, and Blockchain*, P.M.S. Choi and S.H. Huang Ed. Singapore: Springer, 2021, pp. 205-220.

[23] B. Güner, U. Malmendier and G. Tate. "Financial expertise of directors," *Journal of Financial Economics*, vol. 88, no. 2, pp. 323-354, 2008, doi: 10.1016/j.jfineco.2007.05.009.

[24] J.R. Booth and D.N. Deli. "On executives of financial institutions as outside directors," *Journal of Corporate Finance*, vol. 5, no.3, pp. 227-250, 1999, doi: 10.1016/S0929-1199(99)00004-8.

[25] W. Drobetz, F. Von Meyerinck, D. Oesch and M.M. Schmid. "Is director industry experience a corporate governance mechanism?," *SSRN Electronic Journal*, 2018, doi: http://dx.doi.org/10.2139/ssrn.2256477

[26] O. Faleye, R. Hoitash and U. Hoitash. "Industry expertise on corporate boards," *Review of Quantitative Finance and Accounting*, vol. 50, pp. 441-479, 2018, doi: 10.1007/s11156-017-0635-z.

[27] J. Krishnan, Y. Wen and W. Zhao. "Legal Expertise on Corporate Audit Committees and Financial Reporting Quality," *The Accounting Review*, vol. 86, no. 6, pp. 2099-2130, 2011, doi: 10.2308/accr-10135.

[28] J.F. Houston, L. Jiang, C. Lin and Y. Ma. "Political Connections and the Cost of Bank Loans," *Journal of Accounting Research*, vol. 52, no. 1, pp. 193-243, 2014, doi: 10.1111/1475-679X.12038

[29] V. Azofra and M. Santamaría. "El gobierno de las cajas de ahorro españolas," *Universia Business Review*, vol. 2, no. 2, pp. 48-59, 2004. Available at: https://www.redalyc.org/articulo.oa?id=43300204.

[30] B. Francis, I. Hasan and Q. Wu. "Professors in the Boardroom and Their Impact on Corporate Governance and Firm Performance," *Financial Management*, vol. 44, no. 3, pp. 547-591, 2015, doi: 10.1111/fima.12069.

[31] Y. Xie, J. Xu and R. Zhu. "Academic Directors and Corporate Innovation," 2021, Available at SSRN: https://ssrn.com/abstract=3954290.

César Vaca

César Vaca has a degree and a master's degree in physics from the Complutense University of Madrid. He has been a professor and researcher at the University of Valladolid since 1992. He is an expert in software development and complex data analysis. He is the author of the software environment that manages, organizes, optimizes and provides support for intelligent decision-making about all human medical resources in the Autonomous Community of Castilla y León in Spain. He has carried out multiple collaborations in the area of industry for the automation of highly complex systems, the creation of digital twins in the field of intelligent manufacturing and the application of Deep Learning in automatic quality control in industrial environments.

Fernando A. Tejerina

Fernando A. Tejerina Gaite, Ph.D., is an Associate Professor of Financial Economics at University of Valladolid (Spain). He has been Researcher Visiting Fellow at Leeds University Business School (UK), the University College Dublin (Ir), the Cass Business School (UK) and the University of Edinburgh (UK). His research is related to corporate governance, international business, family firms and innovation. He has earned research awards on five international organizations (University of Monterrey–Mx., JAAB-New York, Global Strategic Management Inc, European Journal of Arts and Sciences and Financial Studies Foundation). He has published several books and books chapters, and many papers in national and international journals. He has reviewed papers for many journals and conferences.

Benjamín Sahelices

Benjamín Sahelices has a PhD in computer science from the University of Valladolid, where he has been a professor and researcher since 1991. He has carried out stays and collaborations with research teams from the University of Tennessee in Knoxville, the University of Illinois at Urbana-Champaign and the University of Edinburgh, as well as in different Spanish universities. His main areas of interest are high performance computing, new memory hierarchies' designs, new paradigms of processing architectures, heterogeneous computing for deep learning and its applications. He has published in prestigious journals and conferences in the field of computer architecture. He regularly collaborates with research groups and companies in the area of industrial manufacturing, astronomy, economics, finance and biomedical engineering. He is a member of the GCME research group at the University of Valladolid.