UNED en TASS 2012: Systema para la Clasificación de la Polaridad y Seguimiento de Temas*

UNED at TASS 2012: Polarity Classification and Trending Topic System

Tamara Martín-Wanton
UNED NLP & IR Group
Juan del Rosal, 16
28040 Madrid, Spain
tmartin@lsi.uned.es

Jorge Carrillo de Albornoz
UNED NLP & IR Group
Juan del Rosal, 16
28040 Madrid, Spain
jcalbornoz@lsi.uned.es

Resumen: Los medios sociales, tales como blogs, foros y redes, ofrecen un excelente escenario para compartir información y conectar personas. La información vertida en estos medios es de gran interés tanto para empresas como para particulares. Sin embargo, el gran volumen en que se presenta limita su utilidad a menos que se disponga de herramientas eficientes para su manejo. En este contexto, dos tareas de Procesamiento de Lenguaje Natural, la *detección de temas* y la *clasificación de polaridad*, adquieren gran relevancia. La detección de temas conlleva explorar la web en busca de contenidos relacionados con un determinado tema o materia. La clasificación de polaridad, por su parte, significa determinar la orientación polar (i.e., positivo o negativo) de un texto. Estas dos tareas son el objetivo de la competición *TASS-SEPLN*. En el presente trabajo, se describe la participación de la UNED en dicha competición. Para la tarea de detección de temas, se presenta un sistema basado en un modelo probabilístico (Twitter-LDA). Para la clasificación de polaridad, se propone un método basado en significados emocionales. Los resultados experimentales muestran que el sistema desarrollado se comporta adecuadamente.
Palabras clave: social media, detección de temas, clasificación de polaridad

Abstract: Social media, such as blogs, forums, and social networks, offer an excellent place for sharing information and connecting people. The information in these media (usually referred to as *user generated content*) is of great interest for both companies and individuals. However, the huge amount of information that is generated need to be efficiently processed to be of real use. In this context, two Natural Language Processing tasks, *topic detection* and *polarity classification*, become highly relevant. Topic detection involves exploring the web in the search for contents related to a given topic. Polarity classification, in turn, is a sentiment analysis task concerned with the problem of determining the polar orientation (i.e., positive or negative) of a text. These two tasks are the focus of the TASS-SEPLN competition. In this paper, we present the participation of the UNED group in such competition. For topic detection, we present a system based on a probabilistic model (Twitter-LDA). For polarity classification, we propose an emotional concept-based method. The experimental results show the adequacy of our approach for the task.
Keywords: social media, topic detection, polarity classification

1. Introduction

The enormous popularity of "social media", such as blogs, forums, or real time social networking's sites offer a place for sharing information as it happens and for connecting people in real time, often making lasting friendships, contacts and spreading a wealth of latest news about real-world events and topics dominating social discussions. This spread of new social media channels has produce a huge amount of the so called *user generated content*, which has motivated many natural language processing task, such as sentiment analysis, topic detection, product comparison

* This research was partially supported by the Spanish Ministry of Education (FPI grant nr BES-2011-044328), the Spanish Ministry of Science and Innovation (Holopedia Project, TIN2010-21128-C02) and the European Union (FP7-ICT-2011-7 - Language technologies - nr 288024 (LiMoSINe)

or opinion summarization. In this paper we focus on sentiment analysis and topic detection.

With more than 140 million active users and 340 million tweets a day (as of March 2012), Twitter presents the most popular and interesting social media channel from a research perspective. However, due to its characteristics, it is the most noisy and intensive stream of new content, which makes users to face a challenge when they want to find the most interesting themes in few time.

Topic exploration is a laborious and time-consuming task, usually involving several searches. Users are particularly interested in emergent topics that arise from recent events but the representation of the data and the search results of the social media sites do not support this kind of information. Detecting and characterizing emerging topics of discussion through analysis of Internet data is of great interest to particular users and for businesses. For example, a market analyst may want to review technical and news-related literature for recent trends that will impact the companies he is watching and reporting on. The manual review of all the available data is simply not feasible. Human experts who are tasked with identifying emerging events need to rely on automated systems as the amount of information available in digital format increases.

On the other hand, sentiment analysis is concerned with the problem of discovering emotional meanings in text. This discipline has gained much attention from the research community in recent years, mainly due to its many practical applications and the increasing availability of user generated content.

One of the most popular sentiment analysis tasks is polarity classification, which attempts to classify texts according to the positivity or negativity of the opinions expressed in them (Pang, Lee, y Vaithyanathan, 2002; Turney, 2002; Esuli y Sebastiani, 2006; Wilson, Wiebe, y Hoffmann, 2009; Wiegand et al., 2010). Determining polarity might seem an easy task, as many words have some polarity by themselves. However, words do not always express the same emotions, and in most cases the polarity of a word depends on the context in which the word is used. So, terms that clearly denote negative feelings can be neutral, or even positive, depending on their context. The degree or strength of polarity is also an interesting point to consider. For sure, any opinionated sentence can be classified into positive or negative, but it is clear that not all sentences express the same negative or positive intensity. So, sentiment analysis systems should include semantic-level analysis in order to solve word ambiguity and correctly capture the meaning of each word according to its context. Also, complex linguistic processing is needed to deal with problems such as the effect of negations and intensifiers. Moreover, understanding the emotional meaning of the different textual units is important to accurately determine the overall polarity of a text and its degree.

In this paper, we present a combined system that has as main objectives analyzing the sentiments of tweets written in Spanish, and grouping them into a set of given topics. To accomplish the first objective we have adapted an existing emotional concept-based system for sentiment analysis to classify tweets in Spanish. The original method makes use of an affective lexicon to represent the text as the set of emotional meanings it expresses, along with advanced syntactic techniques to identify negations and intensifiers, their scope and their effect on the emotions affected by them. Besides, the method addresses the problem of word ambiguity, taking into account the contextual meaning of terms by using a word sense disambiguation algorithm. For the second objective, detection of topics, we first build for each topic of the task a lexicon of words that best describe it, thus representing each topic as a ranking of discriminative words. Moreover, a set of events is retrieved based on a probabilistic approach that was adapted to the characteristics of Twitter. To determine which of the topics corresponds to each event, the topic with the highest statistical correlation was obtained comparing the ranking of words of each topic and the ranking of words most likely to belong to the event.

The paper is organized as follows. Section 2 introduces previous work in topic detection and sentiment analysis. Section 3 presents the combined system. Section 4 describes and discusses the results obtained by the system in the TASS-SEPLN challenge. Finally, section 5 provides concluding remarks and outlines future work.

2. Related work

The topic of a tweet is a latent feature and can be inferred by analyzing its content. Modeling Twitter content requires methods that are suitable for short texts with heterogeneous vocabulary. Recent work shows that one such method is Latent Dirichelet Allocation (LDA) (Blei, Ng, y Jordan, 2003) and its extensions (Weng et al., 2010). A direct application is to use the traditional LDA model to discover topics from tweets by treating each tweet as a single document, but it is probable that this method does not work well taking in to account that tweets are very short (often containing only a single sentence).

To overcome this difficulty, some previous studies proposed to consider all the tweets of a user as a single document (Weng et al., 2010; Hong y Davison, 2010). This treatment can be regarded as an application of the author-topic model (Steyvers et al., 2004) to tweets, where each document (tweet) has a single author. However, the aggregated tweets of a single user may have a diverse range of topics, so this model does not exploit the following important property of the tweets: a single tweet is usually about a single topic. We apply a modified author-topic model called Twitter-LDA introduced by (Zhao et al., 2011), which assumes a single topic assignment for an entire tweet.

Concerning polarity classification, this task is usually formulated as a supervised ML problem with two classes (i.e. positive and negative), but sometimes considers a more fine-grained classification (e.g. strongly-negative, negative, neutral, positive and strongly-positive). Traditional approaches consider the text as a bag of word frequencies, n-grams, or even more complex lexical features, such as phrases and information extraction patterns (Pang, Lee, y Vaithyanathan, 2002; Dave, Lawrence, y Pennock, 2003; Riloff, Patwardhan, y Wiebe, 2006). Approaches based on word frequencies have the main drawback of being highly dependent on the application domain.

An alternative to word-based learning is sentiment-based learning. That is, instead of representing the text as a bag of words, the text is modeled as a set of polar expressions (Das y Chen, 2001; Wilson, Wiebe, y Hoffmann, 2009). *Polar expressions* are words that contain a prior polarity. For example, *like* or *good* are positive polar expressions, while *hate* or *bad* are negative ones. However, the approaches above work with words instead of senses, disregarding the contextual meaning of such words, and the fact that a word may present various senses of which some of them could have different polarities. On the other hand, even though the use of polarity-based lexicons is quite frequent, few works employ more fine-grained emotional resources. There seems to be an assumption that emotional classification exclusively depends on the polar orientation of the words or concepts within the text, regardless of the sentiments or emotions they express. However, it is clear, for instance, that the words *cancer* and *cold*, though having a negative orientation, express different emotions: a *cancer* is usually associated with *fear* and *sadness*, while a *cold* is better associated with *displeasure* or *dislike.*

Finally, in spite of their importance for sentiment analysis tasks, linguistic modifiers such as negation or intensifiers have attracted less attention and are usually addressed in very naive fashion. For example, negation is mostly considered a simple polarity shifter (Das y Chen, 2001), while intensifiers are all considered as amplifiers or diminishers that contain a fixed value for all positive words and another value for all negative words respectively (Polanyi y Zaenen, 2006).

3. UNED system

In this section we present the system for topic detection and polarity classification.

For the task of topic detection, our system has three stages. In the first one (Section 3.1.1), the system uses a corpus of tweets labeled with topics to obtain a ranking of important words for every topic. Note that this stage can be done off-line. The second stage (Section 3.1.2) consists in, given a set of tweets, obtaining clusters of tweets that discuss the same event, for example: reviews on a novel, recommendations of a book, comments about an author. And finally, in the third stage (Section 3.1.3) we identify, for each event, to which of the 10 topics it belongs to.

In polarity classification, our main concern is to analyze the applicability of a complex emotional concept-based approach intended for classifying product reviews, to classify tweets in Spanish. To this aim, we have adapted the approach presented in (Carrillo de Albornoz, Plaza, y Gervás, 2010) for polarity and intensity classification of opinionated texts. The main idea of this method is to extract the WordNet concepts in a sentence that entail an emotional meaning, assign them an emotion within a set of categories from an affective lexicon, and use this information as the input to a machine learning algorithm. The strengths of this approach, in contrast to other more simple strategies, are: (1) the use of WordNet and a word sense disambiguation algorithm, which allows the system to work with concepts rather than terms, (2) the use of emotions instead of terms as classification attributes, and (3) the processing of negations and intensifiers to invert, increase or decrease the intensity of these emotions. This system has been shown to outperform previous systems which aim to solve the same task.

3.1. Topic detection

3.1.1. Topic representation

Intuitively, the words that best describe a topic are the words that occur relatively more frequently in the tweets that are labeled with this topic than in the tweets labeled with a different topic. Based on this intuition, we obtain a weighted vector of important words for a topic using the Kullback-Leibler Divergence (KLD). The KLD measures the relative entropy between two probability distributions. We calculate for every topic t the KLD scores of it's lexical units as:

$$KLD_t(w) = P_R(t) \times \log \frac{P_R(w)}{P_N(w)}$$

Where: $P_R(w)$ - probability of the lexical unit w occurring in the relevant documents (tweets labeled with the topic t) , and calculated as $f_R(w)/R$, where $f_R(w)$ - frequency of occurrence of w in the relevant set, R - number of terms in the relevant set; $P_N(w)$ - probability of the lexical unit w occurring in the non-relevant documents (tweets labeled with a different topic), and calculated as $f_N(w)/N$, where $f_N(w)$ - frequency of occurrence of w in the non-relevant set, N - number of terms in the non-relevant set. In this way, we have for each topic the ranking of words that best describes it.

3.1.2. Event detection

In order to obtain the events we use an approach based on a latent variable topic model, namely Latent Dirichelet Allocation (LDA) (Blei, Ng, y Jordan, 2003). It is an unsupervised machine learning technique which uncovers information about latent topics across a corpora. We use a variant of LDA proposed by Zhao et al. (Zhao et al., 2011) that is adapted to the characteristics of Twitter: tweets are short (140-character limit) and a single tweet tends to be about a single topic.

The model is based on the following assumptions. There is a set of topics T in Twitter, each represented by a word distribution. Each user has her topic interests modeled by a distribution over the topics. When a user wants to write a tweet, first chooses a topic based on her topic distribution. Then the user chooses a bag of words one by one based on the chosen topic. However, not all words in a tweet are closely related to the topic of that tweet; some are background words commonly used in tweets on different topics. Therefore, for each word in a tweet, the user first decides whether it is a background word or a topic word and then chooses the word from its respective word distribution. The process is described as follows:

1. Draw $\phi^{\mathcal{B}} \sim Dir(\beta)$, $\pi \sim Dir(\gamma)$
2. For each topic $t \in T$,
   - **(a)** draw $\phi^t \sim Dir(\beta)$
3. For each user $u \in U$,
   - **(a)** draw $\theta \sim Dir(\alpha)$
   - **(b)** for each tweet $d_{u,m}$
     - **i.** draw $z_{u,m} \sim Multi(\theta)$
     - **ii.** for each word $w_{u,m,n}$
       - **A.** draw $y_{u,m,n} \sim Bernoulli(\pi)$
       - **B.** draw $w_{u,m,n} \sim Multi(\phi^{\mathcal{B}})$ if $y_{u,m,n} = 0$ and $w_{u,m,n} \sim Multi(\phi^{z_{u,m}})$ if $y_{u,m,n} = 1$

where: ϕ^t denotes the word distribution for topic t; $t^{\mathcal{B}}$ the word distribution for background words; θ^u denotes the topic distribution of user u and π denotes a Bernoulli distribution that governs the choice between background words and topic words. After applying the TwitterLDA model, a topic is represented as a vector of probabilities over the space of words.

3.1.3. Trending topics

Lastly, we need to obtain a mapping between the topics of the task and the events retrieved by the TwitterLDA method. Noting that, we have two rankings: the ranking of words that best describes each topics obtained from the training data, and for each event of TLDA, the probability that the words belong to that event, this may be constructed as the importance of that word on the event, i.e. a ranking of words of the events. Therefore, using a measure of correlation of rankings for each event, we can obtain the topic to which it relates.

A rank correlation is the relationship between different rankings of the same set of items (ranking of words). A rank correlation coefficient measures the degree of similarity between two rankings. We use one of the most popular rank correlation statistic: the *Kendall* rank correlation coefficient, commonly referred to as *Kendall's tau* (τ) coefficient. Given two rankings on the same domain (on the same set of objects), Kendall's rank correlation coefficient τ is defined as:

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)}$$

where n_c is the number of concordant pairs and n_d is the number of discordant pairs. A concordant (discordant) pair is an ordered pair of objects, which has the same (opposite) order in both rankings. Kendall's τ is normalized in the interval $\langle -1, 1 \rangle$. In the case of maximum similarity between two rankings $\tau = 1$ (rankings are identical). In the case of maximum dissimilarity $\tau = -1$ (one ranking is reverse of the other).

Thus, for each event retrieved by TwitterLDA, we calculate the correlation with each of the topics and choose the topic that has greater value.

3.2. Sentiment Analysis

The original method presented in (Carrillo de Albornoz, Plaza, y Gervás, 2010) has been modified to improve the scope detection approach for negation and intensifiers in order to deal with the effect of subordinate sentences and special punctuation marks. Also, the presented approach uses the SentiSense affective lexicon (Carrillo de Albornoz, Plaza, y Gervás, 2012), which is a lexicon specifically designed for opinionated texts. Sentisense attaches an emotional category from a set of 14 emotions to WordNet concepts. SentiSense also include the antonym relationship between emotional categories, which allows us to capture the effect of some linguistic modifiers such as negation. We have adapted the system to work with Spanish texts, as the original system is conceived for English. The method comprises four steps that are described below:

3.2.1. Pre-processing: POS Tagging and Concept Identification

The first step aims to translate each text to its conceptual representation in order to work at the concept level in the next steps and avoid word ambiguity. To this end, the input text is split into sentences and the tokens are tagged with their POS using the Freeling library (Carreras et al., 2004). In this step, the syntax tree of each sentence is also retrieved using the Freeling chunk parser. With this information, the system next maps each token to its appropriate WordNet concept using the UKB algorithm (Agirre y Soroa, 2009) as included in the Freeling library. Besides, to enrich the emotion identification step, the hypernyms of each concept are retrieved from WordNet.

3.2.2. Emotion Identification

Once the concepts are identified, the next step maps each WordNet synset to its corresponding emotional category in the SentiSense affective lexicon, if any. The emotional categories of the hypernyms are also retrieved. We hypothesize that the hypernyms of a concept entail the same emotions than the concept itself, but decreasing the intensity of the emotion as we move up in the hierarchy. So, when no entry is found in the SentiSense lexicon for a given concept, the system retrieves the emotional category associated to its nearest hypernym, if any. However, only a certain level of hypernymy is accepted, since an excessive generalization introduces some noise in the emotion identification. This parameter has been empirically set to 3. In order to accomplish this step for Spanish texts we have automatically translated the SentiSense lexicon to the Spanish language. To do this, we have automatically updated the synsets in SentiSense to their WordNet 3.0 version using the WordNet mappings. In particular, for nouns and verbs we use the mappings provided by the WordNet team[1] and for adjectives and adverbs, the UPC mappings[2]. In this automatic process we have only found 15 labeled synsets without a direct mapping, which were removed in the new SentiSense version. Finally, in order to translate the SentiSense English version to Spanish we use the Multilingual Central Repository (MRC) (Gonzalez-Agirre, Laparra, y Rigau, 2012). The MCR is an open source database that integrates WordNet versions for five different languages: English, Spanish, Catalan, Basque and Galician. The Inter-Lingual-Index (ILI) allows the automatic translation of synsets from one language to another.

3.2.3. Post-processing: Negation and Intensifiers

In this step, the system has to detect and solve the effect of negations and intensifiers over the emotions discovered in the previous step. This process is important, since these linguistic modifiers can change the polarity and intensity of the emotional meaning of the text. Clearly, the text *Recio no tiene indicios potentes para denunciar a los responsables de los ERE* entails different polarity than the text *Recio tiene indicios potentes para denunciar a los responsables de los ERE*, and sentiment analysis systems must be aware of this fact.

To this end, our system first identifies the presence of modifiers using a list of common negation and intensification tokens. In such a list, each intensifier is assigned a value that represents its weight or strength. The scope of each modifier is determined using the syntax tree of the sentence in which the modifier arises. We assume as scope all descendant leaf nodes of the common ancestor between the modifier and the word immediately after it, and to the right of the modifier. However, this process may introduce errors in special cases, such as subordinate sentences or those containing punctuation marks. In order to avoid this, our method includes a set of rules to delimit the scope in such cases. These rules are based on specific tokens that usually mark the beginning of a different clause (e.g., *porque*, *hasta*, *por qué*, *aunque*, etc.). Since some of these delimiters are ambiguous, their POS is used to disambiguate them. Once the modifiers and their scope are identified, the system solves their effect over the emotions that they affect in the text. The effect of negation is addressed by substituting the emotions assigned to the concepts by their antonyms. In the case of the intensifiers, the concepts that fall into the scope of an intensifier are tagged with the corresponding percentage weight in order to increase or diminish the intensity of the emotions assigned to the concepts.

In order to adapt the present method to Spanish texts, a list of common negation tokens in Spanish (such as *no*, *nunca*, *nada*, *nadie*, etc.) and common intensifiers (*más*, *menos*, *bastante*, *un poco*, etc.) were developed (based on the original list of negation and intensifier signals from (Carrillo de Albornoz, Plaza, y Gervás, 2010)). In order to determine the scope of each modifier, the syntax tree as generated by the FreeLing library is used.

3.2.4. Classification

In the last step, all the information generated in the previous steps is used to translate each text into a Vector of Emotional Intensities (VEI), which will be the input to a machine learning algorithm. The VEI is a vector of 14 positions, each of them representing one of the emotional categories of the SentiSense affective lexicon. The values of the vector are generated as follows:

- For each concept, C_i, labeled with an emotional category, E_j, the weight of the concept for that emotional category, $weight(C_i; E_j)$, is set to 1.0.
- If no emotional category was found for the concept, and it was assigned the category of its first labeled hypernym, $hyper_i$, then the weight of the concept is computed as:

$$weight(C_i; E_j) = 1/(depth(hyper_i) + 1)$$

- If the concept is affected by a negation and the antonym emotional category, $E_a nton_j$, was used to label the concept, then the weight of the concept is multiplied by $\alpha = 0{,}6$. This value has been empirically determined in previous studies. It is worth mentioning that the experiments have shown that α values

[1]WordNet mappings. http://wordnet.princeton .edu/wordnet/download/.

[2]Universidad Politécnica de Cataluña mappings. http://nlp.lsi.upc.edu/web/index.php?option= com_content&task=view&id=21 &Itemid=57.

below 0.5 decrease performance sharply, while it drops gradually for values above 0.6.

- If the concept is affected by an intensifier, then the weight of the concept is increased/decreased by the intensifier percentage, as shown in:

$$weight(C_i; E_j) = weight(C_i; E_j) * (100 + \%)$$

- Finally, the position in the VEI of the emotional category assigned to the concept is incremented by the weight previously calculated.

4. Evaluation and discussion

This section presents the evaluation of our system in the context of the TASS-SEPLN competition.[3] The data set consists of tweets, written in Spanish by nearly 200 well-known personalities and celebrities of the world. The set is divided into two sets: training (7,210 tweets) and test (60,798 tweets). Each tweet is tagged with its global polarity, indicating whether the text expresses a positive, negative or neutral sentiment, or no sentiment at all. 5 levels have been defined: strong positive (P+), positive (P), neutral (NEU), negative (N), strong negative (N+) and one additional no sentiment tag (NONE). Each tweet of the corpus has been semiautomatically assigned to one or several of 10 possible topics: *sports*, *music*, *literature*, *soccer*, *politics*, *economy*, *art*, *entertainment*, *music* and *technology*.

4.1. Topic detection

The aim of this tasks is to automatically identify the topic of each tweet. We run Twitter-LDA with 500 iterations of Gibbs sampling. After trying a few different numbers of topics, we empirically set the number of topics to 100. We set α to 50,0/$|T|$, β to a smaller value of 0.01 and λ to 20 as (Zhao et al., 2011) suggested. Table 1 shows the results obtained by our system. Due to space constraints we only show the result reached by our system and the system that was in first place. It may be seen that the result of our system is not satisfactory. We believe that this behavior is due to the vocabulary that was obtained for each topic in the stage of representation (see Section 3.1.1). For example, the topic *literature* has only 99 tweets and most are references or comments about any book or novel. Most of the time the topic can be deduced from a single word of the tweet: novel, book, author readings, literary. Therefore, when building the vocabulary there are few words that really belong to this topic.

An important improvement to enrich the vocabulary could be to add words from an external resource (e.g., WordNet Domains (Magnini y Cavaglia, 2000)), and name of entities related to the domain (could be extracted from Wikipedia).

System	**Precision**	**Rank**
L2F - INESC	65,37 %	1
UNED	30,98 %	15

Table 1: Trending topic coverage

4.2. Polarity classification

In the TASS-SEPLN competition, polarity classification is evaluated as two different tasks. The first task consists in automatically classifying each tweet in one of the 5 polarity levels previously mentioned. However, prior to this classification, the task requires to filter those tweets which do not express any sentiment (i.e., those tagged as NONE). To perform this filtering, our system simply considers an extra class, NONE, so that it classifies the tweets into six classes. Next, the tweets classified as NONE are ignored for evaluation purposes.

The results of the two variants of the UNED system for this task are shown in Table 2. The first uses the *logistic* model in Weka as the ML algorithm, while the second uses the J48 algorithm. As it may be observed, accuracy for both algorithms is over 52 %, which is a very high accuracy considering the complexity of the task and the high number of polarity classes that are taken into account. Therefore, our results in this task are quite satisfactory, as evidenced by the fact that our runs are ranked 7th and 8th in the competition (of 20 systems)[4]. Besides, the results of the two runs are quite similar, regard-

[3] The task guidelines may be found in http://www.daedalus.es/TASS/tasks.php

[4] The competition ranking may be found in http://www.daedalus.es/TASS/participation.php

System	Accuracy	Rank
Elhuyar F.	65.29	1
UNED-Logistic	53,82 %	7
UNED-J48-Graft	52,54 %	8

Table 2: 5-classes polarity detection

System	Accuracy	Rank
Elhuyar F.	71.12	1
UNED-Logistic	59,03 %	7
UNED-J48-Graft	58,77 %	8

Table 3: 3-classes polarity detection

less of the ML algorithm that is used, which seems to indicate that our emotional-based representation is correctly capturing the polarity of the text.

The second task consists in classifying each tweet in 3 polarity classes. To this end, only the tweets tagged as positive, neutral and negative are considered. The results of the two UNED runs for this task are shown in Table 3. As expected, the results are better than those obtained in the previous task, since the number of polarity classes is lower, and thus the task is simpler. Again, the two variant of our system are ranked 7th and 8th among the 20 participants. However, it is worth mentioning that, even if we consider these results to be quite positive, the original system have presented significantly better accuracy when evaluated over other data sets (in particular, sets of different product reviews). This is due to two main facts: first, the systems is expected to work better when classifying product reviews than more general texts (as the tweets in hands), since product reviews express the user's satisfaction or dissatisfaction with the different product attributes, and therefore employ a highly emotive language. Second, the coverage of the affective lexicon, SentiSense, for the evaluation data sets is quite poor (only around 12 % of the words are labeled), and therefore we find that an important number of tweets are not labeled with any emotion. This was expected, since SentiSense is specially designed for processing product reviews. Therefore, taking this low coverage into account, we expect that expanding the coverage of SentiSense will allow us to significantly improve the classification results.

5. *Conclusions*

This paper presents the contribution of the UNED group to the tasks of sentiment analysis and trending topics at the TASS whorkshop. The results have shown that the method for determining the polarity of the tweets performs reasonably well taking into account that the system was originally conceived for English texts. This may have influenced the results because it was necessary to make an automatic translation of the SentiSense affective lexicon and also to use the Spanish version of Wordnet, which has considerably less coverage than the English version. However, the topic detection obtained quite poor results and we believe that what should be improved is the representation of the topics. Thus, as future work we plan to use other sources and/or resources in order to retrieve better discriminative words for each topic.

Bibliografía

Agirre, E. y A. Soroa. 2009. Personalizing pagerank for word sense disambiguation. En *Proceedings of the 12th conference of the European chapter of the Association for Computational Linguistics.*

Blei, David M., Andrew Y. Ng, y Michael I. Jordan. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, Marzo.

Carreras, X., I. Chao, LL. Padró, y M. Padró. 2004. Freeling: An open-source suite of language analyzers. En *Proceedings of the 4th International Conference on Language Resources and Evaluation.*

Carrillo de Albornoz, J., L. Plaza, y P. Gervás. 2010. A hybrid approach to emotional sentence polarity and intensity classification. En *Proceedings of the 14th Conference on Computational Natural Language Learning*, páginas 153–161.

Carrillo de Albornoz, J., L. Plaza, y P. Gervás. 2012. Sentisense: An easily scalable concept-based affective lexicon for sentiment analysis. En *proceedings of the 8th International Conference on Language Resources and Evaluation.*

Das, Sanjiv y Mike Chen. 2001. Yahoo! for Amazon: Extracting Market Sentiment from Stock Message Boards. En *Proceedings of the Asia Pacific Finance Association Annual Conference (APFA).*

Dave, Kushal, Steve Lawrence, y David M. Pennock. 2003. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. En *Proceedings of the 12th international conference on World Wide Web*, WWW '03, páginas 519–528, New York, NY, USA. ACM.

Esuli, Andrea y Fabrizio Sebastiani. 2006. Determining term subjectivity and term orientation for opinion mining. En *Proceedings EACL-06, the 11th Conference of the European Chapter of the Association for Computational Linguistics*, Trento, Italy.

Gonzalez-Agirre, A., E. Laparra, y G. Rigau. 2012. Multilingual central repository version 3.0: upgrading a very large lexical knowledge base. En *Proceedings of the Sixth International Global WordNet Conference.*

Hong, Liangjie y Brian D. Davison. 2010. Empirical study of topic modeling in twitter. En *Proceedings of the First Workshop on Social Media Analytics*, SOMA '10, páginas 80–88, New York, NY, USA. ACM.

Magnini, Bernardo y Gabriela Cavaglia. 2000. Integrating subject field codes into WordNet. En *Proceedings of LREC 2000.*

Pang, Bo, Lillian Lee, y Shivakumar Vaithyanathan. 2002. Thumbs up? sentiment classification using machine learning techniques. En *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, páginas 79–86, Philadelphia, USA.

Polanyi, Livia y Annie Zaenen. 2006. Contextual valence shifters. En W. Bruce Croft James Shanahan Yan Qu, y Janyce Wiebe, editores, *Computing Attitude and Affect in Text: Theory and Applications*, volumen 20 de *The Information Retrieval Series*. Springer Netherlands, páginas 1–10.

Riloff, Ellen, Siddharth Patwardhan, y Janyce Wiebe. 2006. Feature subsumption for opinion analysis. En *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, EMNLP '06, páginas 440–448, Stroudsburg, PA, USA. Association for Computational Linguistics.

Steyvers, Mark, Padhraic Smyth, Michal Rosen-Zvi, y Thomas Griffiths. 2004. Probabilistic author-topic models for information discovery. En *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '04, páginas 306–315, New York, NY, USA. ACM.

Turney, Peter. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. En *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL 2002)*, páginas 417–424, Philadelphia, USA.

Weng, Jianshu, Ee P. Lim, Jing Jiang, y Qi He. 2010. TwitterRank: finding topic-sensitive influential twitterers. En *Proceedings of the third ACM international conference on Web search and data mining*, WSDM '10, páginas 261–270, New York, NY, USA. ACM.

Wiegand, Michael, Alexandra Balahur, Benjamin Roth, Dietrich Klakow, y Andrés Montoyo. 2010. A survey on the role of negation in sentiment analysis. En *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, NeSp-NLP '10, páginas 60–68, Stroudsburg, PA.

Wilson, Theresa, Janyce Wiebe, y Paul Hoffmann. 2009. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational Linguistics*, 35(3):399–433.

Zhao, Wayne Xin, Jing Jiang, Jianshu Weng, Jing He, Ee-Peng Lim, Hongfei Yan, y Xiaoming Li. 2011. Comparing twitter and traditional media using topic models. En *Proceedings of the 33rd European conference on Advances in information retrieval*, ECIR'11, páginas 338–349, Berlin, Heidelberg. Springer-Verlag.