

GTI-Gradient at TASS 2015: A Hybrid Approach for Sentiment Analysis in Twitter*

GTI-Gradient en TASS 2015: Una aproximación híbrida para el análisis de sentimiento en Twitter

<p>Tamara Álvarez-López Jonathan Juncal-Martínez Milagros Fernández-Gavilanes Enrique Costa-Montenegro Francisco Javier González-Castaño GTI Research Group, AtlantTIC University of Vigo, 36310 Vigo, Spain {talvarez, joni, milagros.fernandez, kike}@gti.uvigo.es, javier@det.uvigo.es</p>	<p>Hector Cerezo-Costas Diego Celix-Salgado Gradient 36310 Vigo, Spain {hcerezo, dcelix}@gradient.org</p>
--	---

Resumen: Este artículo describe la participación en el workshop TASS 2015 del grupo de investigación GTI, del centro AtlantTIC, perteneciente a la Universidad de Vigo, y el centro tecnológico Gradient. Ambos grupos han desarrollado conjuntamente una aproximación híbrida para el análisis de sentimiento global en Twitter, presentado en la tarea 1 del TASS. Se propone un sistema basado en clasificadores y en aproximaciones sin supervisión, construidas mediante léxicos de polaridad y estructuras sintácticas. La combinación de los dos tipos de sistemas ha proporcionado resultados competitivos sobre los conjuntos de prueba propuestos.

Palabras clave: Léxico polar, análisis de sentimiento, dependencias sintácticas.

Abstract: This paper describes the participation of the GTI research group of AtlantTIC, University of Vigo, and Gradient (Galician Research and Development Centre in Advanced Telecommunications), in the TASS 2015 workshop. Both groups have worked together in the development of a hybrid approach for sentiment analysis, at a global level, of Twitter, proposed in task 1 of TASS. A system based on classifiers and unsupervised approaches, built with polarity lexicons and syntactic structures, is presented here. The combination of both approaches has provided highly competitive results over the given datasets.

Keywords: Polarity lexicon, sentiment analysis, dependency parsing.

1 Introduction

In recent years, research on the field of *Sentiment Analysis* (SA) has increased considerably, due to the growth of user content generated in social networks, blogs and other platforms on the Internet. These are considered valuable information for companies, which seek to know or even predict the acceptance of their products, to design their marketing campaigns more efficiently. One of these sources of information is Twitter, where users are allowed to write about any

topic, using colloquial and compact language. As a consequence, SA in Twitter is specially challenging, as opinions are expressed in one or two short sentences. Moreover, they include special elements such as hashtags or mentions. Henceforth, additional treatments must be applied when analyzing a tweet.

Numerous contributions on this subject can be found in the literature. Most of them are supervised machine learning approaches, although unsupervised semantic can also be found in this field. The first ones are usually classifiers built from features of a “bag of words” representation (Pak and Paroubek, 2010), whilst the second ones try to model linguistic knowledge by using polarity dictionaries (Brooke, Tofiloski, and Taboada,

* This work was supported by the Spanish Government, co-financed by the European Regional Development Fund (ERDF) under project TACTICA, and RedTEIC (R2014/037).

2009), which contain words tagged with their semantic orientation. These strategies involve lexics, syntax or semantics analyt-ics (Quinn et al., 2010) with a final aggregation of their values.

The TASS evaluation workshop aims at providing a benchmark forum for comparing the latest approaches in this field. In this way, our team only took part in task 1 related to SA in Twitter. This task encompasses four experiments. The first consists of evaluating tweet polarities over a big dataset of tweets, with only 4 tags, *positive* (P), *negative* (N), *neutral* (NEU) or *no opinion* (NONE) expressed. In the second experiment, the same evaluation is requested over a smaller selection of tweets. The third and fourth experiments propose the same two datasets, respectively, but with 6 different possible tags, including *strong positive* (P+) and *strong negative* (N+). In addition, a training set has been provided, in order to build the models (Villena-Román et al., 2015).

The rest of this article is structured as follows: Section 2 presents in detail the system proposed. Section 3 describes the results obtained and some experiments performed over the target datasets. Finally, Section 4 summarizes the main findings and conclusions.

2 System overview

Our system is a combination of two different approaches. The first approach is an unsupervised approach, based on sentiment dictionaries, which are automatically generated from the set of tweets to analyze (a set of positive and negative seeds, created manually, are necessary to start the process). The second is a supervised approach, which employs *Conditional Random Fields* (CRFs) (Sutton and McCallum, 2011) to detect the scope of potential polarity shifters (e.g. intensification, reversal verbs and negation particles). This information is combined to conform high-level features which are fed to a statistical classifier to finally obtain the polarity of the message.

In this way, both approaches have been previously adapted to the English language and submitted to the *SemEval-2015* sentiment analysis task, achieving good rankings and results separately (Fernández-Gavilanes et al., 2015; Cerezo-Costas and Celix-Salgado, 2015). Because both have shown particular advantages, we decided to build a

hybrid system. The following subsections explain the two approaches, as well as the strategy followed to combine them.

2.1 Previous steps

The first treatments to be applied over the set of tweets rely on *natural language processing* (NLP) and are common to both approaches. They include preprocessing, lexical and syntactic analysis and generation of sentiment lexicons.

2.1.1 Preprocessing

The language used on Twitter contains words that are not found in any dictionary, because of orthographic modifications. The aim here is to normalize the texts to get closer to formal language. The actions executed in this stage are the substitution of emoticons, which are divided in several categories, by equivalent Spanish words, for example, :) is replaced by *e.feliz*; the substitution of frequent abbreviations; the removal of repeated characters and the replacement of specific Twitter words such as hashtags, as well as mentions or URLs, by *hashtag*, *mencion* and URL tags, respectively.

2.1.2 Lexical and syntactic analysis

After the preprocessing, the input text is morphologically tagged to obtain the *part-of-speech* (PoS) associated with a word, as adjectives, adverbs, nouns and verbs. Finally, a dependency tree is created with the syntactic functions annotated. These steps are performed with the *Freeling tool* (Padró and Stanilovsky, 2012).

2.1.3 Sentiment lexicons

Sentiment lexicons have been used in many supervised and unsupervised approaches for sentiment detection. They are not so common in Spanish as in English, although there are some available, such as SOCAL (Brooke, Tofiloski, and Taboada, 2009), SpanishDAL (Dell’ Amerlina Ríos and Gravano, 2013) and eSOL lexicon (Molina-González et al., 2013). Some of them are lists of words with an associated number, which represents the polarity, and others are just lists of negative and positive words.

However, these dictionaries are not contextualized, so we generate additional ones automatically from the words in the syntactic dependencies of each tweet, considering verbs, nouns and adjectives. Then, we apply a polarity expansion algorithm based on

graphs (Cruz et al., 2011). The starting point of this algorithm is a set of positive and negative words, used as seeds, extracted from the most negative and positive words in the general lexicons. This dictionary will contain a list of words with their polarity associated, which is a real number in $[-5, 5]$. Finally, we merge each general lexicon with the automatically created ones, obtaining several dictionaries, depending on the combination applied, to feed our system.

As explained in the next sections, the dictionaries obtained must be adapted for using them in the supervised approach. In this case, only a list of positive and negative words is required, with no associated values.

2.2 Supervised approach

This subsection presents the supervised approach for the tagging of Spanish tweets. After the previous steps, lexical, PoS and CRF labels are jointly combined to build the final features that define the input to a logistic regression classifier.

The system works in two phases. First, a learning phase is applied in which the system learns the parameters of the supervised model using manually tagged data. Second, the supervised model is only trained with the training vector provided by the organization.

2.2.1 Strategy initialization

This strategy uses several dictionaries as an input for different steps of the feature extraction process. Hence, a polarity dictionary, previously created and adapted, containing positive and negative words, is provided as input in this step. Certain polarity shifters play an important role in the detection of the polarity of a sentence. Previous attempts in the academic literature followed different approaches, like hand-crafted rules (Sidorov et al., 2013) or CRFs (Lapponi et al., 2012). We employ CRFs to detect the scope of the polarity shifters such as denial particles (e.g. *sin* (without), *no* (no)) and reversal verbs, (e.g. *evitar* (avoid), *solucionar* (solve)). In order to obtain the list of reversal verbs and denial particles, basic syntactic rules and a manual supervision were applied to the final system. A similar approach can be found in Choi and Cardie (2008).

Additional dictionaries are used in the system (e.g. adversative particles or superlatives) but their main purpose is to give support of the learning steps with the polarity

shifters.

2.2.2 Polarity modifiers

Polarity shifters are specific particles (e.g. *no* (no)), words (e.g. *evitar* (avoid)), or constructions (e.g. *fuera de* (out of)) that modify the polarity of the words under their influence. Detecting these scopes of influence closely related to the syntactic graphs is difficult due to the unreliability of dependency and syntactic parsers on Twitter. To solve this problem we trained sequential CRFs for each problem we wanted to solve. CRFs are supervised techniques that assign a label to each component (in our case the words of a sentence).

Our system follows a similar approach to Lapponi, Read and Ovreid (2012) but it has been enhanced to track intensification, comparisons within a sentence, and the effect of adversative clauses (e.g. sentences with *pero* (but) particles). We refer the reader to Cerezo-Costas and Celix-Salgado (2015) to see the input features employed by the CRFs.

2.2.3 Classifier

All the characteristics from previous steps are included as input features of a statistical classifier. The lexical features (word, stem, word and stem bigrams and flags extracted from the polar dictionaries) are included with PoS and the labels from the CRFs. The learning algorithm employed was a logistic regressor. Due to the size of the feature space and its sparsity, l1 (0.000001) and l2 (0.00005) regularization was applied to learn the most important features and discard the least relevant for the task.

2.3 Unsupervised approach

The unsupervised approach is based on generated polarity lexicons applied to the syntactic structures previously obtained. The final sentiment result of each tweet is expressed as a real number, calculated as follows: first, the words in the dependency tree are assigned a polarity from the sentiment dictionary; second, a polarity value propagation based on Caro and Grella (2013) is performed on each dependency tree from the lower nodes to the root, by means of propagation rules explained later. The real end value is classified as P, N, NEU or NONE, according to defined intervals.

2.3.1 Intensification rules

Usually, adverbs act as intensifiers or diminishers of the word that follows them.

For example, there is a difference between *bonito* (beautiful) and *muy bonito* (very beautiful). The first one has a positive connotation, whose polarity is increased by the adverb *muy* (very). So, its semantic orientation is altered. Therefore, the intensification is achieved by assigning a positive or negative percentage in the intensifiers and diminishers (Zhang, Ferrari, and Enjalbert, 2012).

2.3.2 Negation rules

If words that imply denial appear in the text, such as *no* (no), *nunca* (never) or *ni* (neither) (Zhang, Ferrari, and Enjalbert, 2012), the meaning is completely altered. For example, there is a difference between *Yo soy inteligente* (I am intelligent) and *Yo no soy inteligente* (I am not intelligent). The meaning of the text changes from positive to negative, due to the negator nexus. Therefore, the negation is identified by detecting the affected scope in the dependency tree, for subsequently applying a negative factor to all affected nodes.

2.3.3 Polarity conflict rules

Sometimes, two words appearing together express opposite sentiments. The aim here is to detect these cases, known as *polarity conflicts* (Moilanen and Pulman, 2007). For example, in *fiesta aburrida* (boring party), *fiesta* (party) has a polarity with a positive connotation, which is reduced by the negative polarity of *aburrida* (boring). Moreover, in *náufrago ileso* (unharmed castaway), *náufrago* (castaway) has a negative polarity, which is reduced by the positive polarity of *ileso* (unharmed), yielding a new positive connotation.

2.3.4 Adversative/concessive rules

There is a point in common between adversative and concessive sentences. In both cases, one part of the sentence is in contrast with the other. While the former express an objection in compliance with what is said in the main clause, the latter express a difficulty in fulfilling the main clause. We can assume that both constructions will restrict, exclude, amplify or diminish the sentiment reflected in them. Some adversative nexus can be *pero* (but) or *sin embargo* (however) (Poria et al., 2014), whereas concessive ones can be *aunque* (although) or *a pesar de* (in spite of) (Rudolph, 1996). For example, in the adversative sentence *Lo había prometido, pero me ha*

sido imposible (I had promised it, but it has been impossible), the most important part is the one with the nexus, whereas in the concessive sentence *A pesar de su talento, han sido despedidos* (In spite of their talent, they have been fired), it is the part without the nexus.

2.4 Combination strategy: the hybrid approach

In order to decide the final polarity of each tweet, we combine both approaches as follows: applying the supervised approach, 15 different outputs are generated, randomizing the training vector and selecting a subset of them for training (leaving out 1500 records in each iteration). Then, another 15 outputs are generated applying the unsupervised approach, using 15 different lexicons, created by combining each general lexicon (SDAL, SOCIAL, eSOL) with the automatically generated one, and also combining 3 or 4 of them. During this process, when a word appears in several dictionaries, we apply a weighted average, varying the relevance assigned to each dictionary, thus providing more output combinations. Afterwards, we apply a majority voting method among the 30 outputs obtained to decide the final tweet polarity. This strategy has shown better performance than only one of the approaches by itself, making the combination of both a good choice for the experiments, as explained in the next section.

3 Experimental results

The performance in task 1 was measured by means of the accuracy (correct tweets according to the gold standard). Table 1 shows the results, where accuracy is represented for each experiment, as well as the results of the top ranking systems, out of 16 participants for the 6-tag subtasks, and 15 participants for the 4-tag subtasks.

It can be noticed that the results for 6 tags are considerably worse than those for 4 tags. It appears that it becomes more difficult for our system, and for any system in general, to detect positive or negative intensities, rather than just distinguishing positive from negative. Furthermore, we can also observe in the results for the smaller dataset that accuracy diminishes notably for both experiments.

As previously said, in order to obtain our results, we combined both approaches, by means of a majority voting method. On the

Team	Accuracy			
	6	6 (1k)	4	4 (1k)
LIF	67.2 ₂	51.6 ₁	72.6 ₁	69.2 ₁
GTI-GRAD	59.2 ₅	50.9 ₂	69.5 ₃	67.4 ₂
ELIRF	67.3 ₁	48.8 ₃	72.5 ₂	64.5 ₅
GSI	61.8 ₃	48.7 ₄	69.0 ₄	65.8 ₃
LYS	56.8 ₆	43.4 ₅	66.4 ₅	63.4 ₉

Table 1: GTI-Gradient accuracy obtained for each experiment, compared to the top ranking systems. The subscripts represent the position in the ranking.

one hand, the outputs resulting from the supervised approach were generated by applying classifiers, with different training records. On the other hand, the unsupervised approach requires the use of several dictionaries, getting a real number polarity for each tweet, and then applying an interval to determine when a tweet carries an opinion or not. This interval is fixed to $[-1, 1]$ for no opinion. In addition, the number of words containing a polarity is taken into account to decide the neutrality of a tweet. That is, if it contains polar words but the total result lies in $[-1, 1]$, this means that there is a contraposition of opinions, so the tweet is tagged as neutral. However, our combined system seemed to work not so well for neutral texts, specially in the bigger datasets. This may be due to the small proportion of neutral tweets through out the whole dataset, as they only represent a 2.15% of the total number of tweets, rising to 6.3% for the small datasets.

For the 6-tag experiments, P+ and N+ tags were determined with the supervised approach. This decision was taken because the unsupervised approach was not able to discriminate efficiently between P and P+ or between N and N+.

Table 2 shows several experiments with the supervised and unsupervised models, as well as with the combined one, so we can appreciate the improvement in the last case. These results were obtained by applying a majority voting method to each approach separately, with 15 outputs, and then to 30 outputs of the combined result.

4 Conclusions

This paper describes the participation of the GTI Research Group (AtlantTIC, Univer-

Approach	Accuracy			
	6	6 (1k)	4	4 (1k)
Supervised	58.4	48.3	66.4	63.8
Unsupervised	47.8	41.8	66.3	65.1
Combined	59.2	50.9	69.5	67.4

Table 2: Comparative accuracy analysis. Both approaches and combined output.

sity of Vigo) and Gradient (Galician Research and Development Centre in Advanced Telecommunications) in TASS 2015 Task 1: Sentiment Analysis at global level. We have presented a hybrid system, combining supervised and unsupervised approaches, which has obtained competitive results and a good position in the final ranking.

The unsupervised approach consists of sentiment propagation rules on dependencies, whilst the supervised one is based on classifiers. This combination seems to work considerably well in this task.

There is still margin for improvement, mostly in neutral tweets detection and more refined distinction of degrees of positivity and negativity.

References

- Brooke, J., M. Tofiloski, and M. Taboada. 2009. Cross-linguistic sentiment analysis: From English to Spanish. In *Proc. of the Int. Conf. RANLP-2009*, pages 50–54, Borovets, Bulgaria. ACL.
- Caro, L. Di and M. Grella. 2013. Sentiment analysis via dependency parsing. *Computer Standards & Interfaces*, 35(5):442–453.
- Cerezo-Costas, H. and D. Celix-Salgado. 2015. Gradient-analytics: Training polarity shifters with CRFs for message level polarity detection. In *Proc. of the 9th Int. Workshop on Semantic Evaluation (SemEval 2015)*, pages 539–544, Denver, Colorado. ACL.
- Choi, Y. and C. Cardie. 2008. Learning with Compositional Semantics as Structural Inference for Subsentential Sentiment Analysis. In *Proc. of the Conf. on Empirical Methods in Natural Language Processing*, pages 793–801.
- Cruz, F. L., J. A. Troyano, F. J. Ortega, and F. Enríquez. 2011. Automatic expan-

- sion of feature-level opinion lexicons. In *Proc. of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, pages 125–131, Stroudsburg, PA, USA. ACL.
- Dell’ Amerlina Ríos, M. and A. Gravano. 2013. Spanish dal: A Spanish dictionary of affect in language. In *Proc. of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 21–28, Atlanta, Georgia. ACL.
- Fernández-Gavilanes, M., T. Álvarez-López, J. Juncal-Martínez, E. Costa-Montenegro, and F. J. González-Castaño. 2015. GTI: An unsupervised approach for sentiment analysis in Twitter. In *Proc. of the 9th Int. Workshop on Semantic Evaluation (SemEval 2015)*, pages 533–538, Denver, Colorado. ACL.
- Lapponi, E., J. Read, and L. Ovreliid. 2012. Representing and Resolving Negation for Sentiment Analysis. In *IEEE 12th Int. Conf. on Data Mining Workshops (ICDMW)*, pages 687–692.
- Lapponi, E., E. Velldal, L. Øvreliid, and J. Read. 2012. Uio 2: Sequence-Labeling Negation Using Dependency Features. In *Proc. of the 1st Conf. on Lexical and Computational Semantics*, volume 1, pages 319–327.
- Moilanen, K. and S. Pulman. 2007. Sentiment composition. In *Proc. of RANLP 2007*, Borovets, Bulgaria.
- Molina-González, M. D., E. Martínez-Cámara, M. T. Martín-Valdivia, and J. M. Perea-Ortega. 2013. Semantic orientation for polarity classification in Spanish reviews. *Expert Syst. Appl.*, 40(18):7250–7257.
- Padró, L. and E. Stanilovsky. 2012. Freeling 3.0: Towards wider multilinguality. In *Proc. of the Language Resources and Evaluation Conf. (LREC 2012)*, Istanbul, Turkey. ELRA.
- Pak, A. and P. Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *Proc. of the Int. Conf. on Language Resources and Evaluation, LREC 2010, Valletta, Malta*.
- Poria, S., E. Cambria, G. Winterstein, and G. Huang. 2014. Sentic patterns: Dependency-based rules for concept-level sentiment analysis. *Knowledge-Based Systems*, 69(0):45–63.
- Quinn, K. M., B. L. Monroe, M. Colaresi, M. H. Crespín, and D. R. Radev. 2010. How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science*, 54(1):209–228.
- Rudolph, E. 1996. *Contrast: Adversative and Concessive Relations and Their Expressions in English, German, Spanish, Portuguese on Sentence and Text Level*. Research in text theory. Walter de Gruyter.
- Sidorov, G., S. Miranda-Jiménez, F. Viveros-Jiménez, A. Gelbukh, N. Castro-Sánchez, F. Velásquez, I. Díaz-Rangel, S. Suárez-Guerra, A. Treviño, and J. Gordon. 2013. Empirical Study of Machine Learning Based Approach for Opinion Mining in Tweets. In Ildar Batyrshin and Miguel González Mendoza, editors, *Advances in Artificial Intelligence*, volume 7629 of *LNCIS*. Springer Berlin Heidelberg, pages 1–14.
- Sutton, C. and A. McCallum. 2011. An Introduction to Conditional Random Fields. *Machine Learning*, 4(4):267–373.
- 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. In *TASS 2015: Workshop on Sentiment Analysis at SEPLN*.
- Zhang, L., S. Ferrari, and P. Enjalbert. 2012. Opinion analysis: The effect of negation on polarity and intensity. In Jeremy Jancsary, editor, *Proc. of KONVENS 2012*, pages 282–290. ÖGAI, September. PATHOS 2012 workshop.