FUZZY SENTIMENT ANALYSIS USING SPANISH TWEETS

Pilar Rey del Castillo

Instituto de Estudios Fiscales, Madrid, España
mpilar.rey@ief.minhap.es

Abstract. *Opinion mining and sentiment analysis are two topics with growing interest in artificial intelligence. Last years' research on these areas has evolved in increasing complexity and sophistication with tasks such as opinion retrieval, sentiment extraction, classification and summarization being carried out using text mining and natural language processing techniques. Possible strategies go from concept-based approaches (using web ontologies or semantic networks) and data mining (passing a labeled training corpus of texts to a machine learning algorithm), to the most crude keyword spotting methods. These last approaches use sentiment lexicons defining the prior polarity of particular terms, and aggregate these scores to determine the text's polarity. Recent literature has taken down the text analysis to sentence-level, as different opinions and sentiments towards the same topic can be present across the span of a document.*

Twitter is a popular social media platform where the number of characters in a message is limited to 140. These messages, called tweets, can convey opinions about a range of topics. On the other hand, fuzzy sets theory is a valuable tool in situations dealing with imprecision and ambiguous or incomplete information. Its use can be of interest to tackle the assignment of polarity to small pieces of text such as the tweets. This paper describes a sentiment analysis performed to Spanish tweets following a fuzzy sets approach and using an existing Spanish lexicon. A profiling method to evaluate the quality of the resulting indicator is proposed later.

Keywords: sentiment analysis, fuzzy sets, tweets.

1. Introduction

Microblogs are contemporary and wealthy sources of data that can be studied to obtain valuable knowledge about the opinions and sentiments of people in general. In particular, the 140-character limit for Twitter messages forces users to extract the content and encode it in a small number of sentences and words, what makes them especially appropriate for completing sentiment analysis tasks. The reason is that some of the problems to cope with − such as the appearance in the text of different and opposing sentiments− are simplified due to the limitation in the text size. Analysis can be performed at the sentence-level, assuming that tweets express opinions about one single topic.

The study of Twitter and others microblogs is frequently performed using a variety of techniques from the natural language processing and text mining fields [1, 2]. Last years' research on these areas has evolved in increasing complexity and sophistication. The current strategies can be grouped into four main categories, from the most simple to the most complicated: keyword spotting, lexical affinity, statistical methods, and concept-based techniques [3]. The methods of keyword spotting classify texts by sentiment categories based on the presence of unambiguous sentiment related words such as happy, sad and unlucky. The lexical affinity approach, apart from detecting obvious sentiment words, assigns other words a certain "affinity" to particular emotions, generally in probabilistic terms. The statistical methods use a corpus of sentiment annotated texts to learn the score of certain keywords, taking also into account the punctuation and word co-occurrence frequencies. Finally, the concept-based approaches help the system catch the sentiment information using semantic networks or web ontologies [4].

Although being the most naïve approach, the procedures using keyword spotting are popular within sentiment analysis today. The most difficult part of the analysis problem in this case becomes building a sentiment lexicon that includes the sentiment polarity for words. The presence of the particular words rather than its frequency of appearance forms an effective basis for polarity classification of texts [5]. In the case of Twitter, Gupta et al. [6] observed that the most frequently used keywords formed less than 10 percent of traditional dictionary appraisal words, what makes an indication that a small set of words may constitute a determining factor for sentiment analysis.

The majority of the sentiment lexical resources have been developed for the English language and are not directly usable on other languages. One of these resources is the lexicon Affective Norms for English Words (ANEW) [7] which provides a set of normative evaluative ratings for 1034 words in the English language. The words in ANEW have been used in a wide range of emotion and attitudes-related studies. Its Spanish adaptation –presented in 2007 [8]– represents a valuable instrument which will be used in this study for the assessment of the tweets. This Spanish version includes, for each word, a set of measures evaluating several of the dimensions along which the human emotions can be arranged. Using the fuzzy sets theory to deal with the inherent imprecision and ambiguity of the language, some of these measures will be used to extract sentiment information from the tweets. The scores which will be computed take into account the fuzziness of the polarity intensity. The final aim is obtaining information of the trend of the sentiment during the observation period for the whole Spanish population. For this purpose, a sentiment indicator that summarises the individual scores in a global measure for each day is computed as a final step.

The remainder of this paper is organised as follows: the next section describes the proposed model for identifying the polarity on the tweets from the Spanish lexicon and how the scores are combined to compute the global sentiment indicator; section 3 then sets out the results obtained with the tweets of a period of about 10 months; and finally, a number of remarks and conclusions are presented in Section 4.

2. Model to perform sentiment analysis from tweets

There has been in recent years extensive research in the domain of natural language processing to estimate the opinion polarity and strength in texts. Most of the research in the area of microblogs such as Twitter has been focused on public opinions and reviews on specific products for deriving useful knowledge to improve sales performance. The aim of this study is to obtain a summary indicator with no specific topic in mind. Because of this, the proposed method does not address the full possible complexity in natural language processing and does not enter into more refined analysis. It just tries to obtain a crude idea about the average mood or sentiment in the country and its evolution across time to relate it later with political, economic or social events.

The description of the proposed model is split into three parts: assignment of sentiment scores to the words in the lexicon; assignment of fuzzy sentiment polarities to tweets; and computation of a summary sentiment indicator.

2.1 Assignment of sentiment scores to the words in the lexicon

The ANEW Spanish adaptation [8] was developed at the University of Santiago de Compostela, Spain. The theoretical background for the construction of this lexicon conceptualizes emotion as having three basic underlying dimensions: valence (ranging from *pleasant* to *unpleasant*), arousal (ranging from *calm* to *excited*), and dominance (ranging from *in control* to *out of control*). The ratings of each word in these dimensions are based on 720 psychology students' assessments of the translation into Spanish of the 1034 words included in the original ANEW. The participants could evaluate each word in a 9-point rating scale for each dimension, using the same method as the original study. The lexicon shows the mean values and the standard deviation of the students' ratings. The results when computing correlations between the American and the Spanish populations in the assessments of the words suggested that the ANEW words are understood in a similar way by both Americans and Spaniards.

There are two dimensions in ANEW which are somehow related to the notion of sentiment. The valence reveals the level of pleasantness associated with the word and can be linked to the direction or quality of the sentiment from completely negative to completely positive, corresponding to the 1 and 9 values respectively. The arousal dimension can be connected with the amount and intensity of the emotion or sentiment, ranging from the minimum 1 to the maximum 9.

The sentiment values which this paper proposes combine and encompass the two dimensions –quality and intensity of the sentiment– in a unique score for each word. As a first step, the arousals are lineally transformed to make them take values between 0 and 1, from minimum to maximum of intensity. Then, the valences are translated to the [-4, 4] interval to reflect the opposition negative/positive by subtracting 5 to the original values. An evaluation of the strength of the sentiment for each word is obtained by multiplying these new valences by their previously computed weights (intensity). Hence, the weight modifies the value of the valence in the appropriate direction, from minimum to whole intensity. The last step is to translate these measures to the interval [0, 1] to obtain the final sentiment value. The expression of the final computed score for each word is:

$$w \,\epsilon\, \text{ANEW} \Rightarrow s(w) = \frac{\left[\left(\frac{arousal(w) - 1}{8}\right).(valence(w) - 5)\right] + 4}{8}$$

where *arousal(w)* and *valence(w)* are the corresponding mean values for the word *w* appearing in ANEW. In this way, words expressing positive sentiments are assigned a higher value whereas words with a negative connotation are assigned lower values.

2.2 Assignment of fuzzy sentiment polarities to tweets

The theory of fuzzy sets to catch the sentiment of a message seems truly appropriate since it has been especially developed to deal with imprecision and incomplete information. For this reason, there are a variety of experiences using fuzzy sets and dealing with different aspects of the problem. For example, Gupta et al. [6] present a fuzzy approximation technique based on a set of rules to evaluate microblogs opinion intensity or degree of polarity of movie reviews; Zhao et al. [9] build a fuzzy support vector machine assigning lower weights to the text samples which make small contributions to classification; Mukkamala et al. [10] present a formal model based on fuzzy sets theory for social data; and Wang et al.[11], based on a sentiment lexicon consisting of a positive key sentiment words list and a negative key sentiment words list, construct a fuzzy classifier to identify polarity of Chinese sentiment words. To the best of our knowledge, the approach presented in this paper is new and may be of independent interest.

Our aim in this step is to compute a score for each tweet reflecting its degree of membership in the fuzzy category of sentiment. Following the theory of fuzzy sets [12], what we are looking for is a sentiment function assigning to each tweet a polarity ranging between zero and one. In order to analyze a tweet to assign its degree of polarity, it is firstly pre-processed by removing special characters. The identification of the tweet sentiment polarity is next simply addressed through the detection of the presence of ANEW words. Tweets not including any of these words are discarded. To assign the overall fuzzy sentiment score to the non-discarded tweets, the individual scores of its sentiment-bearing words should be combined in a certain way. The most common procedure of combining the intensity values of the phrases/words in texts is computing its average [6, 13] and this is the method chosen in our approach. Thus, the expression of the fuzzy sentiment polarity for a non-discarded tweet *m* is:

$$f(m) = \frac{\sum_{i=1}^{r} s(w_i)}{r}$$

where w_i , $i = 1, ..., r$ are the words from ANEW detected in *m*.

2.3 Computation of a summary sentiment indicator

A summary indicator is calculated to obtain information about the trend of the sentiment during the observation period for the whole Spanish population. It is computed in the simplest way as the mean value of the tweets available for the selected period:

$$I_t = \frac{\sum_{k=1}^{n} f(m_k)}{n}$$

where m_k , $k = 1, ..., n$ are the tweets in period t. In this way, indicators can be computed for different period durations (year, month, day, hour ...).

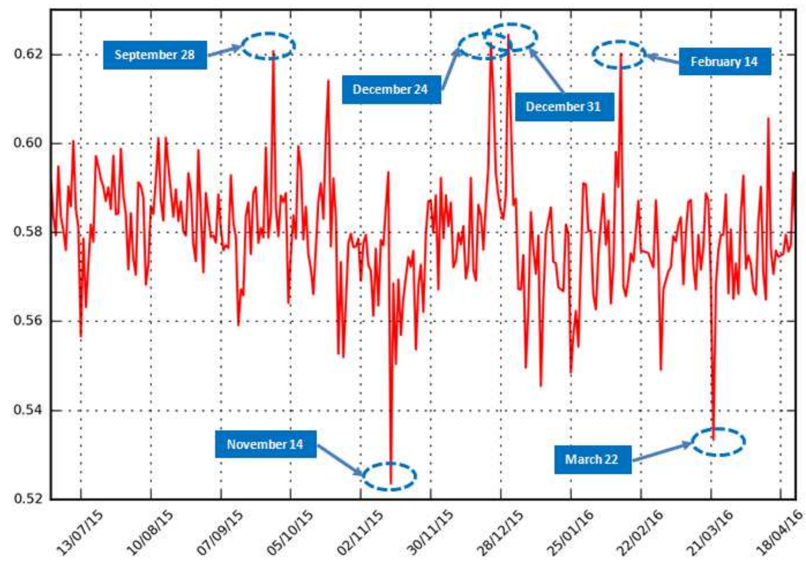
3. Results using Spanish tweets

Our aim is not to obtain the most accurate information about the sentiment towards a specific target such as a movie, a person or a product but just a raw idea of the average general sentiment in the country or, in other words, the general positiveness. A sampling framework has been chosen for this purpose. Thus, a random sample of geographically referenced tweets is gathered each day at 14:00h and during 15 minutes using the Twitter streaming API. The location of the provided tweets is inside a rectangle covering the Iberian Peninsula and the Balearic islands, and their language is Spanish. The procedure supplies around 1000 tweets daily which is considered an appropriate sample size to start with. Following the procedures described in the previous section, fuzzy sentiment indicators are calculated daily, monthly, etc. Figure 1 shows the computed daily indicators for the period between July 2015 and April 2016.

The concept whose evolution we are trying to measure through these indicators –the average collective mood– is an abstract one, lacking from a precise definition and, for this reason, the assessment of its quality and accuracy results a difficult task. Something that can be evaluated firstly is whether the values obtained make sense, e. g., whether meaning can be extracted from the highest and lowest values in the daily time series. In this regard, from Figure 1:

- September 28, 2015: one day after Catalan parliamentary elections.
- November 14, 2015: one day after Paris terrorist attacks.
- December 24, 2015: celebration of Christmas Eve.
- December 31, 2015: End of the Year celebration.
- February 14, 2016: celebration of Saint Valentine’s Day.
- March 22, 2016: Brussels terrorist attacks.

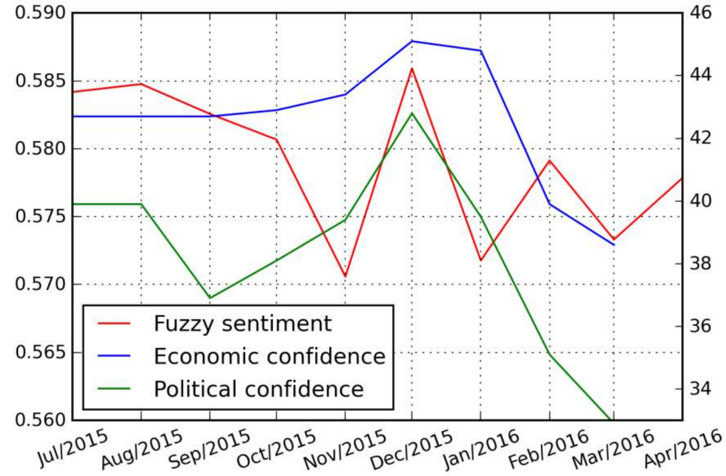
Fig 1. Fuzzy sentiment indicator by day



Some of the atypical values are obtained one day after a particular event because, as was mentioned, the tweets are collected at 14:00 hours. On the other hand, it is clear that any kind of celebration can produce an improvement in the general mood in a day, and this is what is seen for all festivity day cases. Likewise, tragic events such as the Paris and Brussels terrorist attacks produce a significant worsening in the average feeling throughout the country.

Another assessment on the quality of the indicators can be obtained by comparing them with others measuring the same or similar concept. There are not known indicators at daily granularity level. On the other hand, the Spanish Centre for Sociological Research publishes the monthly Barometer Indicators [14] from 1996. Some confidence indicators gathering the judgments on the current economic and political situation from an opinion poll are included in this publication. The phenomena monitored by these indicators are somehow related to or may have an influence on the average mood in the population and, thus, can be compared with the fuzzy indicator for the common months available. In Figure 2 the fuzzy sentiment indicator is shown jointly with the economic and political confidence indicators in the second axis. From the time series analysis perspective, a possible relation cannot be established for such a short period of time: there is not even one year of observations. More detailed analysis may be made when having more observations from a larger period. Besides, the attitude surveys for computing the Barometer Indicators are collected just during one week in the month while the data collection for the fuzzy sentiment indicators is made throughout the whole month.

Fig 2. Monthly indicators



A last evaluation of the fuzzy indicator can be made by deriving background characteristics from the data itself, what is sometimes referred to as *profiling* according to Buelens et al. [15]. This approach will be introduced in what follows. For this purpose, the geographical area is split into points whose latitude is less, and whose latitude is higher or equal to 40 degrees North, respectively. The number 40 has been arbitrary chosen for convenience, because it divides the sample of tweets into two parts with similar sizes for each day.

Fig 3. Partition of Spain by the 40º parallel of latitude

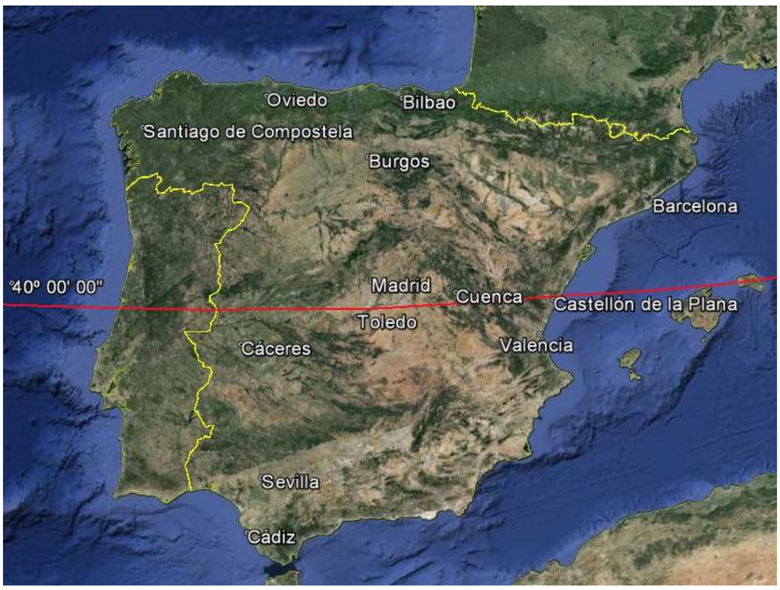


Figure 3 shows a map including the 40th parallel North and the location of some of the cities. The fuzzy sentiment analysis can be performed in each one of these geographical areas and the resulting indicators are shown in Figure 4 and 5.

Fig 4. Fuzzy sentiment indicators by day

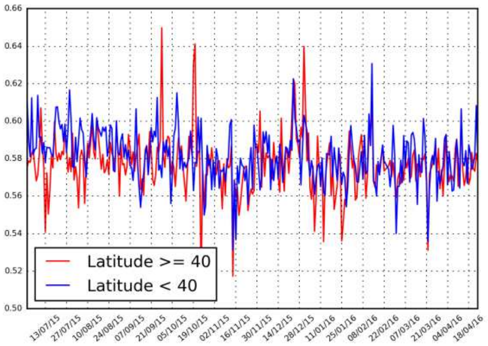
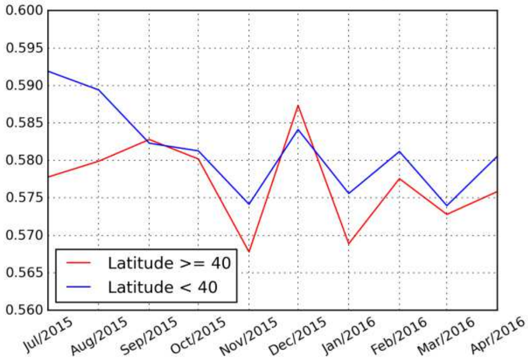


Fig 5. Fuzzy sentiment indicators by month



In Figure 4 it can be seen that the indicators have the same type of evolution. Something that can also be appreciated is that the fuzzy sentiment is slightly lower in the area with higher latitude, with only a few exceptions. This fact is confirmed in Figure 5 for the monthly averages and, as a significant pattern, it can be said that the collective mood is more positive in the less than 40 degrees North area. The result is in accordance with the general understanding that people from the South in Spain are more cheerful and have more positive feelings.

4. Final remarks

The previous sections describe a procedure to compute fuzzy sentiment indicators describing the evolution of the average mood or sentiment using tweets produced in Spain and written in Spanish. The main findings can be summarised in the following comments and remarks:

1. The ANEW Spanish lexicon offers a useful tool to perform sentiment analysis with text messages from microblogs such as Twitter.
2. The presented procedure provides with indicators of the average mood in the country at an unknown until now time granularity level.
3. There are not enough observations for evaluating the quality of the computed fuzzy sentiment indicators. From the time series perspective, the noise component of the indicators [16] could probably be reduced by increasing the number of collected tweets each day.
4. A possible use of the computed indicators is in assessing the impact of specific social, political or economic events –by means of intervention analysis [17]– when the number of data points is large enough.

References

[1] Pang, B. and Lee, l. (2008), Opinion Mining and Sentiment Analysis, Foundations and Trends in Information Retrieval, 2 (1–2), pp. 1–135.

[2] Wagner J., Arora, P., Cortes, S., Barman, U., Bogdanova, D., Foster, J. and Tounsi, L. (2014), Aspect-based Polarity Classification for SemEval Task 4, Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pp. 223 - 229, Dublin, Ireland, August 23-24.

[3] Cambria, E., Schuller, B., Xia, Y. and Havasi, C. (2013), New Avenues in Opinion Mining and Sentiment Analysis, Intelligent Systems IEEE, 28 (2), pp. 15-20, DOI: 10.1109/MIS.2013.30.

[4] Cambria, E. (2013), An Introduction to Concept-level Sentiment Analysis. In: Castro, F., Gelbukh, A., González, M. (eds.) MICAI 2013, Part II. LNCS, 8266, pp. 478–483, Springer, Heidelberg.

[5] Pang, B., Lee, L. and Vaithyanathan, S. (2002), Thumbs Up? Sentiment Classification Using Machine Learning Techniques, in Proceedings of the Conference in Empirical Methods in Natutal Language Processing (EMNLP), pp. 79-86.

[6] Gupta, N., Abhinav, K. R. and Annappa (2013), Fuzzy Sentiment Analysis on Microblogs for Movie Revenue Prediction, in IEEE International Conference on Emerging Trends in Communication, Control, Signal Processing & Computing Applications (C2SPCA), DOI: 10.1109/CS2PA.2013.6749350.

[7] Bradley, M. M. and Lang, P. J. (1999), Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings, Gainesville, FL, Center for Research in Psychophysiology, University of Florida.

[8] Redondo, J., Fraga, I., Padrón, I. and Comesaña, M. (2007), The Spanish Adaptation of ANEW (Affective Norms for English Words), Behavior Research Methods, 39 (3), pp. 600-605.

[9] Zhao, C., Wang, S. and Li, D. (2014), Fuzzy Sentiment Membership Determining for Sentiment Classification, 2014 IEEE International Conference on Data Mining Workshops, Shenzhen, China, December 14, pp.1191-1198, DOI: 10.1109/ICDMW.2014.137.

[10] Mukkamala, R. R., Hussain, A., and Vatrapu, R. (2014), Fuzzy-Set Based Sentiment Analysis of Big Social Data, IEEE 18th International Enterprise Distributed Object Computing Conference, Ulm, Germany, 1-5 Sept., DOI: 10.1109/EDOC.2014.19.

[11] Wang, B., Huang, Y., Wu, X., and Ki, X. (2015), A Fuzzy Computing Model for Identifying Polarity of Chinese Sentiment Words, Computational Intelligence and Neuroscience, 23, pp. 1-13, DOI:10.3115/1220355.1220555.

[12] Zadeh, L. (1965), Fuzzy sets, Information and Control, 8, pp. 338–353.

[13] Hogenboom, A., Bal, D., Frasincar, F., Bal, M., de Jong, F.M.G. and Kaymak, U. (2013), Exploiting Emoticons in Sentiment Analysis. In: Proceedings of the 28th Annual ACM Symposium on Applied Computing, SAC 2013, 18-22 Mar 2013, Lisbon, Portugal. pp. 703-710.

[14] http://www.cis.es/cis/opencms/EN/11_barometros/indicadores.html

[15] Buelens, B., Daas, P., Burger, J., Puts, M., and Van den Brakel, J. (2014), Selectivity of Big Data, Discussion paper 201411, Statistics Netherlands.

[16] Tiao, G. C. and Hillmer, S. C. (1978), Some Consideration of Decomposition of a Time Series, Biometrika, 65, pp. 497-502.

[17] Tiao, G. C. (1985), Autoregressive Moving Average Models, Intervention Problems and Outlier Detection in Time Series, in E. J. Hannan, P. R. Krishnaiah and M. M. Rao, Handbook of Statistics, vol. 5: Time Series in the Time Domain, pp. 85-118, Amsterdam, North-Holland Ed.