GTI at TASS 2016: Supervised Approach for Aspect Based Sentiment Analysis in Twitter*

GTI en TASS 2016: Una aproximación supervisada para el análisis de sentimiento basado en aspectos en Twitter

Tamara Álvarez-López, Milagros Fernández-Gavilanes, Silvia García-Méndez, Jonathan Juncal-Martínez, Francisco Javier González-Castaño
GTI Research Group, AtlantTIC
University of Vigo, 36310 Vigo, Spain
{talvarez,mfgavilanes,sgarcia,jonijm}@gti.uvigo.es, javier@det.uvigo.es

Resumen: Este artículo describe la participación del grupo de investigación GTI, del centro AtlantTIC, perteneciente a la Universidad de Vigo, en el TASS 2016. Este taller es un evento enmarcado dentro de la XXXII edición del Congreso Anual de la Sociedad Española para el Procesamiento del Lenguaje Natural. En este trabajo se propone una aproximación supervisada, basada en clasificadores, para la tarea de análisis de sentimiento basado en aspectos. Mediante esta técnica hemos conseguido mejorar las prestaciones de ediciones anteriores, obteniendo una solución acorde con el estado del arte actual.
Palabras clave: Análisis de sentimiento, aspectos, SVM, aprendizaje automático, Twitter

Abstract: This paper describes the participation of the GTI research group of AtlantTIC, University of Vigo, in TASS 2016. This workshop is framed within the XXXII edition of the Annual Congress of the Spanish Society for Natural Language Processing event. In this work we propose a supervised approach based on classifiers, for the aspect based sentiment analysis task. Using this technique we managed to improve the performance of previous years, obtaining a solution reflecting the actual state-of-the-art.
Keywords: Sentiment analysis, aspects, SVM, machine learning, Twitter

1 Introduction

The social media activity is being profused in the recent years, users post opinions and comments in Twitter and in other social platforms. Due to this, there is a huge amount of information available that could be useful for business, in order to design marketing campaigns or to apply any kind of business analysis.

As a consequence, the research on text mining and also on the field of *Sentiment Analysis* (SA) has grown considerably these days. SA is the part of *Natural Language Processing* (NLP) responsible for determining the polarity of a text or a whole sentence. The SA applied to Twitter has to be conducted in a restricted scenario due to the maximum length of the post. However, tweets have other elements we have to consider, like hashtags, mentions and retweets. More concretely, *aspect-based sentiment analysis* (ABSA) consists of extracting opinions, i.e. determining the sentiment polarity, from specific entities in the text (Liu, 2012). Therefore, this task becomes a challenge on the field of NLP.

The TASS Workshop (García-Cumbreras et al., 2016) and the SEPLN conference offer an opportunity for participants to know about the latest advances on the field of NLP for Spanish language.

Many approaches applied to SA can be found in the literature, where it is possible to distinguish between knowledge based approaches (Brooke, Tofiloski, and Taboada, 2009; Fernández-Gavilanes et al., 2016), using grammars and thesaurus and others based on machine learning approaches (Mo-

* This work was partially supported by the Ministerio de Economía y Competitividad under project COINS (TEC2013-47016-C2-1-R) and by Xunta de Galicia (GRC2014/046).

ISSN 1613-0073

hammad, Kiritchenko, and Zhu, 2013). In the last years we can also find deep learning approaches (Bengio, 2009), applied to this task.

We present our supervised *machine learning* (ML) system which consists of a *Support Vector Machine* (SVM) classifier. Our objective is to conduct the SA process at an aspect level, task 2, determining the polarity of a specific given part of a sentence.

The article is structured as follows. Section 2 is a review of the research involving SA in the Twitter domain. Then, the Section 3 describes the applied approach and the implemented system. In Section 4, we show the experimental results of our system. Finally, in Section 5 we present the conclusions and future works.

2 Related work

A large amount of literature related to *Opinion Mining* (OM) and SA can be found (Pang and Lee, 2008; Martínez-Cámara et al., 2016). Most of the systems are applied to Twitter. However others are applied to social media platforms within the micro-blog context. Due to this, the approaches are varied technically and in connection with the purpose.

Two main approaches exist in SA: supervised and unsupervised learning ones. Supervised systems implement classification methods like SVM, *Logistic Regression* (LR), *Conditional Random Fields* (CRF), *K-Nearest Neighbors* (KNN), etc. Cui, Mittal, and Datar (2006) affirmed that SVM are more appropriate for sentiment classification than generative models, due to their capability for working with ambiguity, that is, dealing with mixed feelings. Supervised algorithms are used when the number of classes, as well as the representative members of each class, are known.

Unsupervised systems are based on linguistic knowledge like lexicons, and syntactic features in order to infer the polarity (Paltoglou and Thelwall, 2012). These last techniques represent a more effective approach in the cross-domain context and for multilingual applications. The unsupervised classification algorithms do not work with a training set, in contrast, some of them use clustering algorithms in order to distinguish groups (Li and Liu, 2010).

As noted earlier, the special case of applying SA to Twitter has been fully addressed (Pak and Paroubek, 2010; Han and Baldwin, 2011). Within the chosen solutions, we highlight the text normalization approach (Fabo, Cuadros, and Etchegoyhen, 2013) and the use of key elements in classification approach (Wang et al., 2011). Others hold the advantages of using deep learning techniques in this task (dos Santos and Gatti, 2014).

According to the purpose of the developed systems, it is possible to find applications like classification of product reviews and political sentiment and election results prediction (Bermingham and Smeaton, 2011), among others.

3 System Overview

In this section we make a brief description of the system submitted for Task 2: Aspect-based sentiment analysis. We developed a supervised system, based on a SVM classifier using different features. In the next subsections we explain the different steps required.

3.1 Preprocessing

Before applying any supervised approach to our corpus, some preprocessing is needed. First of all, we have to normalize the text, since in Twitter language we can find abbreviations, mentions, hashtags, URLs or misspellings. In order to do that, we replace the URLs with the "URL" tag and we replace the abbreviations or misspellings with the correct entire word. For mentions and hashtags, we keep them unchanged but deleting the "@" or "#" symbols. Moreover, when a hashtag is composed of several words, we split and treat them as different tokens.

After this, a lexical analysis is carried out. It consists of lemmatization and POS tagging, which are performed by means of Freeling tool (Atserias et al., 2006).

Once we have analysed lexically the texts, we decided to separate the sentences by the different aspects. For doing that, the scope of each aspect is determined, applying the following rules, which are adapted from our English aspect based sentiment anaylisis system (Alvarez-López et al., 2016)

- If there is only one aspect in the sentence, we keep the sentence unchanged, and introduce it entirely as input for the next step.

- If there are multiple aspects, we separate the sentences by punctuation marks, conjunctions or other aspects found.
- If there are several aspects with no words between them, we consider that they belong to the same context, and assign the same polarity to all of them.

3.2 SVM classifier

In this section we describe the strategy followed to determine the sentiment (positive, negative or neutral) for each aspect predefined in corpus.

We develop a SVM classifier, using the *libsvm* library (Chang and Lin, 2011). The inputs for the SVM will be the sentences separated by contexts, as explained in the previous subsection. The features extracted are the following:

- *Word tokens* of nouns, adjectives and verbs in the sentence.
- *Lemmas* of verbs, nouns and adjectives that appear in each sentence.
- *POS tags* of nouns, adjectives and verbs.
- *N-grams* of different length, grouping the words in each sentence.
- *Aspects* appearing in the sentence. We join "aspect"-"entity", defined in each target as a feature.
- *Negations*. We create a negation dictionary, which contains several particles indicating negation, such as "*no*", "*nunca*", etc.

The previous features are all binary ones, assigning the value 1 if the current feature is present in the tweet and the value 0, if not.

4 Experimental Results

The Task 2: Sentiment Analysis at the aspect level consists of assigning a polarity label to each aspect, which were initially marked in the STOMPOL corpus (Martínez-Cámara et al., 2016) raised by the TASS organization. In this way, this corpus provides both polarity labels and the identification of the aspects that appear in each tweet. The aim is to be able to correctly assign to each aspect a positive, negative or neutral polarity.

In this regard, the STOMPOL corpus consists of a set of Spanish tweets related to a number of political issues, such as health or economy, among others. These issues are framed in the political campaign of Andalusian elections in 2015, where each aspect relates to one or several entities that correspond to one of the main political parties in Spain (PP, PSOE, IU, UPyD, Cs and Podemos). The corpus is composed by 1,284 tweets, and has been divided into a training set (784 tweets) and a set of evaluation (500 tweets).

In order to evaluate the performance of the various features for polarity classification at an aspect-based level, we perform a series of ablation experiments as shown in Table 1. We start with the word token baseline classifier, and then add all four sets of features that help to increase performance as measured by accuracy. As we might expect, including the aspect feature has the most marked effect on the performance of polarity classification, although all the features contributed to improving overall performance on STOMPOL corpus.

Type	Accuracy	Improvement
Word token	56.12	
+Lemmas	57.64	+1.52%
+POS tags	58.26	+0.62%
+Aspects	59.94	+1.68%
+Negations	60.60	+0.66%

Table 1: Results for polarity feature ablation experiments on STOMPOL corpus

Due to the low participation of research teams in task 2 this year, we decided to compare our proposal to the systems presented this year and also to that ones of last year, because of the use of the same dataset.

For this reason, Table 2 compares results for our approach with different official ones submitted in 2015 and 2016 TASS editions. In this way, we compared our results for a ML approach based on well-known squared-regularised logistic regression with a snippet of length 4 (Lys-2) described in Vilares et al. (2015), a clustering method focused on grouping authors with similar sociolinguistic insights (TID-spark) described in Park (2015), a recurrent neural network composed of a single long short term memory and a logistic function (Lys-1) described in Vilares et al. (2015), a ML approach based on a

SVM with a snipped of length 5,7 and 10 (ELiRF) described in Hurtado, Plà, and Buscaldi (2015), and the best performing run of the actual task 2 TASS edition (ELiRF-UPV).

Experiment	Task edition	Accuracy
ELiRF-UPV	2016	63.3
ELiRF	2015	63.3
GTI	**2016**	**60.6**
LyS-1	2015	59.9
TID-spark	2015	55.7
Lys-2	2015	54.0

Table 2: Results of different approaches in 2015/2016 TASS editions on STOMPOL corpus

Comparing the results, the performance of our current model is close from the top ranking systems of this and last year.

5 Conclusions and future works

This paper describes the participation of the GTI group in the TASS 2016, Task 2: Aspect-Based Sentiment Analysis. We developed a supervised system based on a SVM classifier for the aspect-based sentiment analysis. The performance of our approach has been compared to that ones submitted this year but also to that ones submitted last year. Experimental results suggest that we need to include explore new features, such as word embedding representations or paraphrase (Zhao and Lan, 2015), in order to improve the performance.

As future work we plan to include new features explained before and to develop a new system which combines different ML classification methods. We are also interested in considering different paradigms of heterogeneous classification, such as deep learning to increase the performance.

References

Alvarez-López, T., J. Juncal-Martınez, M. Fernández-Gavilanes, E. Costa-Montenegro, and F. J. González-Castano. 2016. Gti at semeval-2016 task 5: Svm and crf for aspect detection and unsupervised aspect-based sentiment analysis. *Proceedings of SemEval*, pages 306–311.

Atserias, J., B. Casas, E. Comelles, M. González, L. Padró, and M. Padró. 2006. Freeling 1.3: Syntactic and semantic services in an open-source NLP library. In *Proceedings of LREC*, volume 6, pages 48–55.

Bengio, Y. 2009. Learning deep architectures for AI. *Found. Trends Mach. Learn.*, 2(1):1–127, January.

Bermingham, A. and A. F. Smeaton. 2011. On using Twitter to monitor political sentiment and predict election results.

Brooke, J., M. Tofiloski, and M. Taboada. 2009. Cross-linguistic sentiment analysis: From english to spanish. In G. Angelova, K. Bontcheva, R. Mitkov, N. Nicolov, and N. Nikolov, editors, *RANLP*, pages 50–54. RANLP 2009 Organising Committee / ACL.

Chang, C.-C. and C.-J. Lin. 2011. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27.

Cui, H., V. Mittal, and M. Datar. 2006. Comparative experiments on sentiment classification for online product reviews. In *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 2*, AAAI'06, pages 1265–1270. AAAI Press.

dos Santos, C. N. and M. Gatti. 2014. Deep convolutional neural networks for sentiment analysis of short texts. In *COLING*, pages 69–78.

Fabo, P. R., 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 co-located with 29th Conference of the Spanish Society for Natural Language Processing (SEPLN 2013), Madrid, Spain, September 20th, 2013.*, pages 59–63.

Fernández-Gavilanes, M., T. Álvarez-López, J. Juncal-Martínez, E. Costa-Montenegro, and F. J. González-Castaño. 2016. Unsupervised method for sentiment analysis in online texts. *Expert Systems with Applications*, 58:57–75.

García-Cumbreras, M. A., J. Villena-Román, E. Martínez-Cámara, M. C. Díaz-Galiano, M. T. Martín-Valdivia, and L. A. Ureña-López. 2016. Overview of tass 2016. In *Proceedings of TASS 2016: Workshop on*

Sentiment Analysis at SEPLN co-located with the 32nd SEPLN Conference (SEPLN 2016), Salamanca, Spain, September.

Han, B. and T. Baldwin. 2011. Lexical normalisation of short text messages: Makn sens a #twitter. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, HLT '11, pages 368–378, Stroudsburg, PA, USA. Association for Computational Linguistics.

Hurtado, L. F., F. Plà, and D. Buscaldi. 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.*, pages 75–79.

Li, G. and F. Liu. 2010. A clustering-based approach on sentiment analysis. In *Intelligent Systems and Knowledge Engineering (ISKE), 2010 International Conference on*, pages 331–337. IEEE.

Liu, B. 2012. *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.

Martínez-Cámara, E., M. A. García-Cumbreras, J. Villena-Román, and J. García-Morera. 2016. Tass 2015 - the evolution of the spanish opinion mining systems. *Procesamiento del Lenguaje Natural*, 56:33–40.

Mohammad, S. M., S. Kiritchenko, and X. Zhu. 2013. Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. In *Proceedings of the seventh international workshop on Semantic Evaluation Exercises (SemEval-2013)*, Atlanta, Georgia, USA, June.

Pak, A. and P. Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In N. C. C. Chair), K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, and D. Tapias, editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may. European Language Resources Association (ELRA).

Paltoglou, G. and M. Thelwall. 2012. Twitter, myspace, digg: Unsupervised sentiment analysis in social media. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3(4):66.

Pang, B. and L. Lee. 2008. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135, January.

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.*, pages 99–104.

Vilares, D., Y. Doval, M. A. Alonso, and C. Gómez-Rodríguez. 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.*, pages 47–52.

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.

Zhao, J. and M. Lan. 2015. Ecnu: Leveraging word embeddings to boost performance for paraphrase in Twitter. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 34–39, Denver, Colorado, June. Association for Computational Linguistics.