Spanish Sentiment Analysis in Twitter at the TASS Workshop.

Ferran Pla · Lluís-F. Hurtado

Abstract This paper describes a Support Vector Machine-based system for different tasks related to the Sentiment Analysis problem in Twitter. We have paid special attention to the parameter optimization of model and to the combination of different system configurations using voting techniques. It is important to say that the tokenization and the feature selection process have been specifically designed for the Spanish Language. We have evaluated this approach in different tasks that were proposed at the five editions of the TASS workshop, from 2012 to 2016. TASS has become a framework for Sentiment Analysis tasks that is focused on the Spanish language. We describe our participation in this competition and the results achieved. We also perform an analysis and a comparison of the other participants in TASS competition. To our knowledge, the results achieved by our system are the best results published until now for the Sentiment Analysis tasks of the TASS workshop.

1 Introduction

The large number of on-line reviews in social media has obliged the scientific community to make great efforts to analyze, structure, and process this in-

Ferran Pla
Universitat Politècnica de València. Camí de Vera, s/n. 46022 València (Spain)
Tel.: +34 963877357
Fax: +34 963877359
E-mail: fpla@dsic.upv.es

Lluís-F. Hurtado
Universitat Politècnica de València. Camí de Vera, s/n. 46022 València (Spain)
Tel.: +34 963877000 (ext. 83500)
Fax: +34 963877359
E-mail: lhurtado@dsic.upv.es

formation. Among other things, these media are used to express opinions and feelings about companies, products, services, hobbies, politics, etc. Therefore, enterprises, organizations, governments, and different groups in general have shown interest in the opinions and feelings that users have about their activities. In this respect, Twitter has become a popular micro-blogging site in which users express their opinions on a variety of topics in real time.

The texts used in Twitter are called tweets, which are short texts of a maximum of 140 characters and language that does not have any restrictions on the form and content. The nature of these texts poses new challenges for researchers in Natural Language Processing (NLP). In some cases, the tweets are written with ungrammatical sentences with a lot of emoticons, abbreviations, specific terminology, slang, etc. Therefore, the usual NLP techniques must be adapted to these characteristics of the language. New approaches must be proposed in order to successfully address this problem. NLP tools like POS taggers, parsers, or Named Entity Recognition (NER) tools usually fail when processing tweets because they are generally trained on grammatical texts and they perform poorly in micro-blogging texts.

Most of the work of Sentiment Analysis (SA) in Twitter is for the English language, and this is also true for the resources and tools available for NLP. Therefore, the TASS[1] workshop aims to be a framework for SA and on-line reputation analysis that is focused on the Spanish language. It provides a forum for discussion and communication in which the latest research work and developments in the field of social media analytics are specifically focused on the Spanish language. To evaluate different approaches in SA and text categorization, TASS provides corpora of Spanish tweets: the *General* corpus, which is annotated with polarity and topics; and the *SocialTV*, *Politics*, and *STOMPOL* corpora, which are also annotated with polarity at the aspect level. These corpora are freely available on the web page of TASS. This is a very important resource that allows researchers to compare their approaches for different SA tasks in Twitter by using the same data.

In this paper, we present a system that is based on the Support Vector Machine formalism for addressing the SA analysis tasks proposed in the different editions of the TASS workshop, at both the global level and at the aspect level in Twitter. In addition, for the topic classification task, the system allows multi-label classification to assign more than one topic to a tweet. The output of the aspect-based Sentiment Analysis system has also been successfully applied to the *political tendency identification* task. We want to highlight the flexibility of our system which has allowed us to participate in all of the Sentiment Analysis tasks proposed at TASS using the same system and achieving the best results of the competition. In addition, our proposed system has demonstrated its usefulness for tasks that are not strictly classification tasks. For this task, as relevant information, we used he polarity of the tweets at the aspect level, which was obtained using the system.

---

[1] Web site with links to the five editions of TASS (Taller de Análisis de Sentimientos en la SEPLN) workshop: http://www.sepln.org/workshops/tass/2016/tass2016.php

The rest of this paper is organized as follows. Section 2 presents some works related to the SA problem.These are mainly focused on the different approaches to the Spanish SA tasks proposed at the TASS competition. In Section 3, we summarize the tasks proposed in the different editions of the TASS competition, the corpora supplied by the organization, and the main resources used by the participating teams. In Section 4, we present the system with which we participated in the different editions of the TASS as well as the peculiarities that we introduced for all of the tasks addressed. In Section 6, we review the best results obtained by the participating teams in the tasks of the TASS workshops focusing on the results achieved by our systems. Finally, in Section 7, we present some conclusions and possible directions for future work.

2 Related Work

Sentiment Analysis has been widely studied in the last decade in multiple domains. Most work focuses on classifying the polarity of the texts as positive, negative, or neutral. The pioneering works in this field used supervised (Pang et al 2002) or unsupervised (knowledge-based) (Turney 2002) approaches. In (Pang et al 2002), the performance of different classifiers on movie reviews was evaluated. In (Turney 2002), some patterns containing POS information were used to identify subjective sentences in reviews to then estimate their semantic orientation. A good presentation of the SA problem and a description of the state-of-the-art of the more relevant approaches to SA can be found in (Liu 2012).

Most works that address the SA problem use polarity lexicons in some way. The construction of these lexicons is another widely explored field of research. Polarity lexicons have usually been constructed for English (Liu et al 2005; Wilson et al 2005), but efforts have also been made to create lexicons for Spanish (Perez-Rosas et al 2012; Saralegi and San Vicente 2013; Martínez-Cámara et al 2013).

Research works about SA in Twitter are much more recent. Twitter appeared in 2006, and the early works in this field are from 2009 when Twitter started to achieve popularity. Some of the most significant works are (Barbosa and Feng 2010), (Jansen et al 2009), and (O'Connor et al 2010). A survey of the most relevant approaches to SA in Twitter can be found in (Martínez-Cámara et al 2014) and (Vinodhini and Chandrasekaran 2012).

The SemEval competition has also dedicated some specific tasks for SA in Twitter.

The approaches and results obtained by the participating teams for the Sentiment Analysis in Twitter can be consulted in the overview papers of each SemEval edition. These overviews and the resources are available in the web site of SemEval[2] .

---

[2] Web sites of the SA tasks at SemEval (International Workshop on Semantic Evaluation): https://www.cs.york.ac.uk/semeval-2013

Next, we shortly review the winner systems of the Sentiment Analysis in Twitter at the message-level in all the SemEval editions.

The first edition in which this task was proposed was SemEval-2013 (Nakov et al 2013). Most of the participating teams used supervised systems such as SVM, Naïve Bayes, Maximum Entropy and, ensemble classifiers. The winner team (Mohammad et al 2013) used an SVM classifier that considered a wide range of features and also polarity lexicons.

In SemEval-2014 (Rosenthal et al 2014), as in the previous edition, the most popular classifiers were SVM, MaxEnt, and Naive Bayes. The highest score was again achieved by the NRC-Canada team (Zhu et al 2014). The major improvement in their system over the 2013 system was the way it handles negation.

In SemEval-2015 (Rosenthal et al 2015) edition, in addition to the previously used machine learning paradigms, approaches based on the use of neural networks were proposed. The winner system Hagen et al (2015) combined in an ensemble four Twitter sentiment classification approaches that participated in previous SemEval editions. The second position was achieved by Unitn team Severyn and Moschitti (2015). They used a deep learning approach based on the use of a convolutional neural network. A Logistic Regression model with a weighting schema of positive and negative labels was used for the third system in the task ranking by Hamdan et al (2015).

In SemEval-2016 (Nakov et al 2016) edition there is a clear dominance of methods based on deep learning. The top two ranked teams (Deriu et al 2016) and (Rouvier and Favre 2016) used convolutional neural network approaches. The third system (Xu et al 2016) also used a convolutional neural network but combined with a Naive Bayes classifier and a support vector machine classifier.

The TASS workshop has proposed different tasks for SA that focus on the Spanish language. An overview of the different tasks proposed, the participating teams, and the results obtained can be found in (Villena-Román et al 2013a,b, 2014, 2015; García Cumbreras et al 2016; García Cumbreras et al 2016). In Section 6, we will show the best results obtained by the participating teams in the last two editions of the TASS workshops. Then, we present the main features of the different SA approaches that participated in the different editions of the TASS competition.

The first edition of TASS competition was celebrated in 2012. Nine groups sent their submissions. There are no proceedings that describe the features of each system, there are, only some technical notes that can be found on the TASS2012[3] web site. The system presented by Saralegi and San Vicente (2012) achieved the best results. It used a sequential minimal optimization (SMO) implementation of the Support Vector Machine algorithm using Weka (Frank et al 2016) and a polarity lexicon that was constructed from automatically

---

http://alt.qcri.org/semeval2014
http://alt.qcri.org/semeval2015
http://alt.qcri.org/semeval2016

[3] http://www.sepln.org/workshops/tass/2012/tass2012.php

translated English Lexicons and words extracted from a training corpus. The system presented by Batista and Ribeiro (2012) used maximum entropy models that combined a large set of features and a sentiment lexicon in Spanish (Perez-Rosas et al 2012). The system presented in (Trilla and Alías 2012) is based on the Multinomial Naive Bayes approach using binary-weighted features. The system presented in (Martín-Wanton and de Albornoz 2012) used WordNet to extract the concepts in a sentence that have an emotional meaning. Then they assigned them an emotion within a set of categories from an affective lexicon. The system uses this information as the input to a logistic regression machine-learning algorithm. In (Castellano et al 2012), information retrieval techniques were used to determine the topic and the polarity of the tweets. The work of Moreno-Ortiz and Pérez-Hernández (2012) combined knowledge sources (e.g., a lexicon of words and multi-words, context rules) and linguistic tools (e.g., *Freeling* (Padró and Stanilovsky 2012)) for lemmatizing and phase identification in order to obtain a global sentiment value for an input tweet.

The second of TASS[4] was celebrated in 2013. A total of fourteen teams participated in this edition (Villena-Román and García-Morera 2013). The work of Fernández et al (2013) combined a ranking algorithm and skipgrams to create new lexicons that take into account the context of the terms. The system presented in (Saralegi and San Vicente 2013) used an SVM classifier that includes lexicons and some linguistic information. Pla and Hurtado (2013a) presented an SVM approach to the SA tasks using specific tools for tokenizing the tweets. The system presented in (Gamallo et al 2013) is based on a Naive Bayes classifier. Vilares et al (2013) used a preprocessing technique to normalize the tweets. Then, they applied some linguistic tools, such as part-of-speech tagging and dependency parsing, to obtain the syntactic structure of the tweets. Balahur and Perea-Ortega (2013) adapted a machine-learning English SA system to the Spanish language. del-Hoyo-Alonso et al (2013) used a Non-Axiomatic Reasoning System (NARS) to discover content words and phases that contain opinions. García and Thelwall (2013) adapted SentiStrength to the Spanish language, by translating and converting a polarity lexicon. They tuned the system for only the political domain. The system presented in (Villar Rodríguez et al 2013) performed a linguistic treatment of phenomena such as negation detection by using Freeling. The work presented in (Castellanos-González et al 2013) used information retrieval techniques. They proposed different types of modeling the tweets taking into account the different types of information. Martínez-Cámara et al (2013) used an unsupervised approach that used linguistic resources such as SentiWordNet, Q-WordNet, and iSOL. Montejo-Ráez et al (2013) proposed a solution based on Information Retrieval by applying Latent Semantic Analysis.

In the TASS2014[5] edition (Villena-Román et al 2014), a total of 14 teams participated, six of which had already participated in previous editions. For this reason, many of the approaches were improvements of the systems pre-

---

[4] http://www.sepln.org/workshops/tass/2013/tass2013.php
[5] http://www.sepln.org/workshops/tass/2014/tass2014.php

sented in previous editions. Hurtado and Pla (2014) presented an improvement of their system based on SVM. They improved the tokenization phase by using linguistic resources (the Freeling toolkit). They also changed the matching learning architecture used in the previous edition by adding a simple majority voting schema. San Vicente Roncal and Saralegi Urizar (2014) improved their SVM classifier by including new polarity lexicons with linguistic features. Vilares et al (2014) improved their machine-learning approach by using several linguistic resources and some information extracted from corpora. Montejo-Ráez et al (2014) used an SVM system in which they included word vectors generated from Spanish Wikipedia. Jiménez Zafra et al (2014) developed an unsupervised classifier based on the use of an opinion lexicon and the use of syntactic heuristics for identifying the scope of Spanish negation words. Perea-Ortega and Balahur (2014) used a machine-learning approach that focused on different feature replacements. The approach of Hernández Petlachi and Li (2014) was based on linguistic rules. The classification was performed according to a dictionary of semantic orientation, where each term is labeled with a use value and an emotional value.

In the TASS2015[6] edition, a total of 17 teams partipated in the competition. Most of the systems used machine-learning approaches. Most classifiers were based on the use of the Support Vector Machine (SVM) formalism. Other systems used Logistic Regressionc (LR), Maximum Entropy (ME) models, Decision Trees (DT), Naive Bayes (NB), and convolutional or recurrent Neural Networks (NN). In addition, many participants implemented a mechanism for combining models in order to take advantage of the complementarity of the models. Our UPV-ELIRF team (Hurtado et al 2015) was the winner of this edition. The system consisted of a voting scheme that combined different feature configurations of SVM-based models using another SVM as a meta-classifier. The LIF team presented a system that combined SVM and Convolutional Neural Networks (CNN), but they did not submit a report. In (Valverde-Tohalino and Tejada-Cárcamo 2015), a voting system considering four classifiers (SMV, ME, DT, and NB) was presented. Mendizabal and Carandell (2015) presented a voting systems considering four classifiers (SVM, AdaBoost, LR, and Random Forest). Other systems that use SVM classifiers can be found in (Araque et al 2015; Sixto-Cesteros et al 2015; Siordia et al 2015; Fernández et al 2015). Some works used Deep Learnig aproaches (Vilares et al 2015; Díaz-Galiano and Montejo-Ráez 2015), Logistic regression classifiers (Álvarez-López et al 2015; Santos-Deas et al 2015), and Maximun Entropy models (del-Hoyo-Alonso et al 2015). There were approaches based on using polarity lexicons and sociolinguistic features exclusively (Martínez-Cámara et al 2015; Park 2015).

In the last edition of TASS2016[7], only seven teams submitted a system. As in the previous edition, the best systems were those based on machine learning that combined different models using a voting strategy (Hurtado and Pla 2016) and (Cerón-Guzmán 2016). Montejo-Ráez and Díaz-Galiano (2016)

---

[6] http://www.sepln.org/workshops/tass/2015/tass2015.php
[7] http://www.sepln.org/workshops/tass/2016/tass2016.php

and Quirós et al (2016) used SVM and Logistic Regression as classifiers, using word embeddings for the representation of the words in the tweets. Finally, Casasola Murillo and Marín Raventós (2016) submitted an unsupervised system based on the system described in (Turney 2002), but with a specific adaptation to the classification of tweets written in Spanish.

In Section 6, we analyze the main features of these systems at TASS2015 and TASS2016 as well as the results obtained by the participating systems.

Unlike for the English language, the best results for the Spanish SA in Twitter task have been obtained by systems based on SVM classifiers. Segura-Bedmar et al (2017) have recently presented the first work that uses convolutional neural networks with the TASS corpus. Although the results are promising, they have not surpassed the best results of the competition which had been achieved using SVM.

3 Tasks and Corpora at the TASS Workshop

The TASS workshop has proposed different tasks for SA that focus on the Spanish language. An overview of the different tasks proposed, the participating teams, and the results obtained can be found in (Villena-Román et al 2013a,b, 2014, 2015; García Cumbreras et al 2016). This section describes the set of tasks proposed during the five editions of the TASS workshop (2012-2016). We also describe the corpora used in the TASS workshops.

3.1 Tasks

The tasks are mainly related to the SA at the global level and the aspect-based SA level problems for the Spanish Language. It also includes a task for topic classification of tweets and a task for political tendency identification of Twitter users. Table 1 presents these tasks and the editions in which they were proposed.

Table 1 Tasks addressed in each edition of the TASS workshop

Task	2012	2013	2014	2015	2016
SA at the global level	X	X	X	X	X
Topic Detection	X	X	X	-	-
Aspect Detection	-	-	X	-	-
Aspect-based SA (Politics corpus)	-	X	-	-	-
Aspect-based SA (Social-TV corpus)	-	-	X	X	-
Aspect-based SA (STOMPOL corpus)	-	-	-	X	X
Political Tendency Identification	-	X	-	-	-

3.1.1 Sentiment Analysis at the global level

This task consists of determining the global polarity of each tweet in the
provided test set. There are two different tasks: one based on six different
polarity labels (P+, P, NEU, N, N+, NONE) and another based on just four
labels (P, N, NEU, NONE).

The first subtask uses N and $N+$ labels for negative polarity with different
intensity, P and $P+$ labels for positive polarity with different intensity, and the
NEU label for neutral polarity. The second subtask has only three polarity
levels: negative polarity (N label), positive polarity (P label), and neutral
polarity (*NEU* label). In both subtasks, an additional label (*NONE*) was used
to represent tweets with no polarity at all. Accuracy (correct tweet polarity
according to the gold standard) was used for rank the systems. This task was
proposed in the five editions of the TASS from 2012 to 2016.

3.1.2 Topic Detection

The aim of this task is to identify the topic or topics of each tweet in the
test set provided by the organizers. The set of topics is composed of 10 topics:
politics, entertainment, economy, music, soccer, films, technology, sports, liter-
ature, and other. Due to the fact that a tweet can be classified into more than
one class (multi-label classification), microaveraged Precision, Recall, and the
F-measure calculated over the full test set has been used to evaluate the sys-
tems (Sebastiani 2002). Systems are ranked by the F_1 measure. This task was
proposed in three editions of TASS, from 2012 to 2014.

3.1.3 Aspect Detection

The aim of this task is to identify the aspects or entities in a given corpus.
This task was proposed at the TASS 2013 workshop for a specific domain.
The participants had to identify different aspects related to a football match
(such as referees, players, teams, etc.) in tweets that were collected during the
celebration of the match.

3.1.4 Aspect-based Sentiment Analysis

The aim of this task is to identify the polarity of each aspect or entity that
appears in the tweets. It was proposed in the 2013, 2014, 2015, and 2016
editions by using different data sets. With the exception the 2014 edition
where participants had to identify the aspects, in the rest of editions, the
organizers provided the participants with a corpus of tweets with the aspects
already detected that were to be labeled with their polarity.

3.1.5 Political Tendency Identification

The objective of this task is to estimate the political tendency of each user from the supplied corpus using four possible values: *Left*, *Right*, *Centre*, and *Undefined*. The participants could use whatever resource or strategy they decided. In this task, the participants must analyze all the tweets published by a user to determine his/her political opinion. In other words, taking into account all of the tweets of a user, you must determine if the user is leftist, conservative, centrist, or has no political opinion. Only one label per user should be provided. This task was only proposed in the TASS 2013 edition.

3.2 Corpora at TASS

To train and test the approaches of the participants, the organizers supplied a set of corpora to be used in the different tasks proposed at the TASS competition. Table 2 summarizes the main features of these corpora and the tasks in which each corpus was used.

Table 2 Corpora at the TASS competition

Corpus	training	test	polarity	aspects	topics	tasks
General	7219	60798	yes	only training	yes	3.1.1 3.1.2 3.1.5
Politics	no	2079	yes	yes	no	3.1.3 3.1.4 3.1.5
SocialTV	1773	1000	no	yes	no	3.1.4
STOMPOL	784	500	no	yes	no	3.1.4

The *General* corpus contains approximately 68000 Twitter messages (*tweets*) written in Spanish (between November 2011 and March 2012) by about 150 well-known personalities of the world of politics, economy, communication, mass media, and culture.

Each tweet includes its ID (*tweetid*), the creation date (*date*), and the user ID (*user*). Due to restrictions in the Twitter terms of service, it is forbidden to redistribute a corpus that includes text contents or information about users. Nevertheless, the actual message content can be easily obtained by making queries to the Twitter API using the *tweetid*. It is encoded in XML, and it is divided into two sets: training (about 10%, 7219 tweets) and test (about 90%, 60798 tweets).

Each message in both the training and the test sets is labeled with its global polarity, indicating whether the text expresses a positive, negative, or neutral sentiment, or no sentiment at all. In addition, the training set is also labeled at the aspect level for some political entities. A set of topics was defined based on the thematic areas covered by the corpus, such as politics, soccer, literature, entertainment, etc. Each tweet is labeled with one or several of these topics.

The *Politics* corpus contains 2079 tweets gathered during the electoral campaign of the 2011 general elections in Spain from Twitter messages mentioning

any of the four principal national-level political parties. The polarity at the entity level for those entities were manually labeled using the labels: positive (P), neutral (NEU), negative (N), and no sentiment (NONE).

The *SocialTV* corpus was collected during the 2014 Final of the Copa del Rey Championship in Spain between Real Madrid and F.C. Barcelona, which was played on 16 April 2014 at Mestalla Stadium in Valencia. All of the tweets were manually labeled with the aspects and their polarity.

The *STOMPOL* corpus was gathered in 2015 April. It is related to different political aspects that appear in political campaigns, such as, economics, the health system, education, political parties and, other aspects. Each aspect is related to one or several political parties in Spain. Each tweet was tagged with the polarity at the aspect level using three tags: P (Positive), NEU (Neutral), and N (Negative).

4 An SVM approach to SA tasks

During our participation in the 2013, 2014, 2015, and 2016 editions of the TASS workshop, we used systems that are based on the SVM formalism (Pla and Hurtado 2013b; Hurtado and Pla 2014; Hurtado et al 2015; Hurtado and Pla 2016). Throughout these editions, we incorporated different modules and different improvements to our system. In this section, we present the last version of our system that we used in the 2015 and 2016 editions of TASS.

Most of the tasks described in Section 3 can be formalized as a classification problem in which we learn a function $f : X \rightarrow Y$ from a data set of labeled samples D, where

- L is a finite set of disjoint labels $L = \{l_1, \cdots, l_{|L|}\}$, and $|L| > 1$
- X is an input space
- $\mathcal{P}(L)$ is the label power set of L
- Y is the set of considered labels ($Y \subseteq \mathcal{P}(L)$), and
- $D = \{(x_1, y_1), \cdots, (x_n, y_n)\}$ is the data set of samples, where $x_i \in X$ and $y_i \in Y$

When $|L| = 2$, the problem is called binary classification, which is a well-studied problem in Text Classification. Since $|L| > 2$ for most practical problems, we need to extend the problem. This is usually done by combining binary classifiers in order to perform multi-class classification. In these cases, the combination is performed by using the well-known one-versus-rest or the one-versus-one strategies.

In the *one-versus-rest* strategy, we assume that we have $|L| > 2$ classes and we need to learn $|L|$ classifiers (multi-class classification). To do this, we use a transformation method of the data set D that consists of the construction of $|L|$ sample sets, in which every set considers those samples that belong to the kth class l_k as positive samples and the rest of the samples belong to the $\neg l_k$ class. From these new data sets, we learn $|L|$ classifiers denoted as f_k, where $k : 1 \cdots |L|$. We denote $P_k(x_i)$ as the probability assigned by the classifier f_k

to the input sample x_i. In order to classify a new sample x_i, we apply all the classifiers to it, and we select the classifier that maximizes Equation 1:

$$\hat{y}_i = \underset{k:1\cdots|L|}{\arg\max}\, P_k(x_i) \tag{1}$$

There are some problems (e.g., topic detection as defined at TASS) in which we need to associate more than one label to an example (multi-label classification). In this case, we have a set of labels $Y \subseteq \mathcal{P}(L)$ that can be assigned to the input sample x_i.

According to the solutions presented in (Tsoumakas and Katakis 2007), two solutions called *label power set* and *binary relevance* are proposed. In our system, we tackle the multi-label classification problem using three different approaches: the *binary relevance* approach, the *label power set* approach, and a *voting strategy* which combines mono-label classifiers.

In the *binary relevance approach*, we constructed one classifier per class following the one-versus-rest strategy. In this approach, $Y = L = \{y_1, \cdots, y_{|L|}\}$. We defined a *threshold*, ϵ, and we applied all the classifiers learned (f_k, where $k : 1 \cdots |L|$) for an input sample x_i. We chose those labels l_k that satisfied that the probability assigned by the classifier $f_k(x_i)$ was greater than ϵ, as shown in Equation 2.

$$y_i = \{l_k \in Y : P_k(x_i) > \epsilon, \forall k : 1 \cdots |L|\} \tag{2}$$

We set the *threshold* to 0.5, but other values of $\epsilon > 0.5$ or non-constant values of ϵ have been considered (Ramón Quevedo et al 2012).

Following the *label power set approach*, a new set of labels is defined. This new set should include all the combinations of labels from the set $\mathcal{P}(L)$. However, we only considered those combinations of labels that were seen in the training set. This approach has the disadvantage of not considering combinations that only appear in the test set.

With this new set of labels, we re-labeled the training set and we constructed a multi-class classifier following the one-versus-rest strategy. Due to the way we have defined the set Y, the classifier can assign a set of labels to an input sample x_i by maximizing Equation 3.

$$\hat{y}_i = \underset{k:1\cdots|Y|}{\arg\max}\, P_k(x_i), \text{where } y_i \in Y \tag{3}$$

In the *voting strategy approach*, using the data set, we learned a set of multi-class mono-label classifiers $\mathcal{C}$. For a new sample, x_i, let $N_k(x_i)$ be the number of mono-label classifiers that assigns the label l_k to the sample x_i. We assigned to the sample x_i all labels l_k that satisfied that $N_k(x_i)$ is greater than a certain value n, where $n \leq |\mathcal{C}|$ following Equation 4.

$$y_i = \{l_k \in L : N_k(x_i) > n, \forall k : 1 \cdots |L|\} \tag{4}$$

An overview of our system for the Sentiment Analysis problems studied is presented in Figure 1. Basically, the system consists of three modules: the Preprocessing module, the Learning module, and the Testing module.
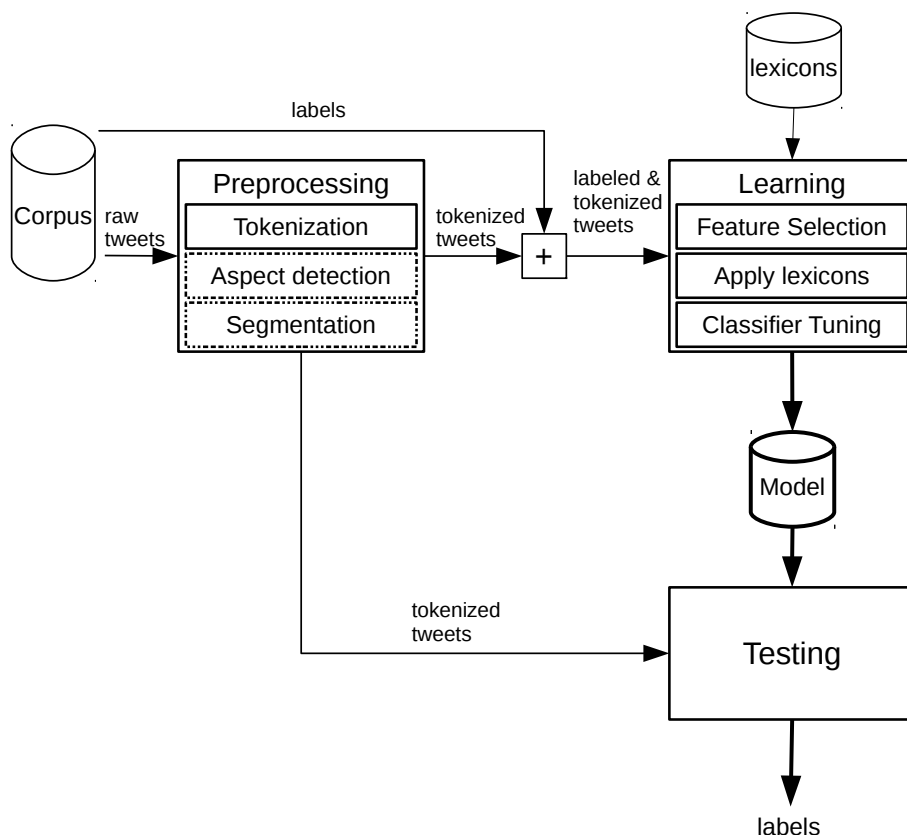
Fig. 1 Sentiment Analysis system description.

The Preprocessing module performs the tokenization of the input tweets. Optionally, aspect detector and tweet segmenter modules can be used in order to perform sentiment analysis at the aspect level. The Learning module constructs an SVM classifier to be used in the different classifications tasks. This module includes the feature selection process and the tuning of the parameters of the model. The learning process can also use some external resources such as polarity lexicons. In tasks such as topic detection, in which a tweet can belong to more than one topic, the module allows us to build multi-label classifiers for that purpose. The Testing module uses the learned model to assign the labels to each input tweet.

4.1 The preprocessing module

Before addressing the SA tasks, it is necessary to make a proper tokenization of the tweets that make up the training corpus and the test corpus. Although there are a lot of freely available tokenizers, they need to be adapted in order

to address the tokenization of a tweet. In our system, we decided to use and adapt available tools for tokenization, lemmatization, NER, and POS tagging. We adapted the package *Tweetmotif* described in (O'Connor et al 2010) in order to be able to process Spanish tweets. We also used *Freeling* (Padró and Stanilovsky 2012) (with the appropriate modifications for handling Twitter messages) for stemming, NER, and POS tagging. We added some functions to group special tokens into single tokens (e.g., *hashtags*, *web* addresses, *url*, *dates*, *numbers*, and some *punctuation* marks).

Saint André de Sangonis es una **d** las poblaciones más bonitas de Languedoc Rosellón **...** @user123 http://www.xxx.es #hashtagx Jajaja **!** **:)** **;)** 12 de junio de 2014.

(*Saint André de Sangonis is one of the most beautiful towns in Languedoc Roussillon ... @ user123 http://www.xxx.es #hashtagx* **LOL! :);)** *June 12, 2014.*)

Lemmatization + NER	POS Tagging	Feature Selected
saint_andré_de_sangonis	Proper_Noum	saint_andré_de_sangonis Proper_Noum
ser	Verb	ser Verb
uno	Determinant	---
de	Preposition	---
el	Determinant	---
población	Noum	población Noum
bonita	Adjective	bonita Adjective
más	Adverb	más Adverb
de	Preposition	---
languedoc_rosellón	Proper_Noum	languedoc_rosellón Proper_Noum
...	Punctuation_mark	---
@user123	<user>	---
http://www.xxx.es	<http>	---
#hashtagx	<hashtag>	<hashtag>
Jajaja	<laugh>	<laugh>
!	Punctuation_mark	! Punctuation_mark
:)	<happy>	<happy>
;)	<wink>	<wink>
12_de_junio_de_2014	<date>	<date>
.	Punctuation_mark	---

Fig. 2 Example of the preprocessing task.

Figure 2 shows an example of a tweet preprocessing. We obtained the lemmas (i. e., *ser*, to be, the infinitive of the verbal form *es*, is) and the NEs (i.e. *saint andré de sangonis*, that is a location). Different terms of Twitter domain are detected (i.e., *user*, *hashtag*, *laugh*). Moreover, the emoticon classifier detects and classifies the emoticons present in the tweet (i.e., *happy* and *wink*). Finally, we show the feature selected from all of this information. We considered the lemmas, their POS, and the special marked vocabulary. Prepositions, determinants, and some punctuation marks were removed.

The prepocessing module also includes two optional modules: the aspect detector module and tweet segmenter module. Both of them are related to the aspect-based SA task. Our aspect detector module is composed of a set of regular expressions to recognize the entities or aspects. This set includes regular expressions that match the aspect observed in the training set. In addition, this set can be extended by adding new regular expressions based on the knowledge of the task (e.g., Name Entities). The segmenter module is applied to define the context for every aspect or entity detected. At the moment, it is basically a window whose length is established during the tuning process. A more sophisticated approach for defining the context has been explored in (Pla and Hurtado 2014a). To determine the polarity at the aspect level, we first detect each aspect and we define its context in the tweet. Then, we use the Sentiment Analysis module but using the aspect as input and its context instead of the whole tweet.

4.2 The learning module

The learning phase is a tuning process that obtains the best set of features and the optimal parameters of the SVM model. The tuning process carried out had two objectives: to choose which features to include in the model, and to perform the estimation of the Support Vector Machine parameters. We conducted this optimization by means of a 10-fold cross validation process.

We tested different set of features: unigrams and n-grams of lemmas and words, emoticons, punctuation marks, polarity lexicons, negation, user name, user type, etc.

We explored different polarity lexicons to be included in our system. The external lexicons used in this work were lists of words and lemmas with their a priori polarity. One of the lexicons used was originally for English (Wilson et al 2005), which was translated into Spanish automatically, and another was the (Perez-Rosas et al 2012) lexicon, which is a list of words that was originally in Spanish. We constructed our original dictionary with these two resources.

Then, we combined this dictionary with the Elhuyar lexicon presented in (Saralegi and San Vicente 2013). Basically, the combination consisted of removing the words that had dual polarity from our dictionary and then combining it with the Elhuyar lexicon (Pla and Hurtado 2014b).

The system was implemented in Python using the *scikit-learn* package (Pedregosa et al 2011) and the LibLinear[8] external library. The developed system supports multi-class classification and multi-label classification using the binary relevance approach, the label power set approach, and the voting strategy approach.

---

[8] http://www.csie.ntu.edu.tw/~cjlin/liblinear/

4.3 The testing module

We performed an evaluation of the different tasks using the test data and the evaluation metrics proposed at the TASS competition for each task. We present the details of this evaluation in Section 6 .

5 Applying the SA system to the Political Tendency Identification task

The objective of the Political Tendency Identification task is to determine the political tendency of each user from their tweets in the test set of the General corpus in four possible values: $Left$, $Right$, $Centre$, and $Undefined$. We use the output of our SA system described in Section 4 to do this identification.

Our approach is based on a simple idea: if a user expresses positive opinions about parties of an ideology, we might think that the user shares the same ideology. In other words, if a user praises right-wing parties and/or criticizes left-wing parties this means that the user's ideology is $Right$. The $Left$ ideology is handled in the same way, but the $Centre$ ideology is a bit more complicated.

Thus, to classify users by their political orientation, we first identify entities associated with political parties by using the aspect detector module and we determine the context for every entity by using the segmenter module. Second, we analyze the polarity of these entities in the tweets of each user by means of our SA system.

We consider three types of entities: entities obtained from our NER module, which are basically proper names, Twitter users, and Twitter hashtags. Among all of the possible entities, we selected those containing the acronym for a political party or the name of a political leader. A total of 864 entities related to political parties and political leaders were detected.

We defined a tendency measure, $Tendency$, which assigned a value of -1 to those entities related to left-wing parties, a value of $+1$ to entities related to right-wing parties and a value of 0 to the entities related to centrist parties.

Next, we show how the political orientation of the users has been numerically calculated. For each user U_i of the General corpus, we obtain the set T_i, which includes all of their tweets that contain political entities. For users who do not have tweets that contain political entities, the $Undefined$ label is assigned.

For each tweet $T_{i_j} \in T_i$, $j = 1 \cdots |T_i|$, we identify the political entities that are contained in it. Let E_{i_j} be the set of entities of the tweet T_{i_j}. We denote each of the entities contained in E_{i_j} as $E_{i_{j_k}} \in E_{i_j}, k = 1 \cdots |E_{i_j}|$.

We obtained the polarity of each entity by using the system described in Section 4. After that, we assigned a numerical value to each polarity. In this respect, we assigned $Polarity = +1$ to the entities with positive polarity (label P), $Polarity = -1$ to the entities with negative polarity (label N), and

$Polarity = 0$ to the entities without polarity, that is, entities labeled with
NEU and NONE labels.

We combined[9] the $Tendency$ and $Polarity$ measures presented above to
define a new measure ($Political\_Tendency$) in order to obtain the political
orientation of each user.

$$Political\_Tendency(U_i) = \frac{\displaystyle\sum_{j=1\cdots|T_i|} \sum_{k=1\cdots|E_{i_j}|} Polarity(E_{i_{j_k}}) \cdot Tendency(E_{i_{j_k}})}{\displaystyle\sum_{j=1\cdots|T_i|} |E_{i_j}|}$$

(5)

From the $Political\_Tendency$ values obtained for each user, we classified
the user tendency taking into account the following: users without political en-
tities in their tweets are classified as $Undefined$; users with $Political\_Tendency$
between -0.05 and +0.05 are classified as $Centre$; users with $Political\_Tendency$
lower than -0.05 are classified as $Left$; and users with $Political\_Tendency$
greater than +0.05 are classified as $Right$. The results obtained using this
heuristic are analyzed in Section 6.

6 The results at TASS

In this section, we review the best results obtained by all the participating
teams in the different tasks of the TASS workshops. Most tasks have been
proposed in various editions and the results have usually improved with each
edition. For this reason, we only present the results obtained by the partici-
pants in the last editions of each task.

Although the results (and a brief description of the proposed models) of
all the participants are presented, in this paper, we will focus our analysis on
the results obtained by our team (UPV-ELiRF). We achieved the best results
in all tasks except in $Sentiment\ Analysis\ at\ the\ global\ level$ task considering 4
labels (we achieved second place - 0.725 compared to 0.726). We addressed all
of classification tasks using the SVM-based approach described in Section 4.

Bellow, we present a task-by-task analysis of the results obtained by all of
the participants in the last editions of all tasks proposed in the TASS work-
shops.

6.1 Sentiment Analysis at the global level

A total of 18 teams (mostly research labs and Universities from Spanish-
speaking countries) participated in the last two editions of the $Sentiment$
$Analysis\ at\ the\ global\ level$ task (TASS2015 and TASS2016). Table 3 shows

---

[9] We have considered multiple combination strategies. In this work, we present the com-
bination with the best results.

information about all 18 teams. This information includes a short description of the approaches used as well as the result of the best run of each team in both subtasks (*6 labels* and *4 labels*). The teams are sorted from highest to lowest according to the results obtained in the *6 labels* subtask. The measure proposed by the TASS organization for the evaluation of this task was *Accuracy*, which was defined as the number of correctly classified tweets divided by the total number of tweets. In other words, a tweet that was not classified by the system was considered to be an error[10].

Even though there were approaches based on using polarity lexicons and sociolinguistic features exclusively, the best results were obtained by machine-learning approaches. Most classifiers were based on the use of Support Vector Machines (SVM) and convolutional or recurrent Neural Networks (CNN and RNN). These classifiers represent the state-of-the-art in Sentiment Analysis tasks. Note that many participants implemented a mechanism for combining models in order to take advantage of the complementarity of the models.

The official results for *Accuracy* ranged from 0.673 to 0.410 for the 6-labels subtask and from 0.726 to 0.446 for the 4-labels subtask. Our team (UPV-ELiRF) obtained first place in the 6-labels subtask and second place in the 4-labels subtask. Our results are almost identical to those achieved by the second-place team (the LIF team) for both the 6-labels subtask and the 4-labels subtask. Although the LIF team used quite different approaches to ours, both teams obtained the best results when exploring the combination of models. We calculated the confidence interval at the 95% level of confidence[11] for this task achieving a value of ± 0.004. The Accuracy achieved by our system is not statistically significant with respect to the second team but the differences with the rest of participants are statistically significant.

The organization of the workshop conducted an analysis of the results of the latest edition (TASS2016) of the *Sentiment Analysis at the global level* with a 6-labels subtask. This analysis shows that there is a set of approximately 6% of the tweets that were not correctly classified by any of the systems proposed by the participants. In addition, only 1% of the tweets were correctly classified by all systems. This allows us to think that although all systems are based on the same paradigms (SVM and NN), they could actually be complementary in some way. As is usual in these tasks, the most difficult samples to classify (those in the 6% impossible set and many more) were tweets that involved sarcasm, irony, negation, or non-verbal information (eg. emojis). More details can be found in (García Cumbreras et al 2016).

Regarding the participation of our group in this task, we should highlight two aspects: the complementarity of models that do not have statistically significant differences and the behavior of the classifier when using polarity lexicons.

---

[10] For this task, *Accuracy* is equal to *Precision*, *Recall*, and $F_{beta=1}$ measures.

[11] The confidence interval is estimated by using the formula $A \pm 1.96\sqrt{\frac{A(1-A)}{N}}$, where A is the *Accuracy* and N is the number of samples in the test data set.

Table 3 The approaches and the best results of the teams participating in the *Sentiment Analysis at the global level* task (TASS2015 & TASS2016).

TEAM	APPROACH	4 labels	6 labels
UPV-ELiRF	Voting system combining different SVM configurations. (Hurtado et al 2015; Hurtado and Pla 2016)	0.725	0.673
LIF	Fusion system combining SVM + CNN + Doc2Vec	0.726	0.672
JACERONG	Logistic Regression classifiers using n-gram and basic features. (Cerón-Guzmán 2016)	0.705	0.620
GSI	SVM with lexical, morphosyntactic, and sentiment features. (Araque et al 2015)	0.690	0.618
DLSI	SMV with Skip-gram models. (Fernández et al 2015)	0.655	0.595
GTI-GRAD	Logistic regression classifier + unsupervised classifier. (Álvarez-López et al 2015, 2016)	0.695	0.592
LyS	Long Short-Term Memory (LSTM) with logistic function. (Vilares et al 2015)	0.664	0.568
DT	SVM + polarity lexicons. (Sixto-Cesteros et al 2015)	0.625	0.560
LABDA	SVM and Log. Regression classifiers using word embedding to represent the tweets. (Quirós et al 2016)	0.653	0.549
ITAINNOVA	Maxim entropy + ensemble classifier. (del-Hoyo-Alonso et al 2015)	0.610	0.549
ITAINNOVA	Maxim entropy + ensemble classifier. (del-Hoyo-Alonso et al 2015)	0.610	0.549
INGEOTEC	SVM classifier. (Siordia et al 2015)	0.633	0.535
BittenPotato	Voting among four classifiers (SVM, AdaBoost, Log. regression, Random Forest). (Mendizabal and Carandell 2015)	0.602	0.535
SINAI-EMMA	Linguistic resources (iSOL, sentiWordNet, Q-WordNet, etc.). (Martínez-Cámara et al 2015)	-	0.502
Columbia University	Adaptation to the Spanish language of the SA system of the CU based on Log. regression. (Santos-Deas et al 2015)	0.597	0.495
SINAI-wd2v	SVM using doc2vec + word2vec embeddings. (Díaz-Galiano and Montejo-Ráez 2015; Montejo-Ráez and Díaz-Galiano 2016)	0.619	0.474
TID-spark	Unsupervised classification based on sociolinguistic information. (Park 2015)	0.594	0.462
GAS-UCR	SVM using sociolinguistic features. (Casasola Murillo and Marín Raventós 2016)	0.446	0.410
UCSP	Voting among four classifiers(SVM, ME, DTrees, NaiveBayes). (Valverde-Tohalino and Tejada-Cárcamo 2015)	0.613	-

In our participations in the TASS workshop, we have always used an SVM with a linear kernel as a classifier. In each edition we tested different combinations of features for SVM models. However, in each edition we tested different combination of features in order to learn the SVM models. Thus, among others, we have tried the following: i) different tokenizations by activating (or not) the detection of Named Entities, the union in a single symbol of hashtag,

numbers, or user mentions; ii) the use of lemmas and words as features; iii) the size of n-grams; iv) negation treatment; v) using part of speech (POS) as features; vi) detection of repeated characters in a word; and vii) the use of polarity lexicons.

Unfortunately, except for the use of polarity lexicons, the results of the various combinations tested did not bring significant differences. However, we could observe that the tweets correctly labeled by each model were slightly different. For this reason, we decided to combine the different models. At first (TASS2014), we used a simple voting scheme, and later (TASS2015 and TASS2016), we used more sophisticated approaches. The best results were obtained using a meta-classifier based on SVM that determined the label of each sample from the output of 192 distinct models. Each one of the 192 models used a combination of the features mentioned above. In this way, we managed to obtain a 0.673 accuracy for the 6-label subtask by combining models that individually had a best accuracy of 0.6316

We also carried out a study of the effect of the use of polarity lexicons on the classification models. To do this, we learned a model similar to the one described above (the combination of 192 classifiers using a meta-classifier) but without using lexicons. In this way, we were able to compare the performance of both models.

The model that did not use lexicons obtained an accuracy of 0.617. This is a decrease of more than 0.055 in performance. A more detailed analysis shows that 12.68% of the test tweets are correctly labeled by the use of lexicons, but more than 7% of the tweets that were correctly labeled without lexicons were mislabeled when using lexicons. This means that the use of lexicons does not provide an incremental benefit. In addition, we also studied what kind of tweets had been labeled better. We thought that the use of lexicons would improve the labeling of tweets with polarity (P+, P, N, and N+ labels), but the result was exactly the opposite -the label that got the biggest increase was the NONE label. The precision of labels with negative polarity (N and N+ labels) decreased by the use of polarity lexicons.

6.2 Topic Detection

The *Topic Detection* task was proposed in the first three editions of the TASS workshop, from 2012 to 2014. It is the only multi-label task of all the organized tasks in the workshop. The metric used for the evaluation of the results was changing in each of the different editions. In the latest edition, the micro F-measure (F_1) was set as the official measure. Therefore, we use this measure to present the results.

The main difference of our proposal with the proposals of the other participants was that we addressed the task as a multi-label problem. We presented three runs which obtained pretty good results at TASS 2014. The three systems addressed the multi-label problem in different ways. The first system (run1) was a multi-label classifier based on the label power set approach; the

second one (run2) used the binary relevance strategy with the same threshold for all the classes; the third system (run3) used a voting scheme from eight mono-label classifiers. Labels predicted by at least three systems were selected.

Table 4 shows the results of all of the participating teams in the last edition of the *Topic Detection* task: ELiRF (Hurtado and Pla 2014), LyS (Vilares et al 2014), and IPN (Hernández Petlachi and Li 2014). *Precision* (P), *Recall* (R), and the F-measure (F_1) are shown. For this task, the differences among the results of the three teams are statistically significant at the 95% level of confidence.

Table 4 Official results for the last edition of the *Topic Detection* task (TASS2014).

System	P	R	F_1
UPV-ELiRF$_{run3}$	0.666	0.748	0.705
UPV-ELiRF$_{run2}$	0.700	0.707	0.703
UPV-ELiRF$_{run1}$	0.682	0.689	0.686
LyS$_{run1}$	0.680	0.600	0.640
LyS$_{run2}$	0.680	0.590	0.630
IPN	0.270	0.330	0.300

When comparing our best result (run3) with the results of the Lys team, the second best team, it can we observed that although they obtained a slightly better *Precision* (0.680 compared to 0.666), our system got much better *Recall* (0.748 compared to 0.600), and, therefore, we got a better F_1 (0.705 compared to 0.640). We believe that, as mentioned above, our better result is because we developed a multi-label system. The low *Precision* of our proposal (0.666) may be due to the selection criterion used. Perhaps, choosing a label with only three votes is too permissive and, consequently, it increases the number of false positives.

6.3 Aspect Detection and Aspect-based Sentiment Analysis

During various editions of the TASS workshop, up to four different *Aspect-based Sentiment Analysis* (ABSA) tasks have been proposed using three different corpora. The corpora used, as described in Subsection 3.2 were: the *Politics* corpus (TASS2013), the *STOMPOL* corpus (TASS2015 and TASS2016), and the *Social-TV* corpus (TASS2014 and TASS2015). In addition, in the 2014 edition, the participants had to find the aspects before labeling them. In all cases, more than one aspect could appear in each tweet.

Note that the sizes of the test sets used for these tasks are much smaller than those used in the previous tasks. Consequently, in some cases, although the differences of the results among participating teams are appreciable, they are not statistically significant at the 95% level of confidence.

The first ABSA task proposed in the TASS workshop used the *Politics* corpus. In this corpus, the aspects refer to political entities (parties, politicians,

or political events). The aspects were already identified in the corpus. The task consisted of establishing the scope of each aspect (context) and determining its polarity.

The organizers only provided the test set, so each team have to learn their models using external resources. We used the training set of the *General* corpus to learn our polarity models. We also developed a heuristic algorithm to determine the scope of each aspect. This algorithm was based on the position of the aspect within the tweet, the relative position of the aspect compared to the other aspects, and the punctuation. Table 5 shows the results obtained in this task.

Table 5 Best official results for the *Aspect-based Sentiment Analysis* task using the *Politics* corpus (TASS2013).

	System	P	R	F_1
UPV-ELiRF	(Pla and Hurtado 2013a)	0.395	0.395	0.395
CITIUS	(Gamallo et al 2013)	0.411	0.378	0.394
SINAI-CESA	(Montejo-Ráez et al 2013)	0.384	0.384	0.384
ETH	(García and Thelwall 2013)	0.307	0.307	0.307

As shown in Table 5, the results were not very good for any of the participants. This was a very difficult task, in which frequently more than one aspect appeared in the same tweet. Nevertheless, we got the best results (0.395) with little difference from the second-place team (0.394). Since our system classified all the aspects, the values of P, R, F_1 are equal. In addition, the corpus that was used to learn the models was very different from the *Politics* corpus (the test set). Therefore, the results were not as expected. When we analyzed the results, we concluded that the algorithm that we developed to determine the scope of each aspect did not include all cases. From then on, we decided to use a simple window around the aspect to fix its scope.

In 2014, the organizers proposed a slightly more complicated ABSA task. Using the Social-TV corpus, the participants first had to detect the aspects (entities related to the Final of the Copa del Rey Championship in Spain), and second, they had to determine the polarity of each detected aspect. The corpus included a small training set (1,773 tweets) and a test set (1,000 tweets). In order to detect the aspects, we defined a set of regular expressions that included instances of entities that appeared in the training set and the names of the players and coaches of the two teams that played the final. The scope of each aspect was determined by a window whose size was fixed during the tuning process. Table 6 shows the results obtained in the detection of the aspects (the first phase).

When our results (UPV-ELiRF) are compared with the results of the second-place team (LyS), it can be observed that the basic difference is *Precision* (0.906 compared to 0.810). We think that this is because our proposal, which is based on regular expressions, is more accurate. It detects approxi-

Table 6 Best official results for the *Aspect Detection* task (TASS2014).

	System	P	R	F_1
UPV-ELiRF	(Hurtado and Pla 2014)	0.906	0.911	0.909
LyS	(Vilares et al 2014)	0.810	0.900	0.850

mately the same number of aspects (*Recall* is very similar), but it minimizes the number of false positives.

For the second phase (the *Aspect-based Sentiment Analysis* phase), we learned the model using only the training set of the *Social-TV corpus*. Both the model parameters and the window size were determined by a 10-fold cross validation tuning process. Table 7 shows the results of the ABSA phase. Since the aspects had to be detected, the evaluation was made in terms of *Precision*, *Recall*, and F_1.

Table 7 Best official results for the *Aspect-based Sentiment Analysis* task using the Social-TV corpus (TASS2014). This task uses the aspects detected in the previous phase.

	System	P	R	F_1
UPV-ELiRF	(Hurtado and Pla 2014)	0.578	0.596	0.587
LyS	(Vilares et al 2014)	0.520	0.580	0.550

It can be seen that, the greatest difference in the performance of the two systems is again *Precision* while *Recall* is more similar. We think that the two polarity systems had a similar performance, but we obtained better overall results because our system detected the aspects better.

In 2015, two ABSA tasks were proposed by the TASS organization. These tasks used the *Social-TV* corpus and the *STOMPOL* corpus. In both cases, the aspects were already marked in the test set and the polarity had be determined using three labels (P, N, NEU). Similarly to the previous task, the scope of the aspects and the model parameters were fixed during a 10-fold cross-validation tuning process.

Table 8 shows the results of the best runs of all of the participants in the ABSA task for the *Social-TV* corpus. These results can be compared with those previously shown in Table 7.

Table 8 Best official results for the *Aspect-based Sentiment Analysis* task using the Social-TV corpus (TASS2015).

	System	Accuracy
UPV-ELiRF	(Hurtado et al 2015)	0.655
GSI	(Araque et al 2015)	0.635
TID	(Park 2015)	0.631
Lys	(Vilares et al 2015)	0.610

Our system was again the winner of the task with 0.655 for *Accuracy*. Overall, the results were better than those achieved the previous year with the same corpus (Table 7). In our case, we improved *Accuracy* from 0.587 to 0.655. Apart from improving our polarity system, this improvement is because the task in the previous edition included the detection of the aspects. Although our proposal was able to identify most aspects (0.911 of *Recall*), there was approximately 10% of undetected aspects that might justify the difference in the performance of the two systems.

The second ABSA task proposed in 2015 used the *STOMPOL* corpus. Table 9 shows the results obtained by all of the participating teams. We achieved the best result. This result is in line with the one obtained for the same task with the *Social-TV* corpus.

Table 9 Best official results for the *Aspect-based Sentiment Analysis* task using the STOMPOL corpus (TASS2015 & TASS2016).

	System	Accuracy
UPV-ELiRF	(Hurtado et al 2015; Hurtado and Pla 2016)	0.633
GTI	(Álvarez-López et al 2016)	0.606
LyS	(Vilares et al 2015)	0.599
TID	(Park 2015)	0.557

6.4 Political Tendency Identification

The *Political Tendency Identification* (PTI) task was only proposed in the TASS edition of 2013. This task consists of determining the political orientation of a user by analyzing his tweets. Three political labels (*Right*, *Left* and *Centre*) were defined. Another label, *Undefined*, was also included to label those users who did not express political opinions. As described in Section 5, instead of training a classifier for each tendency, we use our Aspect-based Sentiment Analysis system to determine the polarity of the opinions that each user has about the political parties. Thus, by considering all the political opinions of a user, we are able to determine the user's political tendency.

The measures used to evaluate the PTI task were *Precision* (P), *Recall* (R), and the F-measure (F_1). These measures were used because participants could leave some users unlabeled. Table 10 shows the results of the best runs of the teams participating in the task.

Our proposal achieved the best results over the other participants that used classifiers. To better understand the results, we have carried out a study of the performance of our system for each political tendency. Table 11 shows a class-level study of the results obtained by our system.

The class that obtained the best results was the *Undefined* class. Users who did not express opinions on political entities were assigned to this class.

Table 10 Best official results for the *Political Tendency Identification* task (TASS2013).

	System	P	R	F_1
UPV-ELiRF	(Pla and Hurtado 2013a)	0.703	0.703	0.703
SINAI-CESA	(Montejo-Ráez et al 2013)	0.583	0.399	0.474
LyS	(Vilares et al 2013)	0.424	0.424	0.424

Table 11 Results at the class level of our proposal for the *Political Tendency Identification* task (TASS2013).

Tendency	P	R	F_1
Left	0.658	0.735	0.694
Centre	0.478	0.393	0.431
Right	0.786	0.698	0.739
Undefined	0.780	0.970	0.865
Global	0.703	0.703	0.703

The *Recall* of this class is very high (0.970), which means that most users without political opinions were identified. However, *Accuracy* was lower (0.780). We think that this is due to the non-detection of political entities, and, therefore, users were labeled as $Undefined$ when, in fact, they did express political opinions.

The class with the worst results was the *Centre* class. It is hard for our system to identify users with this tendency. Therefore, a different method needs to be found in order to detect users who are centrist (i.e., they criticize or praise both right-wing and left-wing parties).

7 Conclusions and Future Work

In this paper, we have presented our approach for the Sentiment Analysis tasks at the TASS competition. TASS has become a benchmark for Sentiment Analysis tasks in the Spanish language. We developed a system based on the Support Vector Machine formalism for classification problems. The system allows us to perform the SA analysis tasks at both the global level and at the aspect level in Twitter. In addition, for the topic classification task, the system allows multi-label classification to assign more than one topic to a tweet. The output of the aspect-based Sentiment Analysis system has been successfully applied to the political tendency identification task.

Our team achieved the best results in all of the tasks. Note that our results were the best for the newest tasks such as Tendency Identification and SA at the aspect-level.

We want to highlight the flexibility of the presented system, which has allowed us to participate in all of the Sentiment Analysis tasks proposed at TASS using the same system and achieving good results. In addition, the proposed system has demonstrated its usefulness even for tasks that are not strictly

classification tasks, such as political tendency identification task, in which we used the polarity of the tweets at the aspect level as relevant information.

Moreover, we tested our system in languages other than the Spanish language. We have also participated in the Sentiment Analysis tasks for English in the SemEval competition with slight modifications to our system. The unique modification consisted of using the SemEval official corpora and changing the polarity lexicons. The results are promising but not as good as those obtained at the TASS workshop.

In light of the obtained results, as future work, we plan to throughly explore more complex ways for combining the systems. We also plan to consider new machine-learning paradigms to be included in the combination. We are currently working on the development of a web-based platform that will allow the free use of the system described in this paper.

References

Álvarez-López T, Juncal-Martínez J, Fernández-Gavilanes M, Costa-Montenegro E, González-Castaño FJ, Cerezo-Costas H, Celix-Salgado D (2015) GTI-Gradiant at TASS 2015: A Hybrid Approach for Sentiment Analysis in Twitter. In: Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN co-located with 31st SEPLN Conference (SEPLN 2015), Alicante, Spain, September 15, 2015., pp 35–40

Álvarez-López T, Fernández-Gavilanes M, García-Méndez S, Juncal-Martínez J, González-Castaño FJ (2016) GTI at TASS 2016: Supervised Approach for Aspect Based Sentiment Analysis in Twitter. In: Proceedings of TASS 2016: Workshop on Sentiment Analysis at SEPLN co-located with 32nd SEPLN Conference (SEPLN 2016), Salamanca, Spain, September 13th, 2016., pp 53–57

Araque O, Corcuera I, Román C, Iglesias CA, Sánchez-Rada JF (2015) Aspect Based Sentiment Analysis of Spanish Tweets. In: Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN co-located with 31st SEPLN Conference (SEPLN 2015), Alicante, Spain, September 15, 2015., pp 29–34

Balahur A, Perea-Ortega JM (2013) Experiments using varying sizes and machine translated data for sentiment analysis in Twitter. In: Proceedings of the TASS workshop at SEPLN 2013, IV Congreso Español de Informática

Barbosa L, Feng J (2010) Robust sentiment detection on Twitter from biased and noisy data. In: Proceedings of the 23rd International Conference on Computational Linguistics: Posters, Association for Computational Linguistics, pp 36–44

Batista F, Ribeiro R (2012) The L2F Strategy for Sentiment Analysis and Topic Classification. Tech. rep., URL http://www.sepln.org/workshops/tass/2012/participation.php

Casasola Murillo E, Marín Raventós G (2016) Evaluación de Modelos de Representación del Texto con Vectores de Dimensiónn Reducida para Análisis de Sentimiento. In: Proceedings of TASS 2016: Workshop on Sentiment Analysis at SEPLN co-located with 32nd SEPLN Conference (SEPLN 2016), Salamanca, Spain, September 13th, 2016., pp 23–28

Castellano A, Cigarrán J, García-Serrano A (2012) UNED @ TASS: Using IR techniques for topic-based sentiment analysis through divergence models. Tech. rep., URL http://www.sepln.org/workshops/tass/2012/participation.php

Castellanos-González A, Cigarrán-Recuero J, García-Serrano A (2013) UNED LSI @ TASS 2013: Considerations about Textual Representation for IR based Tweet Classification.

In: Proceedings of the TASS workshop at SEPLN 2013, IV Congreso Español de Informática

Cerón-Guzmán JA (2016) JACERONG at TASS 2016: An Ensemble Classifier for Sentiment Analysis of Spanish Tweets at Global Level. In: Proceedings of TASS 2016: Workshop on Sentiment Analysis at SEPLN co-located with 32nd SEPLN Conference (SEPLN 2016), Salamanca, Spain, September 13th, 2016., pp 35–39

del-Hoyo-Alonso R, Hupont I, Lacueva F (2013) Affective Polarity word discovering by means of Artificial General Intelligence techniques. In: Proceedings of the TASS workshop at SEPLN 2013, IV Congreso Español de Informática

del-Hoyo-Alonso R, de la Vega Rodrigalvarez-Chamorro M, Vea-Murguía J, Montañes-Salas RM (2015) Ensemble Algorithm with Syntactical Tree Features to Improve the Opinion Analysis. In: Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN co-located with 31st SEPLN Conference (SEPLN 2015), Alicante, Spain, September 15, 2015., pp 53–58

Deriu J, Gonzenbach M, Uzdilli F, Lucchi A, De Luca V, Jaggi M (2016) Swisscheese at semeval-2016 task 4: Sentiment classification using an ensemble of convolutional neural networks with distant supervision. In: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), Association for Computational Linguistics, San Diego, California, pp 1124–1128, URL http://www.aclweb.org/anthology/S16-1173

Díaz-Galiano MC, Montejo-Ráez A (2015) Participación de SINAI DW2Vec en TASS 2015. In: Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN co-located with 31st SEPLN Conference (SEPLN 2015), Alicante, Spain, September 15, 2015., pp 59–64

Fernández J, Gutiérrez Y, Tomás D, Gómez JM, Martínez-Barco P (2015) Evaluating a Sentiment Analysis Approach from a Business Point of View. In: Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN co-located with 31st SEPLN Conference (SEPLN 2015), Alicante, Spain, September 15, 2015., pp 93–98

Fernández J, Gutiérrez Y, Gómez JM, Martínez-Barco P, Montoyo A, Muñoz R (2013) Sentiment Analysis of Spanish Tweets Using a Ranking Algorithm and Skipgrams. In: Proceedings of the TASS workshop at SEPLN 2013, IV Congreso Español de Informática

Frank E, Hall MA, Witten IH (2016) The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", fourth edition edn. Morgan Kaufmann

Gamallo P, García M, Fernández-Lanza S (2013) TASS: A Naive-Bayes strategy for sentiment analysis on Spanish tweets. In: Proceedings of the TASS workshop at SEPLN 2013, IV Congreso Español de Informática

García Cumbreras MÁ, Martínez Cámara E, Villena-Román J, García Morera J (2016) TASS 2015—The Evolution of the Spanish Opinion Mining Systems. Procesamiento del Lenguaje Natural

García Cumbreras MÁ, Villena Román J, Martínez Cámara E, Díaz Galiano MC, Martín Valdivia MT, Ureña López LA (2016) Overview of TASS 2016. In: Proceedings of TASS 2016: Workshop on Sentiment Analysis at SEPLN co-located with 32nd SEPLN Conference (SEPLN 2016), Salamanca, Spain, September 13th, 2016., pp 13–21

García D, Thelwall M (2013) Political alignment and emotional expression in Spanish Tweets. In: Proceedings of the TASS workshop at SEPLN 2013, IV Congreso Español de Informática

Hagen M, Potthast M, Büchner M, Stein B (2015) Webis: An ensemble for twitter sentiment detection. In: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Association for Computational Linguistics, Denver, Colorado, pp 582–589, URL http://www.aclweb.org/anthology/S15-2097

Hamdan H, Bellot P, Bechet F (2015) Lsislif: Crf and logistic regression for opinion target extraction and sentiment polarity analysis. In: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Association for Computational Linguistics, Denver, Colorado, pp 753–758, URL http://www.aclweb.org/anthology/S15-2128

Hernández Petlachi R, Li X (2014) Análisis de sentimiento sobre textos en Español basado en aproximaciones semánticas con reglas lingüísticas. In: Proceedings of the TASS workshop at SEPLN 2014

Hurtado LF, Pla F (2014) ELiRF-UPV en TASS 2014: Análisis de Sentimientos, Detección de Tópicos y Análisis de Sentimientos de Aspectos en Twitter. In: Proceedings of the TASS workshop at SEPLN 2014

Hurtado LF, Pla F (2016) ELiRF-UPV en TASS 2016: Análisis de Sentimientos en Twitter. In: Proceedings of TASS 2016: Workshop on Sentiment Analysis at SEPLN co-located with 32nd SEPLN Conference (SEPLN 2016), Salamanca, Spain, September 13th, 2016., pp 47–51

Hurtado LF, Pla F, Buscaldi D (2015) ELiRF-UPV en TASS 2015: Análisis de Sentimientos en Twitter. In: Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN co-located with 31st SEPLN Conference (SEPLN 2015), Alicante, Spain, September 15, 2015., pp 75–79

Jansen BJ, Zhang M, Sobel K, Chowdury A (2009) Twitter power: Tweets as electronic word of mouth. Journal of the American society for information science and technology 60(11):2169–2188

Jiménez Zafra SM, Martínez Cámara E, Martín Valdivia MT, Ureña López LA (2014) SINAI-ESMA: An unsupervised approach for Sentiment Analysis in Twitter. In: Proceedings of the TASS workshop at SEPLN 2014

Liu B (2012) Sentiment Analysis and Opinion Mining. A Comprehensive Introduction and Survey. Morgan & Claypool Publishers

Liu B, Hu M, Cheng J (2005) Opinion Observer: Analyzing and Comparing Opinions on the Web. In: Proceedings of the 14th International Conference on World Wide Web, ACM, New York, NY, USA, WWW '05, pp 342–351, DOI 10.1145/1060745.1060797, URL http://doi.acm.org/10.1145/1060745.1060797

Martínez-Cámara E, Martín-Valdivia MT, Ureña-López LA, Montejo-Raéz A (2014) Sentiment analysis in Twitter. Natural Language Engineering 1(1):1–28

Martínez-Cámara E, García-Cumbreras MÁ, Martín-Valdivia MT, López LAU (2015) SINAI-EMMA: Vectores de Palabras para el Análisis de Opiniones en Twitter. In: Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN co-located with 31st SEPLN Conference (SEPLN 2015), Alicante, Spain, September 15, 2015., pp 41–46

Martín-Wanton T, de Albornoz JC (2012) UNED at TASS 2012: Polarity Classification and Trending Topic System. Tech. rep., URL http://www.sepln.org/workshops/tass/2012/participation.php

Martínez-Cámara E, Ángel García-Cumbreras M, Martín-Valdivia MT, Ureña-López LA (2013) SINAI-EMML: Combinación de Recursos Lingüíticos para el Análisis de la Opinión en Twitter. In: Proceedings of the TASS workshop at SEPLN 2013, IV Congreso Español de Informática

Martínez-Cámara E, Martín-Valdivia MT, Molina-González MD, Ureña-López LA (2013) Bilingual Experiments on an Opinion Comparable Corpus. In: Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, p 87–93

Mendizabal I, Carandell J (2015) BittenPotato: Tweet Sentiment Analysis by Combining Multiple Classifiers. In: Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN co-located with 31st SEPLN Conference (SEPLN 2015), Alicante, Spain, September 15, 2015., pp 71–74

Mohammad S, Kiritchenko S, Zhu X (2013) Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. In: Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), Association for Computational Linguistics, Atlanta, Georgia, USA, pp 321–327, URL http://www.aclweb.org/anthology/S13-2053

Montejo-Ráez A, Díaz-Galiano MC (2016) Participación de SINAI en TASS 2016. In: Proceedings of TASS 2016: Workshop on Sentiment Analysis at SEPLN co-located with 32nd SEPLN Conference (SEPLN 2016), Salamanca, Spain, September 13th, 2016., pp 41–45

Montejo-Ráez A, Díaz-Galiano MC, García-Vega M (2013) LSA based approach to TASS 2013. In: Proceedings of the TASS workshop at SEPLN 2013, IV Congreso Español de Informática

Montejo-Ráez A, García-Cumbreras M, Díaz-Galiano M (2014) Participación de SINAI Word2Vec en TASS 2014. In: Proceedings of the TASS workshop at SEPLN 2014

Moreno-Ortiz A, Pérez-Hernández C (2012) Lexicon-Based Sentiment Analysis of Twitter Messages in Spanish. Tech. rep., URL http://www.sepln.org/workshops/tass/2012/participation.php

Nakov P, Kozareva Z, Ritter A, Rosenthal S, Stoyanov V, Wilson T (2013) SemEval-2013 Task 2: Sentiment Analysis in Twitter

Nakov P, Ritter A, Rosenthal S, Stoyanov V, Sebastiani F (2016) SemEval-2016 Task 4: Sentiment Analysis in Twitter. In: Proceedings of the 10th International Workshop on Semantic Evaluation, Association for Computational Linguistics, San Diego, California, SemEval '16, p 1–18

O'Connor B, Krieger M, Ahn D (2010) TweetMotif: Exploratory Search and Topic Summarization for Twitter. In: Cohen WW, Gosling S (eds) Proceedings of the Fourth International Conference on Weblogs and Social Media, ICWSM 2010, Washington, DC, USA, May 23-26, 2010, The AAAI Press, DOI http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1540

Padró L, Stanilovsky E (2012) FreeLing 3.0: Towards Wider Multilinguality. In: Proceedings of the Language Resources and Evaluation Conference (LREC 2012), ELRA, Istanbul, Turkey

Pang B, Lee L, Vaithyanathan S (2002) Thumbs up? Sentiment Classification using Machine Learning Techniques. In: IN PROCEEDINGS OF EMNLP, pp 79–86

Park S (2015) Sentiment Classification Using Sociolinguistic Clusters. In: Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN co-located with 31st SEPLN Conference (SEPLN 2015), Alicante, Spain, September 15, 2015., pp 99–104

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12:2825–2830

Perea-Ortega JM, Balahur A (2014) Experiments on feature replacements for polarity classification of Spanish tweets. In: Proceedings of the TASS workshop at SEPLN 2014

Perez-Rosas V, Banea C, Mihalcea R (2012) Learning Sentiment Lexicons in Spanish. In: Chair) NCC, Choukri K, Declerck T, Doğan MU, Maegaard B, Mariani J, Odijk J, Piperidis S (eds) Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12), European Language Resources Association (ELRA), Istanbul, Turkey

Pla F, Hurtado LF (2013a) ELiRF-UPV en TASS-2013: Análisis de sentimientos en Twitter. In: Proceedings of the TASS workshop at SEPLN 2013, IV Congreso Español de Informática

Pla F, Hurtado LF (2013b) ELiRF-UPV en TASS-2013: Análisis de sentimientos en Twitter. XXIX Congreso de la Sociedad Espanola para el Procesamiento del Lenguaje Natural (SEPLN 2013) TASS pp 220–227

Pla F, Hurtado LF (2014a) Political Tendency Identification in Twitter using Sentiment Analysis Techniques. In: Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, Dublin City University and Association for Computational Linguistics, Dublin, Ireland, pp 183–192, URL http://www.aclweb.org/anthology/C14-1019

Pla F, Hurtado LF (2014b) Sentiment analysis in Twitter for Spanish. In: International Conference on Applications of Natural Language to Data Bases/Information Systems, Springer International Publishing, pp 208–213

Quirós A, Segura-Bedmar I, Martínez P (2016) LABDA at the 2016 TASS Challenge Task: Using Word Embeddings for the Sentiment Analysis Task. In: Proceedings of TASS 2016: Workshop on Sentiment Analysis at SEPLN co-located with 32nd SEPLN Conference (SEPLN 2016), Salamanca, Spain, September 13th, 2016., pp 29–33

Ramón Quevedo J, Luaces O, Bahamonde A (2012) Multilabel Classifiers with a Probabilistic Thresholding Strategy. Pattern Recogn 45(2):876–883

Rosenthal S, Nakov P, Ritter A, Stoyanov V (2014) SemEval-2014 Task 9: Sentiment Analysis in Twitter. In: Nakov P, Zesch T (eds) Proceedings of the 8th International Workshop on Semantic Evaluation, Dublin, Ireland, SemEval '14

Rosenthal S, Nakov P, Kiritchenko S, Mohammad S, Ritter A, Stoyanov V (2015) SemEval-2015 Task 10: Sentiment Analysis in Twitter. In: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Association for Computational Linguistics, Denver, Colorado, pp 451–463, URL http://www.aclweb.org/anthology/S15-2078

Rouvier M, Favre B (2016) Sensei-lif at semeval-2016 task 4: Polarity embedding fusion for robust sentiment analysis. In: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), Association for Computational Linguistics, San Diego, California, pp 202–208, URL http://www.aclweb.org/anthology/S16-1030

San Vicente Roncal I, Saralegi Urizar X (2014) Looking for Features for Supervised Tweet Polarity Classification. In: Proceedings of the TASS workshop at SEPLN 2014

Santos-Deas M, Biran O, McKeown K, Rosenthal S (2015) Spanish Twitter Messages Polarized through the Lens of an English System. In: Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN co-located with 31st SEPLN Conference (SEPLN 2015), Alicante, Spain, September 15, 2015., pp 81–86

Saralegi X, San Vicente I (2012) TASS: Detecting Sentiments in Spanish Tweets. Tech. rep., URL http://www.sepln.org/workshops/tass/2012/participation.php

Saralegi X, San Vicente I (2013) Elhuyar at TASS 2013. In: Proceedings of the TASS workshop at SEPLN 2013, IV Congreso Español de Informática

Sebastiani F (2002) Machine Learning in Automated Text Categorization. ACM Comput Surv 34(1):1–47, DOI 10.1145/505282.505283, URL http://doi.acm.org/10.1145/505282.505283

Segura-Bedmar I, Quiròs A, Martìnez P (2017) Exploring Convolutional Neural Networks for Sentiment Analysis of Spanish tweets. In: Proceedings of EACL (15th Conference of the European Chapter of the Association for Computational Linguistics), Association for Computational Linguistics, pp 1014–1022

Severyn A, Moschitti A (2015) Unitn: Training deep convolutional neural network for twitter sentiment classification. In: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Association for Computational Linguistics, Denver, Colorado, pp 464–469, URL http://www.aclweb.org/anthology/S15-2079

Siordia OS, Moctezuma D, Graff M, Miranda-Jiménez S, Téllez ES, Villaseñor E (2015) Sentiment Analysis for Twitter: TASS 2015. In: Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN co-located with 31st SEPLN Conference (SEPLN 2015), Alicante, Spain, September 15, 2015., pp 65–70

Sixto-Cesteros J, Almeida A, López-de-Ipiña D (2015) DeustoTech Internet at TASS 2015: Sentiment Analysis and Polarity Classification in Spanish Tweets. In: Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN co-located with 31st SEPLN Conference (SEPLN 2015), Alicante, Spain, September 15, 2015., pp 23–28

Trilla A, Alías F (2012) Sentiment Analysis of Twitter messages based on Multinomial Naive Bayes. Tech. rep., URL http://www.sepln.org/workshops/tass/2012/participation.php

Tsoumakas G, Katakis I (2007) Multi-label classification: An overview. Int J Data Warehousing and Mining 2007:1–13

Turney PD (2002) Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In: ACL, pp 417–424, DOI http://www.aclweb.org/anthology/P02-1053.pdf

Valverde-Tohalino J, Tejada-Cárcamo J (2015) Comparing Supervised Learning Methods for Classifying Spanish Tweets. In: Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN co-located with 31st SEPLN Conference (SEPLN 2015), Alicante, Spain, September 15, 2015., pp 87–92

Vilares D, Alonso MA, Gómez-Rodríguez C (2013) LyS at TASS 2013: Analysing Spanish tweets by means of dependency parsing, semantic-oriented lexicons and psychometric word-properties. In: Proceedings of the TASS workshop at SEPLN 2013, IV Congreso Español de Informática

Vilares D, Doval Y, Alonso MA, Gómez-Rodríguez C (2014) LyS at TASS 2014: A Prototype for Extracting and Analysing Aspects from Spanish tweets. In: Proceedings of the TASS workshop at SEPLN 2014

Vilares D, Doval Y, Alonso MA, Gómez-Rodríguez C (2015) LyS at TASS 2015: Deep Learning Experiments for Sentiment Analysis on Spanish Tweets. In: Proceedings of

TASS 2015: Workshop on Sentiment Analysis at SEPLN co-located with 31st SEPLN Conference (SEPLN 2015), Alicante, Spain, September 15, 2015., pp 47–52

Villar Rodríguez E, Torre Bastida AI, García Serrano A, González Rodríguez M (2013) TECNALIA-UNED @ TASS: Uso de un enfoque lingüístico para el análisis de sentimientos. In: Proceedings of the TASS workshop at SEPLN 2013, IV Congreso Español de Informática

Villena-Román J, García Morera J, García Cumbreras MÁ, Martínez Cámara E, Martín Valdivia MT, Ureña López LA (2013a) Workshop on Sentiment Analysis at SEPLN 2013: An overview. In: Proceedings of the TASS workshop at SEPLN 2013, Villena-Román, Julio; García Morera, Janine; García Cumbreras, Miguel Ángel; Martínez Cámara, Eugenio; Martín Valdivia, M. Teresa; Ureña López, L. Alfonso

Villena-Román J, Lana-Serrano S, Martínez-Cámara E, González-Cristóbal JC (2013b) TASS-Workshop on Sentiment Analysis at SEPLN. Procesamiento del Lenguaje Natural 50:37–44

Villena-Román J, García Morera J, García Cumbreras MÁ, Martínez Cámara E, Martín Valdivia MT, Ureña López LA (2014) Workshop on Sentiment Analysis at SEPLN: Overview. In: Proceedings of the TASS workshop at SEPLN 2014, Villena-Román, Julio; García Morera, Janine; García Cumbreras, Miguel Ángel; Martínez Cámara, Eugenio; Martín Valdivia, M. Teresa; Ureña López, L. Alfonso

Villena-Román J, García Morera J, García Cumbreras MÁ, Martínez Cámara E, Martín Valdivia MT, Ureña López LA (2015) Overview of TASS 2015. In: TASS 2015: Workshop on Sentiment Analysis at SEPLN co-located with 31st SEPLN Conference (SEPLN 2015).http://ceur-ws.org/Vol-1397/, Villena-Román, Julio; García Morera, Janine; García Cumbreras, Miguel Ángel; Martínez Cámara, Eugenio; Martín Valdivia, M. Teresa; Ureña López, L. Alfonso, pp 13–21

Villena-Román J, García-Morera J (2013) Workshop on Sentiment Analysis at SEPLN 2013: An over view. In: Proceedings of the TASS workshop at SEPLN 2013, IV Congreso Español de Informática

Vinodhini G, Chandrasekaran R (2012) Sentiment Analysis and Opinion Mining: A Survey. International Journal 2(6)

Wilson T, Hoffmann P, Somasundaran S, Kessler J, Wiebe J, Choi Y, Cardie C, Riloff E, Patwardhan S (2005) OpinionFinder: A system for subjectivity analysis. In: Proceedings of HLT/EMNLP on Interactive Demonstrations, Association for Computational Linguistics, pp 34–35

Xu S, Liang H, Baldwin T (2016) Unimelb at semeval-2016 tasks 4a and 4b: An ensemble of neural networks and a word2vec based model for sentiment classification. In: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), Association for Computational Linguistics, San Diego, California, pp 183–189, URL http://www.aclweb.org/anthology/S16-1027

Zhu X, Kiritchenko S, Mohammad S (2014) Nrc-canada-2014: Recent improvements in the sentiment analysis of tweets. In: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Association for Computational Linguistics and Dublin City University, Dublin, Ireland, pp 443–447, URL http://www.aclweb.org/anthology/S14-2077