TASS 2015, septiembre 2015, pp 29-34 recibido 17-07-15 revisado 24-07-15 aceptado 28-07-15

Aspect based Sentiment Analysis of Spanish Tweets

Análisis de Sentimientos de Tweets en Español basado en Aspectos

Oscar Araque, Ignacio Corcuera, Constantino Román,
Carlos A. Iglesias y **J. Fernando Sánchez-Rada**
Grupo de Sistemas Inteligentes, Departamento de Ingeniería de Sistemas Telemáticos,
Universidad Politécnica de Madrid (UPM), España
Avenida Complutense, nº 30, 28040 Madrid, España
{oscar.aiborra, ignacio.cplatas, c.romang}@alumnos.upm.es
{cif, jfernando}@dit.upm.es

Resumen: En este artículo se presenta la participación del Grupo de Sistemas Inteligentes (GSI) de la Universidad Politécnica de Madrid (UPM) en el taller de Análisis de Sentimientos centrado en tweets en Español: el TASS2015. Este año se han propuesto dos tareas que hemos abordado con el diseño y desarrollo de un sistema modular adaptable a distintos contextos. Este sistema emplea tecnologías de Procesado de Lenguaje Natural (NLP) así como de aprendizaje automático, dependiento además de tecnologías desarrolladas previamente en nuestro grupo de investigación. En particular, hemos combinado un amplio número de rasgos y léxicos de polaridad para la detección de sentimento, junto con un algoritmo basado en grafos para la detección de contextos. Los resultados experimentales obtenidos tras la consecución del concurso resultan prometedores.
Palabras clave: Aprendizaje automático, Procesado de lenguaje natural, Análisis de sentimientos, Detección de aspectos

Abstract: This article presents the participation of the Intelligent Systems Group (GSI) at Universidad Politécnica de Madrid (UPM) in the Sentiment Analysis workshop focused in Spanish tweets, TASS2015. This year two challenges have been proposed, which we have addressed with the design and development of a modular system that is adaptable to different contexts. This system employs Natural Language Processing (NLP) and machine-learning technologies, relying also in previously developed technologies in our research group. In particular, we have used a wide number of features and polarity lexicons for sentiment detection. With regards to aspect detection, we have relied on a graph-based algorithm. Once the challenge has come to an end, the experimental results are promising.
Keywords: Machine learning, Natural Language Processing, Sentiment analysis, Aspect detection

1 Introduction

In this article we present our participation for the TASS2015 challenge (Villena-Román et al., 2015a). This work deals with two different tasks, that are described next.

The first task of this challenge, Task 1 (Villena-Román et al., 2015b), consists of determining the global polarity at a message level. Inside this task, there are two evaluations: one in which 6 polarity labels are considered (P+, P, NEU, N, N+, None), and another one with 4 polarity labels considered (P, N, NEU, NONE). P stands for *positive*, while N means *negative* and NEU is *neutral*. The "+" symbol is used for intensification of the polarity. It is considered that NONE means absence of sentiment polarity. This task provides a corpus (Villena-Román et al., 2015b), which contains a total of 68.000 tweets written in Spanish, describing a diversity of subjects.

The second and last task, Task 2 (Villena-Román et al., 2015b), is aimed to detect the sentiment polarity at an aspect level using three labels (P, N and NEU). Within this task, two corpora (Villena-Román et al., 2015b) are provided: SocialTV and STOMPOL corpus. We have restricted ourselves to the SocialTV corpus in this edition. This corpus contains 2.773 tweets captured during the celebration of the *2014 Final of Copa del*

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

rey championship[1]. Along with the corpus a set of aspects which appear in the tweets is given. This list is essentially composed by football players, coaches, teams, referees, and other football-related concepts such as crowd, authorities, match and broadcast.

The complexity presented by the challenge has taken us to develop a modular system, in which each component can work separately. We have developed and experimented with each module independently, and later combine them depending on the Task (1 or 2) we want to solve.

The rest of the paper is organized as follows. First, Section 2 is a review of the research involving sentiment analysis in the Twitter domain. After this, Section 3 briefly describes the general architecture of the developed system. Following that, Section 4 describes the module developed in order to confront the Task 1 of this challenge. After this, Section 5 explains the other modules necessaries to address the Task 2. Finally, Section 6 concludes the paper and presents some conclusions regarding our participation in this challenge, as well as future works.

2 Related Work

Centering the attention in the scope of TASS, many researches have experimented, through the TASS corpora, with different approaches to evaluate the performance of these systems. Vilares et al. (2014) present a system relying in machine learning classification for the tasks of sentiment analysis, and a heuristics based approach for aspect-based sentiment analysis. Another example of classification through machine learning is the work of Hurtado and Pla (2014), in which they utilize Support Vector Machine (SVM) with remarkable results. It is common to incorporate linguistic knowledge to this systems, as proposed by Urizar and Roncal (2013), who also employ lexicons in its work. Balahur and Perea-Ortega (2013) deal with this problem using dictionaries and translated data from English to Spanish, as well as machine-learning techniques. An interesting procedure is performed by Vilares, Alonso, and Gómez-Rodríguez (2013): using semantic information added to psychological knowledge extracted from dictionaries, they combine these features to train a machine learning algorithm. Fernández et al. (2013) employ a ranking algorithm using bigrams and added to this a skipgrams scorer, which allow them to create sentiment lexicons that are able to retain the context of the terms. A different approach is by means of the Word2Vec model, used by Montejo-Ráez, García-Cumbreras, and Díaz-Galiano (2104), in which each word is considered in a 200-dimensional space, without using any lexical or syntactical analysis: this allows them to develop a fairly simple system with reasonable results.

3 System architecture

One of ours main goals is to design and develop an adaptable system which can function in a variety of situations. As we have already mentioned, this has taken us to a system composed of several modules that can work separately. Since the challenge proposes two different tasks (Villena-Román et al., 2015b), we will utilize each module when necessary.

Our system is divided into three modules:

- **Named Entity Recognizer (NER)** module. The NER module detects the entities within a text, and classifies them as one of the possibles entities. In the Section 5 a more detailed description of this module and the set of entities given is presented, as it is used in the Task 2.
- **Aspect and Context detection** module. This module is in charge of detecting the remaining aspects -aspects that are not entities and therefore can not be detected as such- and the contexts of all aspects. In the Section 5 this module is described in greater detail since it is only used for tackling the Task 2.
- **Sentiment Analysis** module. As the name suggests, the goal of this module is to classify the given texts using sentiment polarity labels. This module is based on combining NLP and machine learning techniques and is used in both Task 1 and 2. It is explained in more detail next.

3.1 Sentiment Analysis module

The sentiment analysis module relies in a SVM machine-learning model that is trained with data composed of features extracted from the TASS dataset: General corpus for

[1] www.en.wikipedia.org/wiki/2014_Copa_del_Rey_Final

the Task 1 and SocialTV corpus for Task 2 (Villena-Román et al., 2015b).

3.1.1 Feature Extraction

We have used different approaches to design the feature extraction. The reference document taken in the development of the features extraction was made by Mohammad, Kiritchenko, and Zhu (2013). With this in mind, the features extracted from each tweet to form a feature vector are:

- *N-grams*, combination of contiguous sequences of one, two and three tokens consisting on words, lemmas and stem words. As this information can be difficult to handle due to the huge volume of N-grams that can be formed, we set a minimum frequency of three occurrences to consider the N-gram.
- *All-caps*, the number of words with all characters in upper cases that appears in the tweets.
- *POS information*, the frequency of each part-of-speech tag.
- *Hashtags*, the number of hashtags terms.
- *Punctuation marks*, these marks are frequently used to increase the sentiment of a sentence, specially on the Twitter domain. The presence or absence of these marks (?!) are extracted as a new feature, as well as its relative position within the document.
- *Elongated words*, the number of words that has one character repeated more than two times.
- *Emoticons*, the system uses a Emoticons Sentiment Lexicon, which has been developed by Hogenboom et al. (2013).
- *Lexicon Resources*, for each token w, we used the sentiment score $score(w)$ to determine:
  1. Number of words that have a $score(w) \neq 0$.
  2. Polarity of each word that has a $score(w) \neq 0$.
  3. Total score of all the polarities of the words that have a $score(w) \neq 0$.

  The best way to increase the coverage range with respect to the detection of words with polarity is to combine several resources lexicon. The lexicons used are: Elhuyar Polar Lexicon (Urizar and Roncal, 2013), ISOL (Martínez-Cámara et al., 2013), Sentiment Spanish Lexicon (SSL) (Veronica Perez Rosas, 2012), SOCAL (Taboada et al., 2011) and ML-SentiCON (Cruz et al., 2014).
- *Intensifiers*, a intensifier dictionary (Cruz et al., 2014) has been used for calculating the polarity of a word, increasing or decreasing its value.
- *Negation*, explained in 3.1.2.
- *Global Polarity*, this score is the sum of the punctuations from the emoticon analysis and the lexicon resources.

3.1.2 Negation

An important feature that has been used to develop the classifier is the treatment of the negations. This approach takes into account the role of the negation words or phrases, as they can alter the polarity value of the words or phrases they precede.

The polarity of a word changes if it is included in a negated context. For detecting a negated context we have utilized a set of negated words, which has been manually composed by us. Besides, detecting the context requires deciding how many tokens are affected by the negation. For this, we have followed the proposal by Pang, Lee, and Vaithyanathan (2002).

Once the negated context is defined there are two features affected by this: N-grams and lexicon. The negation feature is added to these features, implying that its negated (e.g. positive becomes negative, +1 becomes -1). This approximation is based on the work by Saurí and Pustejovsky (2012).

4 Task 1: Sentiment analysis at global level

4.1 Experiment and results

In this competition it is allowed for submission up to three experiments for each corpus. With this in mind, three experiments have been developed in this task attending to the lexicons that adjust better to the corpus:

- *RUN-1*, there is one lexicon that is adapted well to the corpus, the ElhPolar lexicon. It has been decided to use only this dictionary in the first run.

- *RUN-2*, in this run the two lexicons that have the best results in the experiments have been combined, the ElhPolar and the ISOL.
- *RUN-3*, the last run is a mix of all the lexicon used on the experiments.

Experiment	Accuracy	F1-Score
6labels	61.8	50.0
6labels-1k	48.7	44.6
4labels	69.0	55.0
4labels-1k	65.8	53.1

Table 1: Results of RUN-1 in the Task 1

Experiment	Accuracy	F1-Score
6labels	61.0	49.5
6labels-1k	48.0	44.0
4labels	67.9	54.6
4labels-1k	64.6	53.1

Table 2: Results of RUN-2 in the Task 1

Experiment	Accuracy	F1-Score
6labels	60.8	49.3
6labels-1k	47.9	43.7
4labels	67.8	54.5
4labels-1k	64.6	48.7

Table 3: Results of RUN-3 in the Task 1

5 *Task 2: Aspect-based sentiment analysis*

This task is an extension of the Task 1 in which sentiment analysis is made at the aspect level. The goal in this task is to detect the different aspects that can be in a tweet and afterwards analyze the sentiment associated with each aspect.

For this, we used a pipeline that takes the provided corpus as input and produces the sentiment annotated corpus as output. This pipeline can be divided into three major modules that work in a sequential manner: first the NER, second the Aspect and Context detection, and third the Sentiment Analysis as described below.

5.1 NER

The goal of this module is to detect the words that represent a certain entity from the set of entities that can be identified as a *person* (players and coaches) or an *organization* (teams).

For this module we used the Stanford CRF NER (Finkel, Grenager, and Manning, 2005). It includes a Spanish model trained on news data. To adapt the model, we trained it instead with the training dataset (Villena-Román et al., 2015b) and a gazette. The model is trained with two labels: *Person* (PER) and *Organization* (ORG). The gazette entries were collected from the training dataset, resulting in a list of all the ways the entities (players, teams or coaches) were named. We verified the performance of the Stanford NER by means of cross-validation on the training data. With this, we obtained an average F1-Score of 91.05%.

As the goal of the NER module is to detect the words that represent a specific entity, we used a list of all the ways these entities were named. In this way, once the Stanford NER detect the general entity our improved NER module search in this list and decides the particular entity by matching the pattern of the entity words.

5.2 Aspect and Context detection

This module aims to detect the aspects that are not entities, and thus have not been detected by the NER module. To achieve this, we have composed a dictionary using the training dataset (Villena-Román et al., 2015b) which contains all the manners that all the aspects -including the entities formerly detected- are named. Using this dictionary, this module can detect words that are related to a specific aspect. Although the NER module already detects entities as players, coaches or teams, this module can detect them too: it treats these detected entities as more relevant than its own recognitions, combining in this way the capacity of aspect/entity detection of the NER module and this module.

As for the context detection, we have implemented a graph based algorithm (Mukherjee and Bhattacharyya, 2012) that allows us to extract sets of words related to an aspect from a sentence, even if this sentence has different aspects and mixed emotions. The context of an aspect is the set of words related to that aspect. Besides, we have extended this algorithm in such a way that allow us to configure the scope of this context detection.

Combining this two approaches -aspect and context detection- this module is able to detect the word or words which identify an aspect, and extract the context of this aspect. This context allows us to isolate the sentiment meaning of the aspect, fact that will be very interesting for the sentiment analysis at an aspect level.

We have obtained an accuracy of 93.21% in this second step of the pipeline with the training dataset (Villena-Román et al., 2015b). As for the test dataset (Villena-Román et al., 2015b) we obtained an accuracy of 89.27%[2].

5.3 Sentiment analysis

The sentiment analysis module is the end of the processing pipeline. This module is in charge of classifying the detected aspects in polarity values through the contexts of each aspect. We have used the same model used in Task 1 to analyse every detected aspect in Task 2, given that the detected aspect contexts in Task 2 are similar to the texts analysed in Task 1.

Nevertheless, though using the same model, it is needed to train this model with the proper data. For this, we extracted the aspects and contexts from the train dataset, process the corresponding features (explained in Section 3), and then train the model with these. In this way, the trained machine is fed contexts of aspects that will classify in one of the three labels (as mentioned: positive, negative and neutral).

5.4 Results

By means of connecting these three modules together, we obtain a system that is able to recognize entities and aspects, detect the context in which they are enclosed, and classify them at an aspect level. The performance of this system is showed in the Table 4. The different RUNs represent separate adjustments of the same experiment, in which several parameters are controlled in order to obtain the better performance.

As can be seen in Table 4, the global performance obtained is fairly positive, as our system ranked first in F1-Score and second in Accuracy.

Experiment	Accuracy	F1-Score
RUN-1	**63.5**	**60.6**
RUN-2	62.1	58.4
RUN-3	55.7	55.8

Table 4: Results of each run in the Task 2

6 Conclusions and future work

In this paper we have described the participation of the GSI in the TASS 2015 challenge (Villena-Román et al., 2015a). Our proposal relies in both NLP and machine-learning techniques, applying them jointly to obtain a satisfactory result in the rankings of the challenge. We have designed and developed a modular system that relies in previous technologies developed in our group (Sánchez-Rada, Iglesias, and Gil, 2015). These characteristics make this system adaptable to different conditions and contexts, feature that results very useful in this competition given the diversity of tasks (Villena-Román et al., 2015b).

As future work, our aim is to improve aspect detection by including semantic similarity based on the available lexical resources in the Linguistic Linked Open Data Cloud. To this aim, we will integrate also vocabularies such as Marl (Westerski, Iglesias, and Tapia, 2011). In addition, we are working on improving the sentiment detection based on the social context of users within the MixedEmotions project.

Acknowledgement

This research has been partially funded and by the EC through the H2020 project MixedEmotions (Grant Agreement no: 141111) and by the Spanish Ministry of Industry, Tourism and Trade through the project Calista (TEC2012-32457). We would like to thank Maite Taboada as well as the rest of researchers for providing us their valuable lexical resources.

References

Balahur, A. and José M. Perea-Ortega. 2013. Experiments using varying sizes and machine translated data for sentiment analysis in twitter.

[2]We calculated this metric using the output granted by the TASS uploading page www.daedalus.es/TASS2015/private/evaluate.php.

Cruz, Fermín L, José A Troyano, Beatriz Pontes, and F Javier Ortega. 2014. Building layered, multilingual sentiment lexicons at synset and lemma levels. *Expert Systems with Applications*, 41(13):5984–5994.

Fernández, J., Y. Gutiérrez, J. M. Gómez, P. Martínez-Barco, A. Montoyo, and R. Muñoz. 2013. Sentiment analysis of Spanish tweets using a ranking algorithm and skipgrams.

Finkel, Jenny Rose, Trond Grenager, and Christopher Manning. 2005. Incorporating non-local information into information extraction systems by gibbs sampling. pages 363–370.

Hogenboom, A., D. Bal, F. Franciscar, M. Bal, F. De Jong, and U. Kaymak. 2013. Exploiting emoticons in polarity classification of text.

Hurtado, Ll. and F. Pla. 2014. ELiRF-UPV en TASS 2014: Análisis de sentimientos, detección de tópicos y análisis de sentimientos de aspectos en twitter.

Martínez-Cámara, E., M. Martın-Valdivia, MD Molina-González, and L. Ureña López. 2013. Bilingual experiments on an opinion comparable corpus. *WASSA 2013*, 87.

Mohammad, Saif M, Svetlana Kiritchenko, and Xiaodan Zhu. 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, pages 321–327.

Montejo-Ráez, A, M.A. García-Cumbreras, and M.C. Díaz-Galiano. 2104. Participación de SINAI Word2Vec en TASS 2014.

Mukherjee, Subhabrata and Pushpak Bhattacharyya. 2012. Feature specific sentiment analysis for product reviews. volume 7181 of *Lecture Notes in Computer Science*, pages 475–487. Springer.

Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *Proc. of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pages 79–86. Association for Computational Linguistics.

Sánchez-Rada, J. Fernando, Carlos A. Iglesias, and Ronald Gil. 2015. A Linked Data Model for Multimodal Sentiment and Emotion Analysis. 4th Workshop on Linked Data in Linguistics: Resources and Applications.

Saurí, Roser and James Pustejovsky. 2012. Are you sure that this happened? assessing the factuality degree of events in text. *Computational Linguistics*, 38(2):261–299.

Taboada, Maite, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307.

Urizar, Xabier Saralegi and Iñaki San Vicente Roncal. 2013. Elhuyar at TASS 2013.

Veronica Perez Rosas, Carmen Banea, Rada Mihalcea. 2012. Learning sentiment lexicons in spanish. In *Proc. of the international conference on Language Resources and Evaluation (LREC)*, Istanbul, Turkey.

Vilares, D., M. A. Alonso, and C. Gómez-Rodríguez. 2013. LyS at TASS 2013: Analysing Spanish tweets by means of dependency parsing, semantic-oriented lexicons and psychometric word-properties.

Vilares, David, Yerai Doval, Miguel A. Alonso, and Carlos Gómez-Rodríguez. 2014. LyS at TASS 2014: a prototype for extracting and analysing aspects from Spanish tweets.

Villena-Román, Julio, Janine García-Morera, Miguel A. García-Cumbreras, Eugenio Martínez-Cámara, M. Teresa Martín-Valdivia, and L. Alfonso Ureña-López, editors. 2015a. *Proc. of TASS 2015: Workshop on Sentiment Analysis at SEPLN*, number 1397 in CEUR Workshop Proc., Aachen.

Villena-Román, Julio, Janine García-Morera, Miguel A. García-Cumbreras, Eugenio Martínez-Cámara, M. Teresa Martín-Valdivia, and L. Alfonso Ureña-López. 2015b. Overview of TASS 2015.

Westerski, Adam, Carlos A. Iglesias, and Fernando Tapia. 2011. Linked Opinions: Describing Sentiments on the Structured Web of Data. In *Proc. of the 4th International Workshop Social Data on the Web.*