TASS 2015, septiembre 2015, pp 23-28 recibido 20-07-15 revisado 24-07-15 aceptado 29-07-15

DeustoTech Internet at TASS 2015: Sentiment analysis and polarity classification in spanish tweets

DeustoTech Internet en TASS 2015: Análisis de sentimientos y clasificación de polaridad en tweets en español

Juan Sixto Cesteros
DeustoTech–Deusto
Institute of Technology
Universidad de Deusto
48007 Bilbao (Spain)
jsixto@deusto.es

Aitor Almeida
DeustoTech–Deusto
Institute of Technology
Universidad de Deusto
48007 Bilbao (Spain)
aitor.almeida@deusto.es

Diego López de Ipiña
DeustoTech–Deusto
Institute of Technology
Universidad de Deusto
48007 Bilbao (Spain)
dipina@deusto.es

Resumen: Este artículo describe nuestro sistema presentado en el taller de análisis de sentimiento TASS 2015. Nuestro sistema aborda la tarea 1 del workshop, que consiste en realizar un análisis automático de sentimiento para determinar la polaridad global de un conjunto de tweets en español. Para ello, nuestro sistema se basa en un modelo supervisado con máquinas de soporte vectorial lineales en combinación con varios léxicos de polaridad. Se estudia la influencia de las diferentes características linguísticas y de diferentes tamaños de n-gramas en la mejora del algoritmo. Así mismo se presentan los resultados obtenidos, las diferentes pruebas que se han realizado, y una discusión sobre los resultados.
Palabras clave: Análisis de sentimientos, clasificación de la polaridad, Twitter

Abstract: This article describes our system presented at the workshop for sentiment analysis TASS 2015. Our system approaches the task 1 of the workshop, which consists on performing an automatic sentiment analysis to determine the global polarity of a set of tweets in Spanish. To do this, our system is based on a model supervised Linear Support Vector Machines combined with some polarity lexicons. The influence of the different linguistic features and the different sizes of n-grams in improving algorithm performance. Also the results obtained, the various tests that have been conducted, and a discussion of the results are presented.
Keywords: Sentiment Analysis, Polarity Classification, Twitter

1 *Introduction*

Since the origin of Web 2.0, Internet contains a very large amounts of user-generated information on an unlimited number of topics. Many entities such as corporations or political groups try to learn about that knowledge to know the opinion of users. Social Media platforms such as Facebook or Twitter have proven to be useful for this tasks,due to the very high volume of messages that these platforms generate in real time and the very large number of users that use them everyday.

Faced with this challenge, in the last years the number of the Sentiment Analysis researches has increased appreciably, especially those based in Twitter and microblogging. It should be taken into account that the performance of these researches is language-dependent, reflecting the considerable differences between languages and the difficulty of establish standard linguistic rules (Han, Cook, and Baldwin, 2013).

In this context, the TASS[1] workshop (Villena-Román et al., 2015) is an evaluation workshop for sentiment analysis focused on Spanish language, organized as a satellite event of the annual conference of the Spanish Society for Natural Language Processing (SEPLN)[2]. This paper is focused on the first task of the workshop consist on determining the global polarity of twitter messages.

This paper presents a global polarity clas-

[1]Taller de Análisis de Sentimientos en la SEPLN
[2]http://www.sepln.org/

Publicado en http://ceur-ws.org/Vol-1397/. CEUR-WS.org es una publicación en serie con ISSN reconocido ISSN 1613-0073

sification in Spanish tweets based on polarity lexicons and linguistic features. It is adapted to Spanish tweet texts, which involve particular linguistic characteristics like short length, limited to 140 characters, slang, spelling and grammatical errors and other user mentions.

The rest of the paper is organized as follows: the sentiment analysis related works are described in Section 2, the developed system's description is presented in Section 3, evaluation and results in Section 4 and conclusion and future work are discussed in Section 5.

2 Related work

There exists a large amount of literature addressing the sentiment analysis field, especially applied to Twitter and microblogging context. General surveys about Opinion Mining and Sentiment Analysis may be found (Pang and Lee, 2008), (Martinez-Camara et al., 2014), although due to the enormous diversity of applications on this field, different approaches to solve problems in numerous scopes have been generated, like user classification (Pennacchiotti and Popescu, 2011), Spam detection in social media (Gao et al., 2010), classification of product reviews (Dave, Lawrence, and Pennock, 2003), demographic studies (Mislove et al., 2011), political sentiment and election results prediction (Bermingham and Smeaton, 2011) and even clinical depression prediction via Twitter (De Choudhury et al., 2013).

Twitter has certain specific characteristics which distinguish them from other social networks, e.g. short texts, @user mentions, #hashtags and retweets. All of these characteristics have been extensively studied (Pak and Paroubek, 2010), (Han and Baldwin, 2011). Some of them have been resolved through the text normalization approach (Ruiz, Cuadros, and Etchegoyhen, 2013) while others have been used as key elements in classification approach (Wang et al., 2011). Indeed, several researches prove that the in-depth knowledge of these characteristics will significantly improve the social media based applications (Jungherr, 2013), (Wang et al., 2013).

For several years we assist to an exponential increase of studies based on sentiment analysis and opinion mining in Twitter. According to the state of art, two main approaches exist in sentiment analysis: supervised learning and unsupervised learning. Supervised systems implement classification models based on classification algorithms, being the most frequent the Support Vector Machine (SVM) (Go, Bhayani, and Huang, 2009), Logistic Regression (LR) (Thelwall, Buckley, and Paltoglou, 2012), Conditional Random Fields (CRF) (Jakob and Gurevych, 2010) and K Nearest Neighbors (KNN) (Davidov, Tsur, and Rappoport, 2010). Unsupervised systems are based on the use of lexicons to calculate the semantic orientation(Turney, 2002) and present a new perspective for classification tasks, most effective in cross-domain and multilingual applications.

During the last TASS workshop in 2014 (Villena-Román et al., 2015), LyS presented a supervised liblinear classifier with several lexicons of Spanish language, whose results are among the best in task 1 (Sentiment Analysis at the tweet level) (Vilares et al., 2014). Further, (San Vicente and Saralegi, 2014) presented a Support Vector Machine (SVM) based on a classifier that merges polarity lexicons with several linguistic features as punctuation marks or negation signs. Finally, the best results in task 1 correspond to (Hurtado and Pla, 2014), who present a Linear-SVM based classifier that addresses the task using a one-vs-all strategy in conjuction with a vectorized list of tf-idf coefficients as text representation.

3 System description

Several tools and datasets have been used during the experiments to develop our final system. Because our system only approaches the Task 1: Sentiment Analysis at global level, this consists in a unique pipeline that reaches the process completely. At the beginning, a naive normalization system is applied to the tweet texts with the purpose to standardize several Twitter own features, like #Hashtags or @User mentions. Then, the Freeling language analysis tool[3] (Padró and Stanilovsky, 2012) is used to tokenize, lemmatize and annotate the texts with part-of-speech tags (pos-tagging).

During this step, based on a list of stop words for Spanish language, this words are annotated to be ignored by polarity ranking steps.

[3] http://nlp.lsi.upc.edu/freeling/

The task has been addressed as an automatic multi-class classification job. For this reason, it has been considered appropriate to focus this problem with a one-vs-all strategy, in a similar way to the presented by (Hurtado and Pla, 2014) in TASS 2014. These binary classifiers have been developed using two different approaches, LinearSVC Machines and Support Vector Regression (SVR) Machines. the comparison of machine-learning based results is shown in Results section.

To represent the text's as vectorized features, two main sources have been used: the polarity lexicon punctuations and the Okapi BM25 ranking function, to represent document's scoring (Robertson et al., 1995). BM25 is a bag-of-words retrieval function that ranks a set of documents based on the query terms appearing in each document. The formula used to implement BM25 in the system is defined below:

$$score(D, Q) = \sum_{i=1}^{n} IDF(q_i) \cdot TF(q_i) \quad (1)$$

$$TF(q_i) = \frac{f(q_i, D) \cdot (k_1 + 1)}{f(q_i, D) + k_1 \cdot (1 - b + b \cdot \frac{|D|}{avgdl})} \quad (2)$$

$$IDF(q_i) = \log \frac{N - n(q_i) + 0{,}5}{n(q_i) + 0{,}5} \quad (3)$$

To calculate the score of a document D, $f(q_i, D)$ is the frecuency of each word lemma (q_i), $|D|$ is the length of the text D in words and $avgdl$ is the average text length. After several experiments over the training corpus, the free parameters k_1 and b have been optimized to $k_1 = 76$ and $b = 0{,}75$. System develops one BM25 dictionary for each one-vs-all classifier.

In conjunction with the document's score, each tweet has been represented using different polarity lexicons in order to classify them into the six (P+, P, NEU, N, N+ and NONE) and the four (P, N, NEU and NONE) polarities. We use several datasets to score the polarity levels of words and lemmas. Owing to different characteristics of each dataset, such as semantic-orientation values, scores are calculated separately and considered as independent attributes in the system.

- **LYSA Twitter lexicon v0.1**. LYSA is an automatically-built polarity lexicon for Spanish language that was created by downloading messages from Twitter, and includes both negative and positive Spanish words (Vilares et al., 2014). The lexicon entries includes a semantic-orientation values ranged from -5 to 5, making it a good resource for multiple sentiment levels identification.
- **ElhPolar dictionary v1.0**. The ElhPolar polarity lexicon for Spanish was created from different sources, and includes both negative and positive words (Saralegi and San Vicente, 2013).
- **The Spanish Opinion Lexicon (SOL)**. The Spanish Opinion Lexicon (SOL) is composed by 1,396 positive and 3,151 negative words, thus in total SOL has 4,547 opinion words[4] (Martínez-Cámara et al., 2013). The lexicon has been elaborated from the Bing Liu's word list using Reverso as translator (M. and L., 2004).
- **Negation Words List**. A list of negation spanish words has been created during the experiments. This list is used as a text feature in order to detect negative sentences and possible polarity inversions.

We also consider other text characteristics as classifier features, like text length in words quantity or a list of sentiments represented by emoticons using the Wikipedia's list of emoticons[5]. To conclude the system's prediction, another automatic classifier has been implemented, trained with the predictions of the binary results to select one label.

4 Results

Our results are relative to the Task 1: Sentiment Analysis at global level of TASS 2015. This task consists on performing an automatic sentiment analysis to determine the global polarity of each message in the provided corpus. There are two different evaluations: one based on 6 different polarity labels (P+, P, NEU, N, N+, NONE) and another based on just 4 labels (P, N, NEU, NONE). Also there are two test sets: complete set and 1k

[4] http://sinai.ujaen.es/sol/

[5] https://en.wikipedia.org/wiki/List_of_emoticons

set, a subset of the first one containing only 1000 tweets with a similar distribution to the training corpus was extracted to be used for an alternate evaluation of the performance of systems.

Tables 1 and 2 show the performance of different tested models using the full and 1k sets. For the rating of the developed system, 3 different systems have been presented for each subtask. Our submitted models consist in different features as follows:

- **Run 1:** Words and lemmas based polarity dictionaries as features, differing between positive and negative scores and between different datasets. Okapi BM25 scores of mono-grams used as features with the lemmas of the tweet texts. Binary classifiers were implemented using LinearSVC Machines and the global classifier uses their predictions (True or False).
- **Run 2:** Words and lemmas based polarity dictionaries as features, differing between positive and negative scores and between different datasets. Okapi BM25 scores of mono-grams and bi-grams used as features with the lemmas of the tweet texts. Binary classifiers were implemented using LinearSVC Machines and the global classifier uses their predictions (True or False).
- **Run 3:** Similar to **Run 2**, with the exception of the binary classifiers that were implemented using Support Vector Regression (SVR) Machines and the global classifier uses their predictions (0 to 1 float values).

	Run	Accuracy
6 Labels	Run1	**0.560**
	Run2	0.557
	Run3	0.545
4 Labels	Run1	0.608
	Run2	**0.625**
	Run3	0.490

Table 1: Accuracy on the 5 levels and 3 levels of different approaches using the General Corpus.

The systems based on SVM present the best accuracy levels, with an appreciably higher performance in all tests than the system based in SVR. This suggests that the precision of the regression values, in contrast with the binary values of the SVM classifiers, has a negative impact on the global classifier. However, the use of mono-grams and bi-grams as features presents different success rates depending of the test. This part of the system must be analysed in-depth in order to comprehend the performance difference between both systems.

	Run	Accuracy
6 Labels (1k)	Run1	0.407
	Run2	**0.408**
	Run3	0.396
4 Labels (1k)	Run1	**0.601**
	Run2	0.583
	Run3	0.571

Table 2: Accuracy on the 5 levels and 3 levels of different approaches using the 1k Test Corpus.

5 *Conclusions and Future work*

This paper describes the participation of the DeustoTech Internet research group in the Task 1: Sentiment Analysis at global level at TASS 2015. In our first participation, our team presents a system based in Support Vector Machines in conjunction with several well established polarity lexicons. Experimental results present a good baseline to continue working through the development of new models and developing an structure able to take full advantage of multiple supervised learning systems.

As future work, we propose to research on different approaches to aboard the measure of sentiment analysis problems, especially those related to sentiment degrees with the aim to detect clearly differences between different sentiment levels (Good vs Very Good, for example).

For further work, we would like to improve the present system including some steps previously to the classifier module, that have been demonstrated to improve the final results like a normalization pipeline based on tweets. Also, the necessity of improving the tokenization module to include features like punctuation signs, web addresses, and named entities has become apparent.

Acknowledgments

The research activities described in this paper are funded by DeustoTech INTERNET, Deusto Institute of Technology, a research institute within the University of Deusto.

References

Bermingham, A. and A. F. Smeaton. 2011. On using Twitter to monitor political sentiment and predict election results. In *Bermingham, Adam and Smeaton, Alan F. (2011) On using Twitter to monitor political sentiment and predict election results. In: Sentiment Analysis where AI meets Psychology (SAAIP) Workshop at the International Joint Conference for Natural Language Processing (IJCNLP), 13th November 2011, Chiang Mai, Thailand.*, Chiang Mai, Thailand, November.

Dave, K., S. Lawrence, and D. Pennock. 2003. Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. In *Proceedings of the 12th International Conference on World Wide Web*, WWW '03, pages 519–528, New York, NY, USA. ACM.

Davidov, D., O. Tsur, and A. Rappoport. 2010. Enhanced sentiment learning using twitter hashtags and smileys. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 241–249. Association for Computational Linguistics.

De Choudhury, M., M. Gamon, S. Counts, and E. Horvitz. 2013. Predicting Depression via Social Media. In *ICWSM*.

Gao, H., J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Y. Zhao. 2010. Detecting and Characterizing Social Spam Campaigns. In *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, IMC '10, pages 35–47, New York, NY, USA. ACM.

Go, A., R. Bhayani, and L. Huang. 2009. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1:12.

Han, B. and T. Baldwin. 2011. Lexical Normalisation of Short Text Messages: Makn Sens a #twitter.

Han, B., P. Cook, and T. Baldwin. 2013. unimelb: Spanish Text Normalisation. In *Tweet-Norm@ SEPLN*, pages 32–36.

Hurtado, LF. and F. Pla. 2014. ELiRF-UPV en TASS 2014: Análisis de sentimientos, detección de tópicos y análisis de sentimientos de aspectos en twitter. *Procesamiento del Lenguaje Natural.*

Jakob, N. and I. Gurevych. 2010. Extracting opinion targets in a single-and cross-domain setting with conditional random fields. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1035–1045. Association for Computational Linguistics.

Jungherr, A. 2013. Tweets and Votes, a Special Relationship: The 2009 Federal Election in Germany. In *Proceedings of the 2Nd Workshop on Politics, Elections and Data*, PLEAD '13, pages 5–14, New York, NY, USA. ACM.

M., Hu and Bing L. 2004. Mining and summarizing customer reviews. *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining.*

Martínez-Cámara, E., M. T. Martín-Valdivia, M. L. Molina-Gonzalez, and L. A. Ureña-López. 2013. Bilingual experiments on an opinion comparable corpus. *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis.*

Martinez-Camara, E., M. T. Martin-Valdivia, L. A. Ureña-Lopez, and A. Montejo-Raez. 2014. Sentiment analysis in Twitter. *Natural Language Engineering*, 20(01):1–28, January.

Mislove, A., S. Lehmann, YY. Ahn, JP. Onnela, and J. N. Rosenquist. 2011. Understanding the Demographics of Twitter Users. In *ICWSM*.

Padró, L. and E. Stanilovsky. 2012. Freeling 3.0: Towards wider multilinguality.

Pak, A. and P. Paroubek. 2010. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In *LREC*.

Pang, B. and L. Lee. 2008. Opinion Mining and Sentiment Analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135, January.

Pennacchiotti, M. and AM. Popescu. 2011. A Machine Learning Approach to Twitter User Classification. In *ICWSM*.

Robertson, S.E., S. Walker, S. Jones, Hancock-Beaulieu M. M., and Gatford M. 1995. Okapi at trec-3. *NIST SPECIAL PUBLICATION SP, 109-109.*

Ruiz, P., M. Cuadros, and T. Etchegoyhen. 2013. Lexical Normalization of Spanish Tweets with Preprocessing Rules, Domain-Specific Edit Distances, and Language Models. In *Proceedings of the Tweet Normalization Workshop at SEPLN 2013.*

San Vicente, I. and X. Saralegi. 2014. Looking for features for supervised tweet polarity classification. *Procesamiento del Lenguaje Natural.*

Saralegi, X. and I. San Vicente. 2013. Elhuyar at TASS 2013. *In Proceedings of "XXIX Congreso de la Sociedad Española de Procesamiento de lenguaje natural". Workshop on Sentiment Analysis at SEPLN (TASS 2013). Madrid. pp. 143-150. ISBN: 978-84-695-8349-4.*

Thelwall, M., K. Buckley, and G. Paltoglou. 2012. Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1):163–173, January.

Turney, P. D. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 417–424. Association for Computational Linguistics.

Vilares, D., Y. Doval, M.A Alonso, and C. Gómez-Rodríguez. 2014. Lys at TASS 2014: A prototype for extracting and analysing aspects from spanish tweets. *Procesamiento del Lenguaje Natural.*

Villena-Román, J., J. García-Morera, M. A. García-Cumbreras, E. Martínez-Cámara, M. T. Martín-Valdivia, and L. A. Ureña-López. 2015. Overview of TASS 2015.

Wang, X., L. Tokarchuk, F. Cuadrado, and S. Poslad. 2013. Exploiting Hashtags for Adaptive Microblog Crawling. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ASONAM '13, pages 311–315, New York, NY, USA. ACM.

Wang, X., F. Wei, X. Liu, M. Zhou, and M. Zhang. 2011. Topic Sentiment Analysis in Twitter: A Graph-based Hashtag Sentiment Classification Approach. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, CIKM '11, pages 1031–1040, New York, NY, USA. ACM.