

# Sentiment Analysis for Twitter: TASS 2015\*

## *Análisis de sentimientos en Twitter: TASS 2015*

**Oscar S. Siordia**

Daniela Moctezuma  
CentroGEO

Col. Lomas de Padierna,  
Delegación Tlalpan,  
CP. 14240, México D.F.

osanchez;dmoctezuma@centrogeo.edu.mx

**Mario Graff**

Sabino Miranda-Jimenez  
Eric S. Tellez

Elio-Atenógenes Villaseñor  
INFOTEC

Ave. San Fernando 37,  
Tlalpan, Toriello Guerra,  
14050 Ciudad de México, D.F.

mario.graff;sabino.miranda;  
eric.tellez;elio.villasenor@infotec.com.mx

**Resumen:** En este artículo se presentan los resultados obtenidos en la tarea 1: clasificación global de cinco niveles de polaridad para un conjunto de tweets en español, del reto TASS 2015. En nuestra metodología, la representación de los tweets estuvo basada en características lingüísticas y de polaridad como lematizado de palabras, filtros de palabras, reglas de negación, entre otros. Además, se utilizaron diferentes transformaciones como LDA, LSI y la matriz TF-IDF, todas estas representaciones se combinaron con el clasificador SVM. Los resultados muestran que LSI y la matriz TF-IDF mejoran el rendimiento del clasificador SVM utilizado.

**Palabras clave:** Análisis de sentimiento, Minería de opinión, Twitter

**Abstract:** In this paper we present experiments for global polarity classification task of Spanish tweets for TASS 2015 challenge. In our methodology, tweets representation is focused on linguistic and polarity features such as lemmatized words, filter of content words, rules of negation, among others. In addition, different transformations are used (LDA, LSI, and TF-IDF) and combined with a SVM classifier. The results show that LSI and TF-IDF representations improve the performance of the SVM classifier applied.

**Keywords:** Sentiment analysis, Opinion mining, Twitter.

## 1 Introduction

In last years the production of textual documents in social media has increased exponentially. This ever-growing amount of available information promotes the research and business activities around opinion mining and sentiment analysis areas. In social media, people share their opinions about events, other people and organizations. This is the main reason why text mining is becoming an important research topic. Automatic sentiment analysis in text is one of most important task in text mining. The task of sentiment classification determines if one document has positive, negative or neutral opin-

ion or any level of each of them. Determining whether a text document has a positive or negative opinion is turning to an essential tool for both public and private companies (Peng, Zuo, y He, 2008). This tool is useful to know “What people think”, which is an important information in order to help to any decision-making process (for any level of government, marketing, etc.) (Pang y Lee, 2008). With this purpose, in this paper we describe the methodology employed for the workshop TASS 2015 (Taller de Análisis de Sentimientos de la SEPLN). The TASS workshop is an event of SEPLN conference, which is a conference in Natural Language Processing for Spanish language. The purpose of TASS is to provide a discussion and a point of sharing about latest research work in the field

\* This work was partially supported by Cátedras CONACYT program

of sentiment analysis in social media (specifically Twitter in Spanish language). In TASS workshop, several challenge tasks are proposed, and furthermore a benchmark dataset is proposed to compare the algorithms and systems of participants (for more details see (Villena-Román et al., 2015)).

Several methodologies to classify tweets from Task 1, Sentiment Analysis at global level of TASS workshop 2015, are presented in this work. This task is to perform an automatic sentiment classification to determine the global polarity (six polarity levels P, P+, NEU, N, N+ and NONE) of each tweet in the provided dataset. With this purpose, several solutions have been proposed in this work.

The paper is organized as follows, a brief overview of related works is shown in Section 2, the proposed methodology is describe in Section 3. Section 4 shows the experimental results and analysis, and finally, Section 5 concludes.

## 2 Related work

Nowadays, several methods have been proposed in the community of opinion mining and sentiment analysis. Most of these works employ twitter as a principal input of data and they aimed to classify entire documents as overall positive or negative polarity levels (sentiment) or rating scores (i.e. 1 to 5 stars).

Such is a case of work presented in (da Silva, Hruschka, y Jr., 2014), which proposes an approach to classify sentiment of tweets by using classifier ensembles and lexicons; where tweets are classified as positive or negative. As a result, this work concludes that classifier ensembles formed by several and diverse components are promising for tweet sentiment classification. Moreover, several state-of-the-art techniques were compared in four databases. The best accuracy result reported was around 75%.

In (Lluís F. Hurtado, 2014) is described the participation of ELiRF research group in TASS 2014 workshop (winners of TASS workshop 2014). Here, the winner approaches used for four tasks are detailed. The proposed methodology uses SVM (Support Vector Machines) with 1-vs-all approach. Moreover, Freeling (Padró y Stanilovsky, 2012) was used as lemmatizer and Tweetmotif<sup>1</sup> to tokenizer to Spanish language. The accuracy

results of classification for task 1 are 64.32% (six labels) and 70.89% (four labels). F1 (F-Measure) is 70.48% in task 2 and 90% in task 3.

Another method to sentiment extraction and classification on unstructured text is proposed in (Shahbaz, Guergachi, y ur Rehman, 2014). Here, five labels were used to sentiment classification: Strongly Positive, Positive, Neutral, Negative and Strongly Negative. The solution proposed combines techniques of Natural language processing at sentence level and algorithms of opinion mining. The accuracy results were 61% for five levels and 75% by reducing to three levels (Positive, negative and neutral).

In (Antunes et al., 2011) an ensemble based on SVM and AIS (Artificial Immune Systems) is proposed. Here, the main idea is that SVM can be enhanced with AIS approaches which can capture dynamic models. Experiments were carried out with the Reuters-21578 benchmark dataset. The reported results show a 95.52% of F1.

An approach of multi-label sentiment classification is proposed in (Liu y Chen, 2015). This approach has three main components: text segmentation, feature extraction and multi-label classification. The features used included raw segmented words and sentiment features based on three sentiment dictionaries: DUTSD, NTUSD and HD. Moreover, here, a detailed study of several multi-label classification methods is conducted, in total 11 state-of-the-art methods have been considered: BR, CC, CLR, HOMER, RAKEL, ECC, MLkNN, and RF-PCT, BRkNN, BRkNN-a and BRkNN-b. These methods were compared in two microblog datasets and the reported results of all methods are around of 0.50 of F1.

In summary, most of works analyzed classify the documents mainly in three polarities: positive, neutral and negative. Moreover, most of works use social media (mainly Twitter) as analyzed documents. In this work, several methods to classify sentiment in tweets are described. These methods were implemented, according with TASS workshop specifications, with the purpose of classify tweets in six polarity levels: P+, P, Neutral, N+, N and None. The proposed method are based on several standard techniques as LDA (Latent Dirichlet Allocation), LSI (Latent Semantic Indexing), TF-IDF matrix in

<sup>1</sup><http://tweetmotif.com/about>

combination with the well-known SVM classifier.

### 3 Proposed solution

In this section the proposed solution is detailed. First, a preprocessing step was carried out, later a Pseudo-phonetic transformation was done and finally the generation of Q-gram expansion was employed.

#### 3.1 Preprocessing step

Preprocessing focuses on the task of finding a good representation for tweets. Since tweets are full of slang and misspellings, we normalize the text using procedures such as error correction, usage of special tags, part of speech (POS) tagging, and negation processing. Error correction consists on reducing words/tokens with invalid duplicate vowels and consonants to valid/standard Spanish words (ruidoooo  $\rightarrow$  ruido; jajajaaa  $\rightarrow$  ja; jijijji  $\rightarrow$  ja). Error correction uses an approach based on a Spanish dictionary, statistical model for common double letters, and heuristic rules for common interjections. In the case of the usage of special tags, twitter’s users (i.e., @user) and urls are removed using regular expressions; in addition, we classify 512 popular emoticons into four classes (P, N, NEU, NONE), which are replaced by a polarity tag in the text, e.g., positive emoticons such as :) , :D are replaced by \_POS, and negative emoticons such as :( , :S are replaced by \_NEG. In the POS-tagging step, all words are tagged and lemmatized using the Freeling tool for Spanish language (Padró y Stanilovsky, 2012), stop words are removed, and only content words (nouns, verbs, adjectives, adverbs), interjections, hashtags, and polarity tags are used for data representation. In negation step, Spanish negation markers are attached to the nearest content word, e.g., ‘no seguir’ is replaced by ‘no\_seguir’, ‘no es bueno’ is replaced by ‘no\_bueno’, ‘sin comida’ is replaced by ‘no\_comida’; we use a set of heuristic rules for negations. Finally, all diacritic and punctuation symbols are also removed.

#### 3.2 Psudo-phonetic transformation

With the purpose of reducing typos and slangs we applied a semi-phonetic transformation. First, we applied the following transformations (with precedence from top to bot-

tom):

$$\begin{aligned}
 cs|xc &\rightarrow x \\
 qu &\rightarrow k \\
 gue|ge &\rightarrow je \\
 gui|gi &\rightarrow ji \\
 sh|ch &\rightarrow x \\
 ll &\rightarrow y \\
 z &\rightarrow s \\
 h &\rightarrow \epsilon \\
 c[a|o|u] &\rightarrow k \\
 c[e|i] &\rightarrow s \\
 w &\rightarrow u \\
 v &\rightarrow b \\
 \Psi\Psi &\rightarrow \Psi \\
 \Psi\Delta\Psi\Delta &\rightarrow \Psi\Delta
 \end{aligned}$$

In our transformation notation, square brackets do not consume symbols and  $\Psi, \Delta$  means for any valid symbols. The idea is not to produce a pure phonetic transformation as in Soundex (Donald, 1999) like algorithms, but try to reduce the number of possible errors in the text. Notice that the last two transformation rules are partially covered by the statistical modeling used for correcting words (explained in preprocessing step). Nonetheless, this pseudo-phonetic transformation does not follow the statistical rules of the previous preprocessing step.

#### 3.3 Q-gram expansion

Along with the placing bag of words representation (of the normalized text) we added the 4 and 5 gram of characters of the normalized text. Blank spaces were normalized and taken into account to the q-gram expansion; so, some q-grams will be over more than one word. In addition of these previous steps, several transformations (LSI, LDA and TF-IDF matrix) were conducted to generate several data models for testing phase.

### 4 Results and analysis

The classifier submitted to the competition was selected using the following procedure. The 7,218 tweets with 6 polarity levels were split in two sets. Firstly, the tweets provided were shuffled and then the first set, hereafter the training set, was created with the first 6,496 tweets (approximately 90% of dataset), and, the second set, hereafter the validation set, was composed by the rest 722

tweets (approximately 10% of dataset). The training set was used to fit a Support Vector Machine (SVM) using a linear kernel<sup>2</sup> with  $C = 1$ , weights inversely proportional to the class frequencies, and using one vs rest multi-class strategy. The validation set was used to select the best classifier using as performance the score F1.

The first step was to model the data using different transformations, namely Latent Dirichlet Allocation (LDA) using an online learning proposed by (Hoffman, Bach, y Blei, 2010), Latent Semantic Indexing (LSI), and TF-IDF.<sup>3</sup> Figure 1 presents the score F1, in the validation set, of a SVM using either LSI or LDA with normalized text, different levels of Q-gram (4-gram and 5-gram), and the number of topics is varied from 10 to 500 as well. It is observed that LSI outperformed LDA in all the configurations tested. Comparing the performance between normalized text, 4-gram, and 5-gram, it is observed an equivalent performance. Given that the implemented LSI depends on the order of the documents more experiments are needed to know whether any particular configuration is statistically better than other. Even though the best configuration is LSI with 400 topics and 5-gram, this system is not competitive enough compared with the performance presented by the best algorithm in TASS 2014.

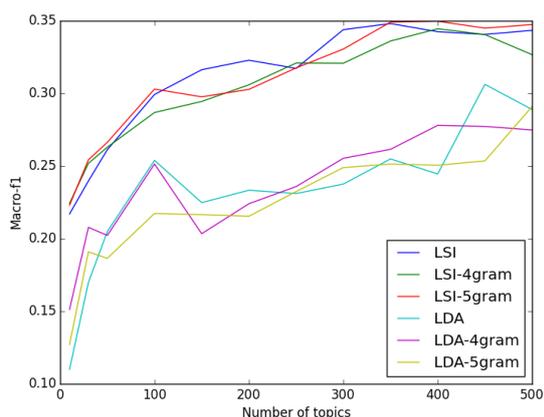


Figure 1: Performance in terms of the score F1 on the validation set for different number of topics using LSI and LDA with different Q-gram.

<sup>2</sup>The SVM was the class LinearSVC implemented in (Pedregosa et al., 2011)

<sup>3</sup>The implementations used for LDA, LSI, and TF-IDF were provided by (Řehůřek y Sojka, 2010).

Table 1 complements the information presented on Figure 1.

The table presents the score F1 per polarity and the average (Macro-F1) for different configurations. The table is divided in five blocks, the first and second correspond to a SVM with LSI (400 topics) and TF-IDF, respectively. It is observed that TF-IDF outperformed LSI; within LSI and TF-IDF it can be seen that 5-gram and 4-gram got the best performance in LSI and TF-IDF, respectively.

The third row block presents the performance when the features are a direct addition of LSI and TF-IDF; here it is observed that the best performance is with 4-gram furthermore it had the best overall performance in N+. The fourth row block complements the previous results by presenting the best performance of LSI and TF-IDF, that is, LSI with 5-gram and TF-IDF with 4-gram. It is observed that this configuration has the best overall performance in P+, N, None and average (Macro-F1). Finally, the last row block gives an indicated of whether the phonetic transformation is making any improvement. The conclusion is that the phonetic transformation is making a difference; however, more experiments are needed in order to know whether this difference is statistically significant.

Based on the score F1 presented on Table 1 the classifier submitted to the competition is a SVM with a direct addition of LSI using 400 topics and 4-gram and LDA with 5-gram. This classifier is identified as INGEOTEC-M1<sup>4</sup> in the competition. The SVM, LSI and LDA were trained with the 7218 tweets and then this instance was used to predict the 6 polarity levels of the competition tweets. This procedure was replicated for the 4 polarity levels competition.

Table 4 presents the accuracy, average recall, precision, and F1 of INGEOTEC-M1 run using the validation set created, a 10-fold crossvalidation on the 7218 tweets and the 1k tweets evaluated by the system's competition. This performance was on the 5 polarity levels challenge. It is observed from the table that the 10-fold crossvalidation gives a much better estimation of the performance

<sup>4</sup>We also submitted another classifier identified as INGEOTEC-E1; however, the algorithm presented a bug that could not be find out on time for the competition.

	P	P+	N	N+	Neutral	None	Average
SVM + LSI							
Text	0.238	0.549	0.403	0.348	0.025	0.492	0.343
4-gram	0.246	0.543	0.404	0.333	0.048	0.533	0.351
5-gram	0.246	0.552	0.462	0.356	0.000	0.575	0.365
SVM + TF-IDF							
Text	0.271	0.574	0.414	0.407	0.103	0.511	0.380
4-gram	0.290	0.577	0.477	0.393	0.130	0.589	0.409
5-gram	0.302	0.577	0.476	0.379	0.040	0.586	0.393
SVM + {LSI + TF-IDF}							
4-gram	0.297	0.578	0.471	<b>0.421</b>	0.142	0.578	0.415
5-gram	0.307	0.567	0.474	0.391	0.040	0.579	0.393
SVM + {LSI with 4-gram + TF-IDF with 5-gram}							
4-5-gram	0.282	<b>0.596</b>	<b>0.481</b>	0.407	0.144	<b>0.595</b>	<b>0.417</b>
SVM + {LSI + TF-IDF without phonetic transformation}							
4-5-gram	<b>0.324</b>	0.577	0.459	0.395	<b>0.150</b>	0.593	0.416

Table 1: Score F1 per polarity level and average (Macro-F1) on the validation set for LSI (with 400 topics) and TF-IDF with different levels of Q-gram. The best performance in each polarity is indicated in boldface.

of the classifier when tested on 1k tweets of the competition (90% of training and 10% of validation).

In summary, in this work the best result reached was a 0.404 of F1. This result was achieved with a combination of LSI with 4-gram + TF-IDF with 5-gram, using a SVM classifier (one-vs-one approach).

	Acc.	Recall	Precision	F1
Val.	0.471	0.428	0.421	0.417
10-fold	0.443	0.397	0.395	0.393
Comp.	0.431	0.411	0.398	0.404

Table 2: Accuracy (Acc.), average recall, average precision and average F1 of the classifier in the validation set (Val.), using a 10-fold cross-validation (7,218 tweets), and as reported by the competition (comp.) on 1k tweets.

## 5 Conclusions

In this contribution, we presented the approach used to tackle the polarity classification task of Spanish tweets of TASS 2015. From the results, it is observed that a combination of different data models, in this case LSI and TF-IDF, improves the performance of a SVM classifier. It also noted that the phonetic transformation makes an improvement; however, more experiments are needed to know whether this improvement is statistically significant. As a result, we obtained a 0.404 of F1 (macro-F1) in sentiment classification task at five levels, with the proposed

solution. This proposed solution uses a combination of LSI with 4-gram + TF-IDF with 5-gram, and a SVM classifier (one-vs-one approach).

## Acknowledgements

This research is partially supported by the Cátedras CONACyT project. Furthermore, the authors would like to thank CONACyT for supporting this work through the project 247356 (PN2014).

## Bibliography

- Antunes, Mário, Catarina Silva, Bernardete Ribeiro, y Manuel Correia. 2011. A hybrid ais-svm ensemble approach for text classification. En *Adaptive and Natural Computing Algorithms*, volumen 6594 de *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, páginas 342–352.
- da Silva, Nádia F.F., Eduardo R. Hruschka, y Estevam R. Hruschka Jr. 2014. Tweet sentiment analysis with classifier ensembles. *Decision Support Systems*, 66(0):170 – 179.
- Donald, E Knuth. 1999. The art of computer programming. *Sorting and searching*, 3:426–458.
- Hoffman, Matthew, Francis R Bach, y David M Blei. 2010. Online learning for latent dirichlet allocation. En *advances*

- in neural information processing systems*, páginas 856–864.
- Liu, Shuhua Monica y Jiun-Hung Chen. 2015. A multi-label classification based approach for sentiment classification. *Expert Systems with Applications*, 42(3):1083 – 1093.
- Lluís F. Hurtado, Ferran Pla. 2014. Elirf-upv en TASS 2014: Análisis de sentimientos, detección de tópicos y análisis de sentimientos de aspectos en twitter. *Proc. of the TASS workshop at SEPLN 2014*.
- Padró, Lluís y Evgeny Stanilovsky. 2012. Freeling 3.0: Towards wider multilinguality. En *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*, Istanbul, Turkey, May. ELRA.
- Pang, Bo y Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, y E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Peng, Tao, Wanli Zuo, y Fengling He. 2008. Svm based adaptive learning method for text classification from positive and unlabeled documents. *Knowledge and Information Systems*, 16(3):281–301.
- Řehůřek, Radim y Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. En *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, páginas 45–50, Valletta, Malta, Mayo. ELRA.
- Shahbaz, M., A. Guergachi, y R.T. ur Rehman. 2014. Sentiment miner: A prototype for sentiment analysis of unstructured data and text. En *Electrical and Computer Engineering (CCECE), 2014 IEEE 27th Canadian Conference on*, páginas 1–7, May.
- Villena-Román, Julio, Janine García-Morera, Miguel A. García-Cumbreras, Eugenio
- Martínez-Cámara, M. Teresa Martín-Valdivia, y L. Alfonso Ureña-López. 2015. Overview of TASS 2015.