Future Generation Computer Systems 102 (2020) 838–846

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

# Enhancing the classification of social media opinions by optimizing the structural information

Carla Vairetti [a,*], Eugenio Martínez-Cámara [c,**], Sebastián Maldonado [a,b], Victoria Luzón [c], Francisco Herrera [c,d]

[a] Facultad de Ingeniería y Ciencias Aplicadas, Universidad de Los Andes, Mons. Álvaro del Portillo 12455, Las Condes, Santiago, Chile
[b] Instituto Sistemas Complejos de Ingeniería (ISCI), Santiago, Chile
[c] Andalusian Research Institute in Data Science and Computational Intelligence, University of Granada, 18071 Granada, Spain
[d] Faculty of Computing and Information Technology, King Abdulaziz University Jeddah, Saudi Arabia

ARTICLE INFO

*Article history:*
Received 15 May 2019
Received in revised form 26 July 2019
Accepted 14 September 2019
Available online 24 September 2019

*Keywords:*
Online review
Sentiment analysis
Support vector machines
Weighting optimization

ABSTRACT

Sentiment Analysis is an extensively studied task, however an important aspect yet to study is the underlying structural information of opinions. An important aspect to tackle is the analysis underlying structural information of opinions. Social media is a great source of user opinions, which are structured in most of the cases in two sections: the title and the content or body of the opinion. We claim that the structure of social media opinions has useful information for the polarity classification task. We propose a model for optimizing the contribution of that underlying structural information for polarity classification. Our model is built by weighting the contribution of each section, title and body. We develop a modified Support Vector Machine that includes a weight parameter, which is optimized via a line-search strategy. We evaluate our proposal on three datasets of reviews from different domains written in two different versions of the Spanish language. The results show that our model outperforms the classification of the joint or individual classification of each section of the opinion. Therefore, our claim holds.



## 1. Introduction

Since the advent of the Web 2.0, the barrier among the producers and consumers of content on the Internet was broken, which means that any user may read and post any kind of content. One of the kind of content published on the Internet is opinions, which are really valuable information for users and for the target organization or topic of the opinion. Accordingly their study and their automatically treatment is crucial. Sentiment Analysis (SA) is the computational task centered on the computational treatment of subjective, sentiment and opinions in texts [1].

The main source of opinions are social media, social networks and microblogging sites, because they are mainly based on the communication and exchange of information and experiences between users. Accordingly, Cambria and Hussain [2] updated the definition of SA as the set of computational techniques for extracting and classifying opinions in user-generated text from online sources. Social networks related to a specific domain, like TripAdvisor,[1] Booking,[2] IMDB[3] or Patient Opinions,[4] allow their users to publish opinions about the target topic of the social network. Likewise, those social networks allow to organize the opinion in two sections: (1) title, which is an excerpt of the opinion; and (2) the entire opinion, which we call the body of the opinion. Users usually highlight and vehemently express their view in the title, and then they detail it on the body of the review. For instance, Martínez Cámara et al. [3] showed the usefulness of processing the title of the review for classifying the sentiment meaning of the entire review. Accordingly, we make two research questions:

1. Is it possible to weight the relevance of each part of the structure of the review?
2. Is it possible to optimize the weight of each part of the review?

We answer the two questions by proposing a modified version of the Support Vector Machine algorithm, which consists of including a new parameter $\gamma$ for weighting the contribution of each

* Corresponding author.
** Corresponding author.
*E-mail addresses:* cvairetti@uandes.cl (C. Vairetti), emcamara@decsai.ugr.es (E. Martínez-Cámara), smaldonado@uandes.cl (S. Maldonado), luzon@ugr.es (V. Luzón), fherrera@decsai.ugr.es (F. Herrera).

https://doi.org/10.1016/j.future.2019.09.023
0167-739X/

[1] https://www.tripadvisor.es/.
[2] https://www.booking.com.
[3] https://www.imdb.com/.
[4] https://www.patientopinion.org.au.

part of the structure of a review. The value of $\gamma$ is optimized by a line-search method. From a general point of view, our method is able to optimize the use of all the underlying information and knowledge of the reviews posted in review sites.

We assess our proposal on three corpora of opinions written in two versions of Spanish (Spain and Chile) and from three different domains: hotels, movies, and restaurants. Since the preprocessing may cause differences in the performance of the proposed model, we also evaluate our proposal with different preprocessing pipelines as we will describe in Section 4.2. The results show that our proposal is able to optimize the contribution of each section of the structure of a review, and it thus takes advantage of the underlying knowledge of reviews posted in online sources.

The reminder of the paper is organized as follow: Section 2 succinctly describe some prominent related works. Subsequently, we detail our proposal in Section 3. Sections 4 and 5 are focused on the experimental setup and the results and their analysis. Finally, the conclusions are exposed in Section 6.

## 2. Related works

Our evaluation is focused on reviews written in Spanish. We describe in Section 2.1 some works about SA in Spanish. Subsequently, in Section 2.2, we expose some definitions related to the use of SVM in Natural Language Processing (NLP) tasks and its use in SA.

### 2.1. Sentiment analysis in Spanish

SA, as other NLP tasks, has been mainly studied for texts written in English, which can be seen in [1,4]. However, there are some works focused on the processing of Spanish written texts. The first works were related to the elaboration of linguistic resources to work on Spanish data. Cruz et al. [5] presented the first corpus of Spanish reviews, specifically in the domain of movies reviews. Brooke et al. [6] described the adaptation of a non-supervised English polarity classification system to the classification of Spanish reviews, by means the translation of English sentiment lexicons into Spanish, and the adaptation of some rules for the treatment of negation, gender, and plural in Spanish. Since the adaptation by hand of resources is not an efficient approach, Perez-Rosas et al. [7] propose an automatic method for projecting English sentiment lexicons into other languages. Due to the method is built upon SentiWordNet [8] and WordNet [9], the method can be used to all those languages that have non-commercial WordNet versions, like Spanish.

The release of new resources for Spanish SA has not been ceased, for instance, the Spanish multi-domain version of the SFU corpus [6], the same corpus with annotations at negation level [10] and the COPOS corpus in the health services review domain [11]. Likewise, new sentiment lexicons have been released such as iSOL [12] or ML-SentiCon [13], as well as methods for automatically adapting those lexicons to different domains [14].

Concerning the classification of Spanish reviews, we highlight the work [3], because it studied the effect of classifying the different parts of a opinion, which is close to our proposal. The authors concluded that only taking into account the title allows reaching acceptable results and close to the ones reached when all the parts of the review are used. In a similar line, Taboada et al. [15] studied different parts of the contents of a review and identified what are the most useful for polarity classification. However, they do not use the structure of the most used review web sites like TripAdvisor, as we do in this paper, and they performed a manual annotation of different discourse sections of a review and the kind of linguistic information that they provided.

Micro-blog sites like Twitter are other instance of social media sites where users post reviews. We also find some works related to Spanish SA on Twitter. The first works were also related to the description and release of new corpora of tweets, for instance, the COST corpus [16], the General TASS corpus [17] and Inter-TASS corpus [18]. The former two corpora were released by the organization of the TASS workshop,[5] which has attracted the attention of the research community by means the organization of a competition built upon the corpora released. The main features of the participant systems in TASS are in [19].

There are more works related to SA in Spanish and if the reader is interested, we recommend to read [20]. However, as far as we know, there are not any work that conduct an analysis of the structure of the opinions from online sources. Therefore, we contribute with a method that optimizes the joint classification of the different sections of an online opinion, which means that our proposal combines linguistic and structure information for enhancing the performance of SA classification systems.

### 2.2. SVM for natural language processing

SVM has been successfully applied in various domains, including computer vision, medical diagnosis, bioinformatics, NLP, among others. Below we discuss some studies that highlight the virtues of SVM classification applied to NLP.

Understanding the relations between chemicals and diseases is relevant in areas such as biomedical research and health care. As for machine learning-based relation extraction (RE), feature-based methods [21–23] and kernel-based methods [24,25] are widely used. Feature-based methods focus on designing effective features including lexical, syntactic and semantic information. However, the traditional lexical and flat syntactic features are 'one-hot' representations, which could not adequately capture the deep semantic and syntactic structure information. In particular the work of Kin and Liu [26] proposed an efficient and scalable system using a linear kernel for drug–drug interaction (DDI) extraction where linear SVM are competitive when equipped with rich lexical and syntactic features.

In the language areas, finding a way to tag every word in a text as a particular part of speech can be done through POS tagging. POS tagging is a very important preprocessing task for language processing activities. POS taggers were proposed for various Indian languages, such as Hindi, Punjabi, Malayalam, Bengali, and Telugu. Linear SVM classification has been used for POS tagging [27]. For example, for Malayalam authors proposed [28] a part-of-speech Tagger for Malayalam language using SVM classification. Their objective was to identify the ambiguities in Malayalam lexical items, and to develop a tag set appropriate for Malayalam. IIT Mumbai developed POS tagging for the Marathi language [29]. In the case of Bengali Language, an SVM-based tagger was proposed in [30].

Text classification via machine learning methods has been widely studied in the recent literature. Sivakumar et al. [31] proposed a hybrid text classifier using $k$-NN and SVM to reduce the parameter impact in classification accuracy. Hassan et al. [32] proposed a method for text categorization in which they compared linear SVM and naïve Bayes. Tsikerdekis et al. [33] proposed a novel approach for multiple identity deception using non-verbal behavior. Srivasatava [34] examined the impact of NLP features (stop words, stemmer and combination of both) on predictive performance of base classifiers and ensembles of Naive Bayesian category.

Many NLP studies have been developed based on Twitter data, such as twitter sentiment analysis, twitter trending topic

[5] http://www.sepln.org/workshops/tass/.

detection, twitter sentiment classification. Several authors have compared machine learning algorithms such as SVM, decision trees, or naïve Bayes, concluding that SVM performed best (see e.g. [35–38]). For example, Li et al. [39] presented an empirical study of skip-gram features for large scale sentiment analysis. To promote model-efficiency and prevent overfitting, authors used feature selection in order to demonstrate the utility of logistic regression incorporating both L1 regularization and L2 regularization for weight distribution.

## 3. Optimizing the weight of each section of the opinion

In this paper we claim that the underlying structural information of social media opinions has useful knowledge for polarity classification. Accordingly, we argue that the linear combination of the features of the different sections of a review, in our case the title and the body, may enhance the classification of the sentiment meaning of reviews posted on social media. Hence, we propose the use of a linear classification system to linearly combine the contribution of each section of the review, and the optimization of that contribution by means the incorporation of a weighting parameter.

In this section we present our proposal, which is built upon SVM and described in Section 3.1. Subsequently, we detail our proposed modification of SVM in Section 3.2 for allowing the weighted contribution of each section of the review.

### 3.1. Support vector machine

We selected SVM [40] as the linear classification algorithm to build our proposal because of the advantages indicated by Maldonado and Weber [41] regarding driving better empirical results compared to other statistical and machine learning approaches. SVM provides theoretical advantages, such as adequate generalization to new objects, thanks to the Structural Risk Minimization (SRM) principle, absence of local minima via convex optimization, and representation that only depends on a few parameters [40].

For the linearly separable case, the SVM determines the optimal hyperplane that separates the convex hulls of both training patterns. The standard SVM aims at finding a classifier of the form $f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b$ that maximizes the distance from it to the nearest training point on each class (the margin). To maximize this measure, SVM minimizes the Euclidean norm of coefficients $\mathbf{w}$ [40]. Additionally, we intend to classify the training vectors $\mathbf{x}_i$ correctly into two different classes $y_i$.

The $l_1$-SVM model [42] is a very efficient SVM extension in which the Euclidean norm is replaced by the $l_1$-norm. Let $\{(\mathbf{x}_i, y_i)\}_{i=1}^m$ be a set of training instances $\mathbf{x}_i \in \Re^n$ with labels $y_i \in \{-1, +1\}$, for $i = 1, \ldots, m$. $l_1$-SVM finds a hyperplane of the form $\mathbf{w}^\top \mathbf{x} + b = 0$ by solving the following problem:

$$\begin{aligned} \min_{\mathbf{w},b,\boldsymbol{\xi}} \ & \|\mathbf{w}\|_1 + C \sum_{i=1}^{m} \xi_i \\ \text{s.t. } & y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i, \ i = 1, \ldots, m, \\ & \xi_i \geq 0, \qquad i = 1, \ldots, m, \end{aligned} \tag{1}$$

where $C > 0$ controls the trade-off between the $l_1$-regularization and model fit, and $\boldsymbol{\xi}$ denotes the vector of slack variables related to each sample. This strategy for model fit is known as hinge loss. The $l_1$-norm or LASSO penalty encourages sparsity, finding a good compromise between complexity reduction and feature selection [42]. Furthermore, Eq. (1) can be cast into a linear programming problem, which can be efficiently solved using coordinate descent methods [43]. Therefore, this method is suitable for large, sparse datasets, such as the ones related to NLP problems.

### 3.2. Optimizing the contribution of each section

The main goal of the proposal is to optimize the contribution of each section of customer opinions (title and body) for the classification of the opinion meaning. We claim that a word that appears in the title of the comment is more important than one in the body, and the title–body aggregation leads to a loss of information. Our proposal is based on the $l_1$-SVM model [42], and we propose optimizing the contribution of each section via a variable $\gamma$ that upweights the words in the title, i.e. $\mathbf{x}(\gamma)_i = \gamma \mathbf{x}_i^t + \mathbf{x}_i^b$. The usual approach is setting $\gamma = 1$, i.e. each word is counted regardless if it appears in the title or the body. This means that the usual approach do not prioritize when a word appears in the title over the body.

Given training samples of the form $\mathcal{D} = \{(\mathbf{x}_1^t, \mathbf{x}_1^b, y_1), \ldots, (\mathbf{x}_m^t, \mathbf{x}_m^b, y_m)\}$, where $\mathbf{x}_i^t \in \Re^{m \times n}$ and $\mathbf{x}_i^b \in \Re^{m \times n}$ are matrices containing the information of the comments for the title and body, respectively, and $y_i = \{-1, 1\}$ are their respective labels, for $i = 1, \ldots, m$. We assume that the ranking prediction problem can be cast into a binary classification task by defining a threshold.

We propose to jointly optimize the variable $\gamma$ with the hyperplane in the SVM model, extending Eq. (1) with Eq. (2):

$$\begin{aligned} \min_{\mathbf{w},b,\boldsymbol{\xi},\gamma} \ & \|\mathbf{w}\|_1 + C \sum_{i=1}^{l} \xi_i \\ \text{s.t. } & y_i(\mathbf{w}^\top \mathbf{x}(\gamma)_i + b) \geq 1 - \xi_i, \ i = 1, \ldots, l, \\ & \xi_i \geq 0, \qquad i = 1, \ldots, l. \end{aligned} \tag{2}$$

The issue of directly optimizing $\gamma$ is that $l_1$-SVM has a highly-optimized implementation (LIBLINEAR [43]), which makes it especially suitable for large datasets. The inclusion of $\gamma$ would not allow the use of the efficient coordinate descent algorithm proposed in Fan et al. [43]. We propose a line search optimization method to find the most adequate value for $\gamma$. Specifically, the search space for $\gamma$ is defined by the set $\gamma \in \{0.25, 0.5, 0.75, 1.0, 2.0, 2.5, 3, 3.5, 4, 4.5, 5\}$. In order to avoid overfitting, we used a ten-fold-cross-validation approach as strategy for setting $\gamma$. The training of the SVM in Eq. (2) is formally defined in Eq. (3).

$$\begin{aligned} \min_{\gamma} \min_{\mathbf{w},b,\boldsymbol{\xi}} \ & \frac{1}{2} \|\mathbf{w}\|_1 + C \sum_{i=1}^{l} \xi_i \\ \text{s.t. } & y_i(\mathbf{w}^\top \mathbf{x}(\gamma)_i + b) \geq 1 - \xi_i, \ i = 1, \ldots, l, \\ & \xi_i \geq 0, \qquad i = 1, \ldots, l. \end{aligned} \tag{3}$$

## 4. Experimental results

In this Section we detail the set up of the evaluation of our proposal. First, we describe the data used in the assessment in Section 4.1. Subsequently, in Section 4.2, we describe the preprocessing pipeline run for preparing the data, and in Section 4.3 we expose the configuration of our proposal.

### 4.1. Data set

Since SA is a highly domain-dependent task, we conducted the evaluation on three datasets of reviews from three different domains, specifically (1) movies, (2) hotels, and (3) restaurants. Moreover, SA, as other NLP tasks, is determined by the use of language. The Spanish language is the native language of Spain and most of the countries of America. Although the spoken and written Spanish versions are the same language, there are some lexical and use of language differences that may mislead a classification system. Accordingly, we conducted an evaluation with two different written versions of Spanish, specifically written

Spanish in Spain and Chile. Hence, we assess the consistency of our proposal on three different review domains and on two dissimilar written version of the same language.

The three datasets used in the evaluation have in common their structure, which is composed of two sections: title and body. The title is a short excerpt that summarizes the review, and the body is the entire review.[6] This structure allows us to study if our claim holds, namely the right combination of the subjective meaning of the title and the body will increase the performance of the polarity classification of the entire opinion. The datasets used are detailed as what follows.

*Corpus of opinions of andalusian hotels (COAH).* The COAH data set[7] [44] is composed of reviews of hotels from Andalusia, which is the most touristic region of Spain. The source of the opinions is the traveling site TripAdvaisor, and the language of the reviews is Spanish written in Spain. The reviews are from 80 different hotels, which assure that the opinions are not skewed to a specific hotel. The reviews are labeled on a scale of five levels of opinion intensity, from strong positive (5) to strong negative (1). The size of the corpus per label is 532 (5), 493 (4), 291 (3), 203 (2), 316 (1), in total 1835. Since we performed a two-class evaluation, we consider the labels 5 and 4 as positive, the labels 1 and 2 as negative, and we did not use the neutral intensity label 3 because it cannot be considered either positive or negative. The statistics of the COAH corpus used are in Table 1. Fig. 1 shows an instance from the training corpus.

*Spanish movie review corpus.* This corpus is also known in the literature as MuchoCine corpus[8] [5], and its reviews were gathered from the movie review site MuchoCine.[9] The language of the reviews is the Spanish language written in Spain. As the COAH corpus, the reviews are labeled in a scale of five levels of opinion intensity, and the size per label is: 461 (5), 890 (4), 1253 (3), 923 (2), 351 (1), in total 3878. As in the COAH corpus, we considered the reviews labeled with 5 and 4 as positive, the ones labeled with 1 and 2 as negative, and we did not consider the neutral intensity label 3. The size of the MuchoCine corpus used in the evaluation is in Table 1. Fig. 2 shows an instance from the training corpus.

*Chilean restaurant corpus (CHIREST).* The CHIREST corpus [45] is composed of reviews about restaurants, and they are written in the Spanish language written in Chile. The source of the reviews in TripAdvisor, and they were published from 2009 to 2014. The corpus is composed of 37,801 reviews from 757 Chilean restaurants, hence the reviews are not skewed to a specific hotel. Like the previous two corpora, the reviews of CHIREST are annotated in a five scale of opinion intensity, and the number of reviews per label are: 15,636 (5), 15,583 (4), 7736 (3), 3729 (2) 3116 (1), in total 39,316. As well as the previous corpora, we did not use the reviews annotated with label 3, and we considered the labels 5 and 4 as positive, and the labels 1 and 2 as negative. The size of the CHIREST corpus used in the evaluation is in Table 1. Fig. 3 shows an instance from the training corpus.

Since we changed the annotation of the three corpora, we show the number of reviews per corpus in Table 1.

**Table 1**
Number of positive and negative reviews per each corpus.

| Corpus | Positive | Negative | Total |
|---|---|---|---|
| COAH | 1025 | 519 | 1544 |
| MuchoCine | 1351 | 1274 | 2625 |
| CHIREST | 31,219 | 6852 | 38,071 |

### 4.2. Pre-processing

The right preparation of the data before its processing is crucial for the classification system to learn and generalize from the training data. We arrange the pre-processing in two steps pipeline: (1) cleaning the data with the aim of normalizing it, removing stop-words and grouping similar ones; and (2) converting the data, in this case, a piece of text, in a feature vector.

Concerning SA, Martínez Cámara et al. [3] showed the relevance of the study of the cleaning of the data, specifically the effect of using *stopper* and *stemmer*, or in other words, the impact of removing stop-words words and grouping the ones with the same stem. Likewise, there is some controversy in how to build the feature representation of an opinion. On the one hand, Pang et al. [46] show that unigrams outperform bigrams, and the other hand Dave et al. [47] defended that bigrams and trigrams may yield better results than unigrams, because they may represent some linguistic patterns that are not possible using only unigrams, like negation or some modifiers that act as opinion shifters. Moreover, we find authors that support the idea of representing opinions with term presence features as [46], and other authors that assert that frequency features are more recommendable as [3].

Since there is not a real consensus in the literature in how to conduct the pre-processing, we evaluate our claim in different pre-processing scenarios, specifically: (1) fourteen cleaning layouts built upon a set of actions, (2) three n-grams representation schemes, and (3) two feature weighting approaches.

*Cleaning layouts.* One of the characteristics of the Spanish language is the use of accents in the words. Since the reviews were crawled from TripAdvisor, they are written in a non-formal writing style by non-professional writers, which means that they are not short of misspellings and grammatical errors. Accordingly, we evaluated the removal of maintaining and removing the accents of all the words. As additional normalization actions, we also evaluated the effect of converting the words to they lowercase form, and to change all the numbers by the wild card `_DIGIT_`. Since the relevance of using *stemmer* and *stopper*, we assessed: (1) the performance of using raw words, (2) the use of the lemma of the words, which allow to group words with the same lemma, (3) the use of a *stemmer*, which group more words than the lemmatizer, and (4) the removal of stop-words by means the use of a *stopper*.

*N-grams.* Since the controversial of using unigrams or other number of n-grams, we evaluated the use of: (1) unigrams, (2) bigrams and (3) unigrams and bigrams.

*Feature weighting.* As the n-grams controversy, there some differences among the use of a term presence or frequency weighting scheme. Consequently, we evaluated the two approaches: (1) the term presence that we call binary, and (2) the term frequency. The presence scheme is developed by a binary strategy, namely, those words that are in the input text are weighted as the value 1, and otherwise 0. Concerning the frequency approach, we used the well-known TF-IDF frequency measure, which gives more relevance to the discriminant words.

The fourteen cleaning layouts scenarios described were performed for each of the three n-grams representations used and the two feature weighting approaches, which make 84 pre-processing scenarios. Hence, we conducted an exhaustive evaluation and we performed all of them for the three corpora and for each of parts of their structure: title, body and jointly for the title and the body.

[6] We clarify that the title is not the output of an extractive summarization system, and it was written by the author of the review.

[7] http://sinai.ujaen.es/coah/.

[8] http://www.lsi.us.es/~fermin/corpusCine.zip.

[9] http://www.muchocine.net/.

Title: *Moderno y buena relación calidad/precio. Muy recomendable*

Body: *Hotel pequeño, bien situado, moderno, limpio y tranquilo. Se encuentra en pleno centro de Jaén, a cinco minutos andando de una de las zonas más famosas de tapas. El trato del personal fue bastante bueno y profesional. La habitación estaba muy limpia y el estilo es bastante moderno, para gente joven. El baño no esta separado totalmente de la habitación, por lo que se oye todo y no da mucha intimidad. No tiene parking y las calles de los alrededores son estrechas y casi siempre están llenas. Por lo demás, ninguna queja. Para repetir.*

Title: `Up-to-date hotel with a good quality price relation. Highly recommended.`

Body: `It is a small, clean, good located, up-to-date, clean and quite hotel. It is located at the downtown of Jaén, five minutes on foot from the most famous tapas zones. The treatment of the staff was very pleasant and professional. The room was clean and the style trendy, mainly for young people. The bath is not completely independent, so the bathroom noise is listened, which does not nearly provide privacy. The hotel doesn't have parking and the surrounding streets are narrow and usually with cars. Otherwise any complaint. To repeat.`

**Fig. 1.** A review instance from the COAH corpus.

### 4.3. Experimental setting for classification

We compared our proposal with three strategies for constructing the data matrix **x**. We consider only the words that appear in the title ($\mathbf{x}_i = \mathbf{x}_i^t$), the words that appear in both title and body with equal weights ($\mathbf{x}_i = \mathbf{x}_i^t + \mathbf{x}_i^b$), and only the words that appear in the body ($\mathbf{x}_i = \mathbf{x}_i^b$). For these alternative strategies, $l_1$-SVM is used. The $\gamma$ values were explored within the range [0, 10]. For the $l_1$-SVM, we used $C = 1$. The Area Under the Curve (AUC) and the F1 score were used as evaluation measures.

Regarding model implementation, we used the LIBLINEAR toolbox [43] for the $l_1$-SVM model. All experiments were performed on an HP Envy dv6 with 16 GB RAM, 750 GB SSD, an Intel Core Processor i7-2620M (2.70 GHz), and using Matlab 2014a and Microsoft Windows 8.1 OS (64-bits).

## 5. Results and analysis

The proposed method was compared with the three alternative techniques for the two feature weighting approaches (binary and TF-IDF), the three n-gram strategies, and the 14 pre-processing scenarios (see Section 4.2). For each of the three corpora and the six representation/n-gram combinations, the best pre-processing strategy is reported. The performances in terms of AUC and F1 score are presented in Tables 2, 3, and 4, for the COAH, Muchocine and CHIREST corpora respectively. The best performance (largest AUC) is highlighted in bold type for the six representation/n-gram combinations. For our proposal, the optimal value for the parameter $\gamma$ is also reported.

Tables 2, 4 and 3 show our proposal outperformed the other methods in all the experiments, showing the largest AUC and F1 score for the six representation/n-gram combinations and the three corpora. The second best performance is achieved with the title/body combination with equal weights, followed by the use of the body of the comments as the sole source attribute construction.

**Table 2**
The AUC and F1 values reached by using the different sections of the corpus COAH without any optimization and with our optimized proposal.

| Corpus | Feat. weight | Unigrams | | Bigrams | | Unigr.+Bigr. | |
|---|---|---|---|---|---|---|---|
| | | AUC | F1 | AUC | F1 | AUC | F1 |
| Title | Binary | 0.84 | 0.89 | 0.71 | 0,65 | 0.85 | 0.89 |
| | TF-IDF | 0.84 | 0.88 | 0.69 | 0.58 | 0.82 | 0.87 |
| Title+Body | Binary | 0.93 | 0.96 | 0.89 | 0.93 | 0.94 | 0.96 |
| | TF-IDF | 0.90 | 0.94 | 0.79 | 0.88 | 0.89 | 0.94 |
| Body | Binary | 0.91 | 0.95 | 0.88 | 0.92 | 0.92 | 0.95 |
| | TF-IDF | 0.89 | 0.93 | 0.77 | 0.87 | 0.88 | 0.93 |
| Our proposal | Binary | **0.95** | **0.97** | **0.91** | **0.94** | **0.96** | **0.97** |
| | TF-IDF | **0.93** | **0.96** | **0.84** | **0.91** | **0.91** | **0.94** |
| $\gamma$ | Binary | 4.5 | | 4.75 | | 3.45 | |
| | TF-IDF | 0.5 | | 4.55 | | 0.58 | |

**Table 3**
The AUC and F1 values reached by using the different sections of the corpus Muchocine without any optimization and with our optimized proposal.

| Corpus | Feat. weight | Unigrams | | Bigrams | | Unigr.+Bigr. | |
|---|---|---|---|---|---|---|---|
| | | AUC | F1 | AUC | F1 | AUC | F1 |
| Title | Binary | 0.74 | 0.75 | 0.65 | 0.66 | 0.74 | 0.74 |
| | TF-IDF | 0.71 | 0.74 | 0.60 | 0.58 | 0.70 | 0.73 |
| Title+Body | Binary | 0.82 | 0.82 | 0.78 | 0.79 | 0.84 | 0.85 |
| | TF-IDF | 0.81 | 0.82 | 0.67 | 0.72 | 0.77 | 0.79 |
| Body | Binary | 0.82 | 0.83 | 0.78 | 0.79 | 0.84 | 0.84 |
| | TF-IDF | 0.79 | 0.80 | 0.65 | 0.71 | 0.76 | 0.76 |
| Our proposal | Binary | **0.85** | **0.85** | **0.80** | **0.80** | **0.86** | **0.87** |
| | TF-IDF | **0.82** | **0.83** | **0.72** | **0.74** | **0.81** | **0.82** |
| $\gamma$ | Binary | 2.0 | | 2.5 | | 2.0 | |
| | TF-IDF | 0.75 | | 1.00 | | 2.0 | |

Regarding the $\gamma$ values, it can be observed that the optimal values ranged between 2 and 5 for the binary representation. This demonstrates the importance of using both sources, but upweighting the information provided by the respondents at the

Title: *Aburrimiento. Esa es la palabra que define el primer tropezón serio del maestro Clint cuando termino de ver Banderas de nuestros padres.*

Body: *Aburrimiento. Esa es la palabra que define el primer tropezón serio del maestro Clint (Deuda de sangre era un poco sosa también, pero es que esta encima es pretenciosa) cuando termino de ver Banderas de nuestros padres. ¿Por dónde empezar? El film tiene una estructura de bucle y las mismas tres escenas se repiten incesantemente durante las dos horas (largas) de metraje. Primero tenemos un flashback del campo de batalla, luego una escena de despacho y luego una escena de campo deportivo. Y así se pasa la película. Luego están los actores, que ni tienen carisma ni emocionan, sobre todo Adam Beach, lamentable, que logra que Ryan Phillippe parezca actor. Los secundarios pasan por ahí, y nos encontramos con Paul Walter o Barry Pepper, (este último parece recién salido de Salvar al soldado Ryan) o con Billy Elliot repitiendo su papel de King Kong. La música es tan repetitiva como la estructura de la narración, es decir, también aburre y llega a cansar. La historia es bastante insulsa y está trilladísima. ¿Lo positivo? Las secuencias bélicas, estupendamente tratadas y rodadas con abrumador clasicismo. Veré las cartas de Iwo Jima, porque no puede ser más aburrida que esta.*

Title: Boredom. This is the word that defines the first important slip of the master Clint after watching Flags of our fathers.

Body: Boredom. This is the word that defines the first important slip of the master Clint (Blood work was a bit boring too, but this is also pretentious) after watching Flags of our fathers. Where to start? The movie has a repetitive structure, because the same three scenes are repeated during the two (long) hours of the film. First, there is a flashback to the battlefield, then a office scene and then a sport field scene. These three scenes are repeated all the film. On the other hand, the actors does not have charisma and they don't thrill, especially Adam Beach, terrible, and his bad performance makes Ryan Phillippe work as an actor. The supporting actors do not have a relevant performance, Paul Walter, Barry Pepper (the former one looks like from Saving Private Ryan) or Billy Elliot repeating his role in King Kong. The music is as repetitive as the plot, in other words, it is boring and exhausting. The stoy is very dull and overused. Any positive? The war scenes, which are fantastically treated and filmed with a overwhelming classicism. I will watch Letters from Iwo Jima, because it cannot be more boring than this film.

**Fig. 2.** A review instance from the MuchoCine corpus.

title of the comment. In contrast, the $\gamma$ values related to the TF-IDF representation are usually below 1. This difference is caused by the nature of the two feature weighting metrics. The binary representation approach is a document level metric, while TF-IDF is a corpus level one. Accordingly, the TF-IDF is providing information from all the opinions in the corpus in each opinion, which means that most frequent words in all the corpus would have a low value of TF-IDF. The title of the reviews are usually short excerpts that usually have similar words, because the goal of the title is to straightly express the polarity of the review. If those words of the title are also frequently used in the body, then the TF-IDF value would be low and the differences among the title and the body will not be relevant and the contribution would be similar, as the $\gamma$ shows. On the other hand, the binary approach only provides document level information, and it does not depend of the frequency of the words. The high frequent use of some words in the title may determine its polarity value, and then those words would be crucial for the classification. Consequently, the $\gamma$ value is higher for the title when the binary approach is used. Regarding the data pre-processing, three pre-processing layouts highlighted: (1) the removal of all the accents, the deletion of the numbers, the transformation to the lowercase form, and the lemmatization of each word; (2) the use of only an *stemmer* and (3) the remove of all the accents, the deletion

Title: *Buena comida y ambientación.*

Body:*Si bien el local no es de lo mejor que he visto, esta caracteristica es suplida por el recibimiento del dueño, la decoracion del local, la atencion de los mozos y su comida. Muy bien preparada y sabrosa, con diferentes intensidades de picor segun se quiera comer y una buena variedad de platos. En general una buena experiencia.*

Title: `Good food and atmosphere.`

Body: `Although I have been in better places, this is overcome by the reception of the owner, the decoration and the service of the waiters and the food. The food is well prepared and tasty with different spicy levels depending on you want and a wide variety of dishes. In general a good experience.`

**Fig. 3.** A review instance from the CHIREST corpus.

**Table 4**
The AUC and F1 values reached by using the different sections of the corpus CHIREST without any optimization and with our optimized proposal.

| Corpus | Feat. weight | Unigrams | | Bigrams | | Unigr.+Bigr. | |
|---|---|---|---|---|---|---|---|
| | | AUC | F1 | AUC | F1 | AUC | F1 |
| Title | Binary | 0.86 | 0.96 | 0.78 | 0.82 | 0.87 | 0.96 |
| | TF-IDF | 0.85 | 0.96 | 0.76 | 0.81 | 0.85 | 0.96 |
| Title+Body | Binary | 0.92 | 0.98 | 0.91 | 0.98 | 0.93 | 0.98 |
| | TF-IDF | 0.92 | 0.98 | 0.86 | 0.97 | 0.91 | 0.98 |
| Body | Binary | 0.90 | 0.98 | 0.89 | 0.97 | 0.91 | 0.98 |
| | TF-IDF | 0.90 | 0.98 | 0.84 | 0.97 | 0.89 | 0.98 |
| Our proposal | Binary | **0.93** | **0.98** | **0.91** | **0.98** | **0.94** | **0.99** |
| | TF-IDF | **0.93** | **0.98** | **0.88** | **0.97** | **0.92** | **0.98** |
| $\gamma$ | Binary | 2.5 | | 3.0 | | 2.5 | |
| | TF-IDF | 0.75 | | 0.75 | | 1.0 | |

of the numbers, the transformation to the lowercase form, and the use of a *stemmer*. Concerning the feature weighting method, we conclude that the binary approach reaches better results than TF-IDF.

## 6. Conclusion

The structure of the opinions posted on domain specific social networks usually have two sections: title and body. Accordingly, we set two research questions: (1) is it possible to weight the relevance of each part of the structure of the review? and (2) is it possible to optimize the weight of each part of the review? In order to answer them, we claim that the right combination of the polarity meaning of each section of an opinion may enhance the performance of a polarity classification system. We thus propose the weighting optimization of the contribution of each section of the opinion by a modification of the training function of the SVM algorithm classification algorithm, which allows to weight the contribution of each section. Likewise, we optimize the training function of the SVM with a line-search method.

We evaluated our proposal on three datasets from a different domain and written in two different versions of the same language, Spanish. Since the pre-processing of the data determines the performance of the classification algorithm, we also evaluated several pre-processing approaches. As the results show, our proposal outperforms the individual and joint classification of the reviews, which allow us to conclude that our claim holds and we positively answered to the research questions that we set.

As future work, we will study the lexical diversity among the parts of the reviews, and the differences between the Spanish spoken in Spain and in Chile, or in other words the set up a cross-lingual evaluation of our proposal with reviews of both versions of Spanish. We will also work on the gathering of reviews written in other Spanish versions from Spanish spoken countries.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This research work is partially supported by the Spanish Government project TIN2017-89517-P, and a grant from the Fondo Europeo de Desarrollo Regional (FEDER). Eugenio Martínez Cámara was supported by the Spanish Government Programme Juan de la Cierva Formación (FJCI-2016-28353). Sebastián Maldonado gratefully acknowledges financial support from CONICYT PIA/BASAL AFB180003 and FONDECYT-Chile, grant 1160738.

## References

[1] Bo Pang, Lillian Lee, Opinion mining and sentiment analysis, Found. Trends Inf. Retr. (ISSN: 1554-0669) 2 (1–2) (2008) 1–135, http://dx.doi.org/10.1561/1500000011.
[2] Erik Cambria, Amir Hussain, Sentic Computing. A Common-Sense-Based Framework for Concept-Level Sentiment Analysis, second ed., Springer, 2015, http://dx.doi.org/10.1007/978-3-319-23654-4.
[3] Eugenio Martínez Cámara, M. Teresa Martín Valdivia, José M. Perea Ortega, L. Alfonso Ureña López, Opinion classification techniques applied to a spanish corpus, Proces. Leng. Nat. 47 (2011) 163–170.
[4] Bing Liu, Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Cambridge University Press, 2015, http://dx.doi.org/10.1017/CBO9781139084789.
[5] Fermín L. Cruz, Jose A. Troyano, Fernando Enríquez, Javier Ortega, Experiments in sentiment classification of movie reviews in spanish, Proces. Leng. Nat. 41 (2008) 73–80.
[6] Julian Brooke, Milan Tofiloski, Maite Taboada, Cross-linguistic sentiment analysis: From english to spanish, in: Proceedings of the International Conference RANLP-2009, Association for Computational Linguistics, Borovets, Bulgaria, 2009, pp. 50–54.
[7] Veronica Perez-Rosas, Carmen Banea, Rada Mihalcea, Learning sentiment lexicons in spanish, in: Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012), European Language Resources Association (ELRA), Istanbul, Turkey, 2012, pp. 3077–3081.
[8] Stefano Baccianella, Andrea Esuli, Fabrizio Sebastiani, Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining, in: Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10), European Languages Resources Association (ELRA), Valletta, Malta, 2010, pp. 2200–2204.
[9] George A. Miller, Wordnet: A lexical database for english, Commun. ACM (ISSN: 0001-0782) 38 (11) (1995) 39–41, http://dx.doi.org/10.1145/219717.219748.

[10] Salud María Jiménez-Zafra, Mariona Taulé, M. Teresa Martín-Valdivia, L. Alfonso Ureña-López, M. Antónia Martí, Sfu reviewsp-neg: a spanish corpus annotated with negation for sentiment analysis. a typology of negation patterns, Lang. Res. Eval. (ISSN: 1574-0218) 52 (2) (2018) 533–569, http://dx.doi.org/10.1007/s10579-017-9391-x.

[11] Flor Miriam Plaza del Arco, M. Teresa Martín Valdivia, Salud María Jiménez Zafra, M. Dolores Molina González, Eugenio Martínez Cámara, Copos: Corpus of patient opinions in spanish. application of sentiment analysis techniques, Proces. Leng. Nat. 57 (2016) 83–90.

[12] Eugenio Martínez Cámara, M. Teresa Martín-Valdivia, M. Dolores Molina-González, L. Alfonso Ureña-López, Bilingual experiments on an opinion comparable corpus, in: Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Association for Computational Linguistics, Atlanta, Georgia, 2013, pp. 87–93.

[13] Fermín L. Cruz, José A. Troyano, Beatriz Pontes, F. Javier Ortega, Ml-senticon: Un lexicón multilingüe de polaridades semánticas a nivel de lemas, Proces. Leng. Nat. 53 (2014) 113–120.

[14] M. Dolores Molina-González, Eugenio Martínez-Cámara, M. Teresa Martín-Valdivia, L. Alfonso Ureña-López, A spanish semantic orientation approach to domain adaptation for polarity classification, Inf. Process. Manage. 51 (4) (2015) 520–531, http://dx.doi.org/10.1016/j.ipm.2014.10.002.

[15] Maite Taboada, Julian Brooke, Manfred Stede, Genre-based paragraph classification for sentiment analysis, in: Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue, in: SIGDIAL '09, Association for Computational Linguistics, Stroudsburg, PA, USA, ISBN: 978-1-932432-64-0, 2009, pp. 62–70.

[16] Eugenio Martínez-Cámara, M. Teresa Martín-Valdivia, L. Alfonso Ure na López, Ruslan Mitkov, Polarity classification for spanish tweets using the cost corpus, J. Inf. Sci. 41 (3) (2015) 263–272, http://dx.doi.org/10.1177/0165551514566564.

[17] Julio Villena-Román, Sara Lana-Serrano, Eugenio Martínez-Cámara, José Carlos González-Cristóbal, Tass - workshop on sentiment analysis at sepln, Proces. Leng. Nat. (ISSN: 1989-7553) 50 (2013) 37–44.

[18] Manuel C. Díaz-Galiano, Eugenio Martínez-Cámara, M. Ángel García Cumbreras, Manuel García Vega, Julio Villena Román, The democratization of deep learning in tass 2017, Proces. Leng. Nat. 60 (2018) 37–44.

[19] Edgar Casasola, Alejandro Pimentel, Gerardo Sierra, Eugenio Martínez Cámara, Gabriela Marín, Comparative analysis of the computational characteristics in modern sentiment analysis systems for spanish, Proces. Leng. Nat. 62 (2019) 69–76.

[20] Eugenio Martínez Cámara, Sentiment Analysis in Spanish (PhD thesis), University of Jaén, 2015.

[21] Quoc-Chinh Bui, Peter M.A. Sloot, Erik M. Van Mulligen, Jan A. Kors, A novel feature-based approach to extract drug–drug interactions from biomedical text, Bioinformatics 30 (23) (2014) 3365–3371.

[22] E. Pons, B. Becker, S.A. Akhondi, Z. Afzal, E.M. van Mulligen, J.A. Kors, Religator: chemical-disease relation extraction using prior knowledge and textual information, in: Proceedings of the Fifth BioCreative Challenge Evaluation Workshop, pp. 247–253, 2015.

[23] Jun Xu, Yonghui Wu, Yaoyun Zhang, Jingqi Wang, Ruiling Liu, Qiang Wei, Hua Xu, Uth-ccb@ biocreative v cdr task: identifying chemicalinduced disease relations in biomedical text. In Proceedings of the Fifth BioCreative Challenge Evaluation Workshop, pp. 254–259, 2015.

[24] Alessandro Moschitti, A study on convolution kernels for shallow semantic parsing, in: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics, 2004, p. 335.

[25] Min Zhang, Jie Zhang, Jian Su, Guodong Zhou, A composite kernel to extract relations between entities with both flat and structured features, in: Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, 2006, pp. 825–832.

[26] Sun Kim, Haibin Liu, Lana Yeganova, W. John Wilbur, Extracting drug–drug interactions from literature using a rich feature-based linear kernel approach, J. Biomed. Inform. 55 (2015) 23–30.

[27] Dan Jurafsky, James H. Martin, Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition, 2009.

[28] Dinesh Kumar, Gurpreet Singh Josan, Part of speech taggers for morphologically rich indian languages: a survey, Int. J. Comput. Appl. 6 (5) (2010) 32–41.

[29] P.J. Antony, K.P. Soman, Kernel based part of speech tagger for kannada, in: Machine Learning and Cybernetics (ICMLC), 2010 International Conference on, Vol. 4, IEEE, 2010, pp. 2139–2144.

[30] Asif Ekbal, Sivaji Bandyopadhyay, Part of speech tagging in bengali using support vector machine, in: Information Technology, 2008. ICIT'08. International Conference on, IEEE, 2008, pp. 106–111.

[31] M. Sivakumar, C. Karthika, P. Renuga, A hybrid text classification approach using knn and svm, Int. J. Innov. Res. Sci. Eng. Technol. 3 (3) (2014) 1987–1991.

[32] Sundus Hassan, Muhammad Rafi, Muhammad Shahid Shaikh, Comparing svm and naive bayes classifiers for text categorization with wikitology as knowledge enrichment, in: Multitopic Conference (INMIC), 2011 IEEE 14th International, IEEE, 2011, pp. 31–34.

[33] Michail Tsikerdekis, Sherali Zeadally, Multiple account identity deception detection in social media using nonverbal behavior, IEEE Trans. Inf. Forens. Secur. 9 (8) (2014) 1311–1321.

[34] Saurabh Kr. Srivasatava, Roshan Kumari, Sandeep Kr. Singh, An ensemble based nlp feature assessment in binary classification, in: Computing, Communication and Automation (ICCCA), 2017 International Conference on, IEEE, 2017, pp. 345–349.

[35] Ajay Deshwal, Sudhir Kumar Sharma, Twitter sentiment analysis using various classification algorithms, in: Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO), 2016 5th International Conference on, IEEE, 2016, pp. 251–257.

[36] Dwi Aji Kurniawan, Sunu Wibirama, Noor Akhmad Setiawan, Real-time traffic classification with twitter data mining, in: Information Technology and Electrical Engineering (ICITEE), 2016 8th International Conference on, IEEE, 2016, pp. 1–5.

[37] Jeonghee Yi, Wayne Niblack, Sentiment mining in webfountain, in: Data Engineering, 2005. ICDE 2005. Proceedings. 21st International Conference on, IEEE, 2005, pp. 1073–1083.

[38] Songbo Tan, Jin Zhang, An empirical study of sentiment analysis for chinese documents, Expert Syst. Appl. 34 (4) (2008) 2622–2629.

[39] Cheng Li, Bingyu Wang, Virgil Pavlu, Javed A. Aslam, An empirical study of skip-gram features and regularization for learning on sentiment analysis, in: European Conference on Information Retrieval, Springer, 2016, pp. 72–87.

[40] Vladimir Vapnik, The Nature of Statistical Learning Theory, Springer science & business media, 2013.

[41] Sebastián Maldonado, Richard Weber, A wrapper method for feature selection using support vector machines, Inform. Sci. 179 (13) (2009) 2208–2217.

[42] Paul S. Bradley, Olvi L. Mangasarian, Feature selection via concave minimization and support vector machines, in: ICML, Vol. 98, 1998, pp. 82–90.

[43] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, Chih-Jen Lin, Liblinear: A library for large linear classification, J. Mach. Learn. Res. 9 (Aug) (2008) 1871–1874.

[44] M. Dolores Molina-González, Eugenio Martínez-Cámara, M. Teresa Martín-Valdivia, L. Alfonso Ureña-López, Cross-domain sentiment analysis using spanish opinionated words, in: Elisabeth Métais, Mathieu Roche, Maguelonne Teisseire (Eds.), Natural Language Processing and Information Systems, Springer International Publishing, Cham, 2014, pp. 214–219.

[45] Sebastián Maldonado, Julio López, Carla Vairetti, An alternative smote oversampling strategy for high-dimensional datasets, Appl. Soft Comput. 76 (2019) 380–389, http://dx.doi.org/10.1016/j.asoc.2018.12.024, URL http://www.sciencedirect.com/science/article/pii/S1568494618307130.

[46] Bo Pang, Lillian Lee, Shivakumar Vaithyanathan, Thumbs up?: Sentiment classification using machine learning techniques, in: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Vol. 10, in: EMNLP '02, Association for Computational Linguistics, Stroudsburg, PA, USA, 2002, pp. 79–86, http://dx.doi.org/10.3115/1118693.1118704.

[47] Kushal Dave, Steve Lawrence, David M. Pennock, Mining the peanut gallery: Opinion extraction and semantic classification of product reviews, in: Proceedings of the 12th International Conference on World Wide Web, in: WWW '03, ACM, New York, NY, USA, ISBN: 1-58113-680-3, 2003, pp. 519–528, http://dx.doi.org/10.1145/775152.775226.

**Carla Vairetti** is currently Assistant Professor and Researcher at the Universidad de Los Andes in Chile. She holds a bachelor's degree in Computer Science Engineering from the UNLP in Argentina, a M.S. degree from the Pontífice University of Chile, in 2013, and a Ph.D. degree from the University of Trento (Italy), in 2017. Her research interests are Business Process Management (BPM): Modeling and Process Simulation, Data Mining, and Business Analytics.

**Eugenio Martínez-Cámara** received his Ph.D. in Computer Science at the University of Jaén (Spain). He is working as postdoctoral researcher at the University of Granada (Spain), where he is researching on the use of Deep Learning in several Natural Language Processing tasks. He collaborates with several journals and conferences as reviewer, and he is member of the scientific committee of the Spanish Society for Natural Language Processing (SEPLN).

**Sebastián Maldonado** received his B.S. and M.S. degree from the University of Chile, in 2007, and his Ph.D. degree from the University of Chile, in 2011. He is currently Associate Professor at the School of Engineering and Applied Sciences, Universidad de los Andes, Santiago, Chile. His research interests include statistical learning, data mining and business analytics. Sebastián Maldonado has published more than 60 scientific contributions including more than 40 Thomson Reuters' ISI papers in the last five years.

**M. Victoria Luzón** is an associate professor in the Software Engineering Department at University of Granada. Her research interests include sentiment analysis, artificial intelligence, computer graphics and cultural heritage. Luzón has a Ph.D. in Industrial Engineering from the University of Vigo, Spain.

**Francisco Herrera** received his M.Sc. in Mathematics in 1988 and Ph.D. in Mathematics in 1991 (University of Granada, Spain). He is currently a professor in the Department of Computer Science and Artificial Intelligence at the University of Granada. Prof. Herrera has supervised 44 Ph.D. thesis and published more than 400 journal papers (h-index = 133, Scholar Google). He currently acts as Editor in Chief of the international journals "Information Fusion"(Elsevier) and "Progress in Artificial Intelligence"(Springer). He has been selected as a Highly Cited Researcher http://highlycited.com/ (fields of Computer Science and Engineering, respectively, 2014 to present, Clarivate Analytics). His current research interests include soft computing (fuzzy modeling, evolutionary algorithms and deep learning), computing with words, information fusion, and data science (data preprocessing, prediction and big data).