

GTI: An Unsupervised Approach for Sentiment Analysis in Twitter

Milagros Fernández-Gavilanes, Tamara Álvarez-López, Jonathan Juncal-Martínez,
Enrique Costa-Montenegro, Francisco Javier González-Castaño

GTI Research Group

AtlantTIC Centre, School of Telecommunication Engineering, University of Vigo
36310 Vigo, Spain

{milagros.fernandez, talvarez, jonijm, kike}@gti.uvigo.es,
javier@det.uvigo.es

Abstract

This paper presents the approach of the GTI Research Group to SemEval-2015 task 10 on Sentiment Analysis in Twitter, or more specifically, subtasks A (Contextual Polarity Disambiguation) and B (Message Polarity Classification). We followed an unsupervised dependency parsing-based approach using a sentiment lexicon, created by means of an automatic polarity expansion algorithm and *Natural Language Processing* techniques. These techniques involve the use of linguistic peculiarities, such as the detection of polarity conflicts or adversative/concessive subordinate clauses. The results obtained confirm the competitive and robust performance of the system.

1 Introduction

The domain of sentiment analysis has received increasing attention in recent years (Liu, 2012), particularly due to the growth of the Internet and content generated by users of social networks and other platforms. Some of these, such as Twitter, allow people to express their opinions using colloquial, compact language. The result is a new form of expression that may in the long term become a source of extremely valuable information. An increasing number of companies are now focusing their marketing campaigns on online comments, sentiments, and opinions of brands from clients or potential clients, and some are even trying to predict the acceptance and rejection of certain products using this information (Jansen et al., 2009).

Even though the approaches used for this purpose are numerous and varied, they can be broadly divided into two categories: supervised machine-learning and unsupervised semantic-based approaches. The former are often classifiers built from features of a “*bag of words*” representation (Hu and Liu, 2004; Pak et al., 2010). In other words, they consist of automatically analyzing n -grams in search of recurrent combinations of opinion words. The latter aim at capturing and modeling linguistic knowledge through the use of dictionaries (Taboada et al., 2011) containing words that are tagged with their semantic orientation. These methods detect the words present in a text using different strategies involving lexics, syntax or semantics (Quinn et al., 2010) and then aggregate their values. Such methods usually combine two or more levels of analysis.

In recent years, work on sentiment classification using different types of texts has shown that specialized methods are required. For example, emotions are not conveyed in the same manner in newspaper articles as in blogs, reviews, forums or other types of user-generated content (Balahur, 2013). Dealing with sentiment in Twitter, thus, requires an analysis of the characteristics of tweets and the design of adapted methods.

This paper presents a method for sentiment analysis in English that uses dependency parsing to determine the polarity of tweets, using a previously created sentiment lexicon and considering the special structure and linguistic content of these postings.

The remainder of this article is structured as follows: Section 2 provides a brief description of the task and some of its subtasks. Section 3 presents in

detail the system proposed for the performance of these tasks, and Section 4 shows the results obtained and discusses them. Finally, Section 5 summarizes the main findings and conclusions.

2 Task Description

This paper describes our contribution to the SemEval-2015 Task 10: Sentiment Analysis in Twitter. Of the five subtasks established, we participated in two:

- **Contextual Polarity Disambiguation (A)**, on determining the polarity of a marked instance of a word or phrase in the context of a given message.
- **Message Polarity Classification (B)**, on classifying the content of a whole message.

This year there were two datasets for testing candidate systems for subtasks A and B: The Official 2015 Test and a Progress Test. The first test consisted of a set of Twitter messages (Rosenthal et al., 2015) whilst the second test was a rerun of SemEval-2014 Task 9 (Rosenthal et al., 2014), which includes Twitter messages and other kinds of texts from different domains. Datasets formed by the datasets given in SemEval-2013 Task 2 (Nakov et al., 2013) were also provided for training and development. In our case, the approach does not involve any training, and all the datasets were used to test the behavior of our system.

3 System Overview

The main objective of the tasks was to detect whether a marked instance of word/phrase in a given context (A) or message (B) expresses *positive*, *negative* or *neutral* sentiment. Most learning- or lexicon-based systems do not usually take into account relations between words, although they try to simulate comprehension of some linguistic constructions, such as negation, but this does not always work correctly due to the complexity of human language. For this reason, in this paper, we propose an alternative system to exploit the information present in dependencies obtained from a parsing analysis, *without the need for any kind of training*. The research we

describe in this section has several linguistic peculiarities that were used to improve sentiment detection performance. Our method, which was fully unsupervised, consisted of four stages, which are each explained in detail below.

3.1 Preprocessing

Working with tweets presents several challenges for natural language processing. The language used on social media sites is quite different from that used in other forums because it often contains words that are not found in a dictionary. One reason is that tweets have particular orthographic and typographical characteristics, such as letter or word duplication. Hence, before applying our approach, it was necessary to start with a data preprocessing stage to normalize the language used, remove noisy elements and generalize the vocabulary used to express sentiment. The aim of the preprocessing module is to bring tweets as close as possible to natural language by eliminating expressions that are not considered part of current usage, in order to minimize the noise in later stages. There are four main steps involved:

- URL links (such as “*http://url*”), hashtags links (such as “*#hashtag*”) and username links (such as “*@username*”) are replaced with “URL”, “HASHTAG” and “USERNAME” placeholders respectively.
- Replicated characters are removed to return the word to its normal form, such as *sweeeet* → *sweet*.
- Emoticons¹ are replaced by one of nine labels: *e_laugh*, *e_happy*, *e_surprise*, *e_positive_state*, *e_neutral_state*, *e_inexpressive*, *e_negative_state*, *e_sad* and *e_sick*. For instance, *:-(* is replaced with *e_sad*.
- Abbreviations² are replaced with their respective full written forms, such as *h8* → *hate*.

¹taken from the list available at <http://www.datagenetics.com/blog/october52012/index.html>

²taken from the lists available at <http://chatslang.com/terms/abbreviations>.

3.2 Lexical and syntactic analysis

In order to derive the syntactic context, each pre-processed social media message must first be broken into tokens and then into sentences. To then ensure that all inflected forms of a word are covered, lemmatization and *part-of-speech* (POS) tagging are performed using the *Freeling Tagger* (Atserias et al., 2006; Padró et al., 2012), or more specifically, its tagger implementation based on HMM (Brants, 2000). *Freeling* is a library that provides multiple language analysis services, including probabilistic prediction of categories of unknown words. POS tagging allows the identification of lexical items that can contribute to the correct recognition of sentiment in message. These items are namely adjectives, adverbs, verbs and nouns.

The resulting lemmatized and POS-annotated messages are fed to a parser that transforms the output of the tagger into a full parse tree. Finally, the tree is converted to dependencies, and the functions are annotated. The entire process is performed with *Freeling Parser* (Padró et al., 2012).

3.3 Sentiment lexicons

Sentiment lexicons, such as SOCAL (Taboada et al., 2011), AFINN (Nielsen et al., 2011) and *NRC Emotion and Hashtag Sentiment Lexicon* (Mohammad et al., 2013; Mohammad et al., 2013b), have been used in many systems for determining the semantic orientation of a phrase within a tweet or sentence. These lexicons contain English word lists sorted by lexical category, i.e. adjectives, verbs, nouns and adverbs. Each word is assigned a score of between -5 and 5.

However, these lexical resources are intrinsically non-contextual, so it is necessary to improve their coverage. To do this, we need to acquire new polarities of subjective words that are not present in generic dictionaries and adapt the scores of the other words using the data available. Consequently, we apply an automatic polarity expansion algorithm based on graphs (Cruz et al., 2011). The graph is generated from the syntactic dependencies provided by the *Freeling Parser*, considering only those involving verbs, nouns and adjectives. The starting point of the algorithm is a subset of negative and positive words, that are fed into the system as seed words. In this regard, we chose the most negative

and positive words in the SOCAL and AFINN lexicons, as they resulted to work quite well for the datasets provided, after carrying out different experiments through the training datasets. Then, we apply the iterative polarity expansion through the created graph, and the result is merged with the unique word list of SOCAL/AFINN lexicons, incorporating 5982 of new words. The next step is to include emoticon labels, together with their polarities, in the resulting sentiment lexicon.

3.4 Sentiment Detection

Once the lexical and syntactic analyses are complete, it is possible to estimate the polarity resulting from a message. In other words, its sentiment can be expressed by a real number, which can be later interpreted as positive, negative or neutral. This value is computed by using the lexical polarities of the words included in the text (provided by the sentiment lexicon we have generated), and subjecting the special parsing structure and its content to linguistic processing which is described below. Once these have been applied, the resulting sentiment is a propagation of the values of linguistic elements within the dependencies, from the leaves to the upper levels until the root is achieved (Caro, 2013). Then, it is classified according to defined interval.

3.4.1 Intensification treatment

Intensifiers and diminishers, such as “*very*” or “*a little*”, are usually adverbs that emphasize or attenuate the semantic orientation of the words or expressions they precede. Intensification is achieved by associating a positive or negative percentage, which implies a graduation depending on its type (Zhang et al., 2012). For instance, in “*very good*”, “*very*” enhances the positivity of “*good*”. Our system detects these structures and uses the parsing to identify the exact scope of the intensification whose semantic orientation will be altered. Superlative adjectives are also taken into account by assuming that they behave like a word accompanied by an intensifier. An example is “*greatest*”, where the superlative implies an intensification of the word “*great*”.

3.4.2 Negation treatment

Negation can be used to deny or reject statements. It is expressed grammatically through a variety of

negator words, such as “no”, “not”, “never” or “neither” (Zhang et al., 2012). In our case, it is first necessary to identify the dependencies in which any of the above negator forms are present to estimate the negation scope. Later, the semantic polarities of the words involved in the affected dependencies are modified using a negative factor.

3.4.3 Polarity conflict treatment

The mere application of polar lexicons, intensifiers, diminishers and negators on a syntactic structure is insufficient. That is, words cannot be considered individually (Moilanen et al., 2007). The meaning and polarity of “*unpleasant dream*” differs for example from those of “*wonderful dream*”. The first statement has a negative connotation while the second has a positive one. In both cases, the word “*dream*” is involved, and we could expect that, regardless of its accompanying terms, it should behave in a specific way, with certain polarity effects or expectations. However, the meaning changes significantly with the addition of “*unpleasant*” or “*wonderful*”. In these cases, our system is able to detect *polarity conflicts*, i.e., it recognizes when a positive adjective modifies a negative noun, or vice-versa, and subsequently reduces the polarity of the elements that cause the conflict.

3.4.4 Adversative/concessive clause treatment

There is a point in common between adversative and concessive subordinate clauses. 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, although it is not impossible. In both cases, one part of the sentence is in contrast with the other part. For this reason, in a context of sentiment analysis, we can assume that both constructions will restrict, exclude, amplify or diminish the sentiment reflected in the clauses. In this regard, it is necessary to clearly distinguish them. In an adversative structure, the argument introduced by items such as “*but*” or “*however*” is usually more important (Winterstein, 2012; Poria et al., 2014), while in a concessive structure, that introduced by items such as “*despite*” or “*in spite of*” is the least important (Rudolph, 1996).

Our approach is able to coherently estimate the sentiment of sentences that involve not only adver-

sative clauses, such as “*Bill Maher may be a little out there, but he does make some points*” (where the speaker is backing the view of Bill in general), but also concessive clauses such as “*Despite going off on Saturday, it looks like Ian Bennett could be fine for Wembley*” (where what appears to be really important is that Ian could go to Wembley).

4 Experimental Results

In this section we describe the experiments we conducted for both subtasks. These experiments were carried out using the datasets provided by the SemEval-2015 task organizers. These datasets are composed of texts extracted from Twitter (including sarcastic tweets), LiveJournal and phone text messages. The performance of each system is measured by means of the *F-score*, calculated as shown in Equation 1,

$$F\text{-score} = (F_P + F_N)/2 \quad (1)$$

where F_P stands for the *F-score* estimated only for positive results. In this case, this value is computed as shown in Equation 2, where P_P represents the precision and R_P the recall, both for positive results. The same is calculated for negative results, denominated F_N .

$$F_P = (2 * P_P * R_P)/(P_P + R_P) \quad (2)$$

Table 1 presents the overall score for subtasks A and B, in *Twitter2015 Test*, as well as *precision*, *recall* and *F-measure* values for positive (P), negative (N) and neutral (NEU) results.

Twitter 2015				
		Precision	Recall	F-score
Task A	P	87.33%	71.26%	78.48%
	N	80.02%	72.47%	76.06%
	NEU	10.25%	34.21%	15.78%
	Overall score: 77.27%			
Task B	P	72.13%	66.09%	68.98%
	N	41.57%	59.45%	48.93%
	NEU	61.72%	57.35%	59.45%
	Overall score: 58.95%			

Table 1: Results of our approach for subtasks A and B.

The approach previously described was applied on both datasets (A and B) in the same way using the

		TASK A					TASK B				
		LJ'14	SMS'13	T'13	T'14	TS'14	LJ'14	SMS'13	T'13	T'14	TS'14
Precision	P	84.65%	82.16%	91.41%	93.23%	88.75%	73.53%	56.57%	68.54%	76.27%	52.17%
	N	85.04%	92.64%	87.53%	78.86%	95.83%	74.52%	58.47%	54.56%	51.94%	87.50%
	NEU	31.62%	16.55%	7.76%	9.79%	10.00%	62.00%	81.64%	67.69%	60.03%	37.50%
Recall	P	76.06%	83.85%	81.38%	76.26%	86.59%	70.26%	71.75%	73.73%	70.06%	72.73%
	N	81.21%	79.80%	79.23%	71.63%	62.16%	64.47%	70.05%	59.73%	63.34%	35.00%
	NEU	51.39%	30.19%	29.38%	52.27%	40.00%	71.05%	67.44%	60.43%	62.18%	69.23%
F-score	P	80.13%	82.99%	86.11%	83.90%	87.65%	71.86%	63.26%	71.04%	73.04%	60.76%
	N	83.08%	85.74%	83.17%	75.07%	75.41%	69.14%	63.74%	57.03%	58.26%	50.00%
	NEU	39.15%	21.38%	12.27%	16.49%	16.00%	66.21%	73.87%	63.85%	61.09%	48.65%
Overall		81.61%	84.37%	84.64%	79.48%	81.53%	70.50%	63.50%	64.03%	65.65%	55.38%

Table 2: Performance of our approach on the progress test A and B.

generated sentiment lexicon and applying the propagation of the sentiment values within the dependencies. After performing several tests on the training datasets provided by organizers, we set the neutral sentiment intervals to $[-0.05, 0.05]$ for subtask A and $[-1.0, 1.0]$ for subtask B.

As can be seen, all our results are adjusted, so we can state that our system has no bias for one particular result, but performs quite well for all three types of answers. However, as can be seen in subtask A, the performance measures for neutral tweets are notably lower than those obtained for positive and negative tweets. This can be explained by the content of the dataset provided, which contained 1006 negative and 1896 positive tweets, but just 190 neutral tweets, which is an insufficient sample for producing reliable estimates on precision. The same problem happened for progress test A, where the proportions of tweets are similarly unbalanced.

Detailed scores for progress tests of subtasks A and B are shown in Table 2. In general, we can say that our system is quite stable, as it generates similar results for the different kinds of texts under evaluation. Also of note are the high percentages obtained for sarcastic tweets, which ranked in the first position in subtask A and in the tenth (test dataset) and sixth positions (progress dataset) in subtask B (as shown in Table 3).

5 Conclusions

This paper describes the participation of the GTI Research Group, AtlantTIC Centre, University of Vigo, in SemEval-2015 task 10: Sentiment Analysis in Twitter. We achieved our results using a *fully* unsupervised approach for message-level and phrase-

Test sets	Task A	Task B
<i>Twitter2015</i>	9/11	22/40
<i>Twitter2015Sarcasm</i>	-	10/40
<i>LiveJournal2014</i>	8/11	18/40
<i>SMS2013</i>	6/11	16/40
<i>Twitter2013</i>	5/11	25/40
<i>Twitter2014</i>	9/11	22/40
<i>Twitter2014Sarcasm</i>	1/11	6/40

Table 3: Position of our approach for each test and task, according to results provided on January 1, 2015.

level sentiment analysis of tweets. Table 3 shows our position in the ranking published for both subtasks A and B for all the different datasets evaluated.

Our approach comprises different processing stages, including the generation of sentiment lexicons, test preprocessing and the application of different methods for determining contextual polarity based on syntactical structure. This makes our approach robust in diverse contexts without the need for previous manual tagging of datasets. To the best of our knowledge, it is the only system presented in this competition whose sentiment analysis method does not require any supervision.

Acknowledgments

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

References

Atserias Jordi, Casas Bernardino, Comelles Elisabeth, González Meritxell, Padró Luis and Padró Muntsa.

2006. *FreeLing 1.3: Syntactic and semantic services in an open-source NLP library*. In *Proc. of the 5th International Conference on Language Resources and Evaluation*, p. 48–55. Genoa, Italy.
- Balahur Alexandra. 2013. *Sentiment Analysis in Social Media Texts*. In *Proc. of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, p. 120–128, ACL. Atlanta, Georgia.
- Brants Thorsten. 2000. *TnT: A Statistical Part-of-speech Tagger*. In *Proc. of the 6th Conference on Applied Natural Language Processing*, p. 224–231. Seattle, Washington.
- Caro Luigi and Grella Matteo. 2013. *Sentiment analysis via dependency parsing*. In *Computer Standards & Interfaces Journal*, p. 442–453, volume 35 (5), New York.
- Cruz Fermín L., Troyano José A., Ortega F. Javier and Enríquez Fernando. 2011. *Automatic Expansion of Feature-level Opinion Lexicons*. In *Proc. of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, p. 125–131. ACL. Portland.
- Jansen Bernard J., Zhang Mimi, Sobel Kate and Chowdury Abdur. 2009. *Twitter power: Tweets as electronic word of mouth*. *Journal of the American Society for Information Science and Technology*, p.2169–2188, volume 60 (1), New York.
- Hu Minqing and Liu Bing. 2004. *Mining and summarizing customer reviews*. In *Proc. of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, p. 168–177. ACM Press.
- Liu Bing. 2012. *Sentiment Analysis and Opinion Mining*. In *Synthesis Digital Library of Engineering and Computer Science*, Morgan & Claypool Publisher.
- Mohammad Saif M. and Turney Peter D. 2013. *Crowdsourcing a Word-Emotion Association Lexicon*. *Journal of Computational Intelligence*, p. 436–465, volume 29 (3).
- Mohammad Saif M., Kiritchenko Svetlana and Zhu Xiaodan. 2013. *NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets*. In *Proc. of the 7th International workshop on Semantic Evaluation Exercises (SemEval-2013)*, p. 282–290. Vienna, Austria.
- Moilanen Karo and Pulman Stephen. 2007. *Sentiment Composition*. In *Proc. of Recent Advances in Natural Language Processing (RANLP)*, p.378–382, Borovets, Bulgaria.
- Nakov Preslav, Rosenthal Sara, Kozareva Zornitsa, Stoyanov Veselin, Ritter Alan and Wilson Theresa. 2013. *SemEval-2013 Task 2: Sentiment Analysis in Twitter*. In *Proc. of the 7th International Workshop on Semantic Evaluation*, p. 312–320. ACL. Atlanta, Georgia, USA.
- Nielsen Finn Årup. 2011. *A new ANEW: Evaluation of a word list for sentiment analysis in microblogs*. In *Proc. of the ESWC2011 Workshop on Making Sense of Microposts*, p. 93–98.
- Padró Lluís and Stanilovsky Evgeny. 2012. *FreeLing 3.0: Towards Wider Multilinguality*. In *Proc. of the 8th International Conference on Language Resources and Evaluation*, p.23–25, Istanbul, Turkey.
- Pak Alexander and Paroubek Patrick. 2010. *Twitter as a Corpus for Sentiment Analysis and Opinion Mining*. In *Proc. of the 7th International Conference on Language Resources and Evaluation (LREC'10)*, p.19–21, Valletta, Malta.
- Poria Soujanya, Cambria Erik, Winterstein Grégoire and Huang Guang-Bin. 2014. *Sentic patterns: Dependency-based rules for concept-level sentiment analysis*. *Journal of Knowledge-Based Systems*, p.45–63, volume 69 (1).
- Quinn Kevin M., Monroe Burt L., Colaresi Michael, Crespin Michael H. and Radev Dragomir R. 2010. *How to Analyze Political Attention with Minimal Assumptions and Costs*. *American Journal of Political Science*, p.209–228, volume 54 (1).
- Rosenthal Sara, Nakov Preslav, Kiritchenko Svetlana, Mohammad Saif M., Ritter Alan and Stoyanov Veselin. 2015. *SemEval-2015 Task 10: Sentiment Analysis in Twitter*. In *Proc. of the 9th International Workshop on Semantic Evaluation*, ACL. Denver, Colorado.
- Rosenthal Sara, Ritter Alan, Nakov Preslav and Stoyanov Veselin. 2014. *SemEval-2014 Task 9: Sentiment Analysis in Twitter*. In *Proc. of the 8th International Workshop on Semantic Evaluation*, p.73–80. ACL. Dublin, Ireland.
- Rudolph Elisabeth. 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. Berlin.
- Taboada Maite, Brooke Julian, Tofiloski Milan, Voll Kimberly and Stede Manfred. 2011. *Lexicon-based Methods for Sentiment Analysis*. *Journal of the Computational Linguistics*, p.267–307, volume 35 (2), MIT Press. Cambridge, MA, USA.
- Winterstein Grégoire. 2012. *What but-sentences argue for: a modern argumentative analysis of but*. *Lingua* 122 (15), p. 1864–1885.
- Zhang Lei, Ferrari Silvia and Enjalbert Patrice. 2012. *Opinion Analysis: the Effect of Negation on Polarity and Intensity*. In *Proc. of KONVENS 2012 (PATHOS 2012 Workshop)*, p. 282–290. Vienna, Austria.