

Exploiting Data of the Twitter Social Network Using Sentiment Analysis

David Gonzalez-Marron^(✉), David Mejia-Guzman, and Angelica Enciso-Gonzalez

Instituto Tecnológico de Pachuca, Carretera México-Pachuca Km 81.5,
42080 Pachuca, Hidalgo, Mexico
{dgonzalez, a_enciso}@itpachuca.edu.mx,
davidmejia.05@hotmail.com

Abstract. Social Networks nowadays are producing an enormous quantity of data, this data transformed into information could be useful for the decision support systems. A new emerging technology denominated as Sentiment Analysis or Opinion Meaning extracts the opinion or sentiment of a particular text. The Twitter social network is a source of valuable information in simple text and appropriated to use this technology. In this paper is described the process used to select the most suitable algorithms to analyze tweets for particular words written in Spanish, also the results obtained by every algorithm are reported.

Keywords: Future internet · Social networks · Sentiment analysis · NoSQL databases

1 Introduction

Recently, the growth rate in the volume of data produced has caused the generation of technology to manage this huge quantity of data, most of this data is stored in the internet, and every year this tendency of growth continues.

Social networks are a good source of information because users tend to use the web to establish communication among people who share similar interests. Many enterprises are interested in knowing what their customers think about the products or services received, and a good source of information is the exploitation of data stored in social networks.

According to [1] a social network is defined as, a social structure conformed by organizations or individuals with similar interests that permits the creation of a public or semipublic profile in a delimited system, enabling the creation of a user's list to share connections. It is possible for developers the extraction of data contained in several social networks, as Facebook, Twitter, Google plus and others, this paper is oriented to Twitter because the data produced is a simple chain of up to 140 characters by tweet.

2 Sentiment Analysis

Also known as “opinion mining” consists in the extraction of information produced by users (post, blog, etc.) for study or classification [2]. The process for this sentiment

analysis is constituted by the next parts [3]: (a) Tokenization, (b) Word correction, (c) Words grammatical tagging, (d) Word categorization and identification, (e) Morphological analysis of text, and (f) Text properties extraction.

In social media the immediacy and spontaneity of opinions using an automatic sentiment analysis tool permits the translation of human emotions into data, this data can preserve the emotional content of the message and can be used by the taking decision group to improve services or product's qualities. The most valuable features in sentiment analysis are: Polarity, intensity and subjectivity. The polarity is considering if the comment is (positive, negative or neutral), the intensity is referring to how strong the emotion is expressed, and the subjectivity is about whether the source is objective, biased or subjective.

A brief description describing what polarity must consider to evaluate texts is realized below.

Polarity: [4]

Positive words: Words that express happiness or gratitude, usually are adjectives or verbs

Verbs examples: Love, adore, like, enjoy

Adjectives examples: Beautiful, pretty, well, fascinating, etc.

Negative words: Words expressing anger or sadness, regularly are adjectives or verbs

Verbs examples: Hate, annoy, upset, angry, etc.

Adjectives examples: Ugly, awful, bad, boring, etc.

Neutral words: Words that not express negative or positive emotions, usually are nouns, articles or pronouns.

Nouns examples: School, house, dog, football, etc.

Articles examples: The, a, an, etc.

Pronouns examples: I, you, them

Deniers, Cancellers or inversors: Words contained in a phrase or sentence that change its meaning

Negation adverbs: No, never, neither, etc.

Adversative conjunctions: but, yet, nevertheless, etc.

Twitter was selected by several reasons, below are described the most important reasons why we selected this social network:

- Because it is a microblogging free social network used to publish, share and interchange information in real time.
- It is used by different people to express their opinion.
- It contains a great number of posts related to different topics.
- The members of this social network represent different parts of a society.
- People from different regions and countries, interact in this networks.
- People post information in several languages, but it is of our particular interest the messages produced in Spanish.

3 Sentiment Analysis Algorithms

The process defined to realize sentiment analysis [4], was to select some resources and APIs, prioritizing those who could analyze Spanish sentences and that could interact with the python language, as is described in Table 1.

Table 1. Analyzing characteristics of the sentiment algorithms

| Sentiment Analysis APIs and Resources | Resource type | Language | Compatibility with Python 2.7 | Selected |
|---|-------------------|---------------------|-------------------------------|----------|
| NLTK | Library of Python | English | No | No |
| Diccionario marcado con emociones y ponderado para el español | Dictionary | Spanish | No apply | No |
| Sentiment140 | API | English and Spanish | Yes | Yes |
| TextBlob | API | English | Yes | No |
| Sentiment Analysis of Meaning Cloud | API | Spanish | Yes | Yes |
| Bitext's API | API | English and Spanish | Yes | Yes |

In order to validate the best performance of Sentiment Algorithms in Spanish, It was developed a set of 100 sentences in Spanish that allowed us to verify the classification efficiency of the three selected twitter API's, in Table 2, a reduced number of six sentences representing typical expected classification are shown, it was detailed the classification process to obtain the sentence type.

Table 2. Reduced set of sentences to validate sentiment analysis in Spanish

| Type sentence | Word type conforming the sentence | Example sentence | Sentence classified words | | | |
|-------------------|-------------------------------------|--|---------------------------|---------|--------------------|-------------------|
| | | | Pos. P. | Neg. P. | Neu. P. | N.I.A. |
| Simple positive | Pos. P + Neu. P. | Ayer fue un excelente día | Ayer | | Ayer, fue, un, día | |
| Positive with NIA | Neg. P. + Neu. P. + N.I.A. | No estoy enojado | | Enojado | Estoy | No |
| Simple negative | Neg. P + Neu. P. | Odio ir a la escuela | | Odio | Ir, a, la, escuela | |
| Negative with NIA | Pos. P. + Neu. P. + N.I.A. | No me siento bien | bien | | Me, siento | No |
| Simple neutral | Neu. P. | Prueba 1 | | | Prueba, 1 | |
| Neutral complex | Pos. P. + Neg. P. + Neu. P. + N.I.A | No estoy triste pero tampoco estoy feliz | feliz | triste | Estoy, estoy | No, pero, tampoco |

Abbreviations description.

Pos. P.: words with positive polarity **Neg. P.:** words with negative polarity **Neu. P.:** words with neutral polarity
N.I.A.: words with the ability to deny, cancel, or change the meaning of one sentence

In Table 3 is described the result obtained by each API selected with the set of 100 Spanish sentences. Once selected the meaning cloud algorithm as the best algorithm to classify sentences in twitter, a set of 1000 text posts on Twitter were analyzed automatically with the algorithm and verified manually in order to analyze the efficiency obtained as can be in Figs. 1 and 2.

Table 3. Performance obtained with selected Twitter sentiment analysis tools

| Sentences type | Sentences to classify | Sentiment140 | | | Sentiment Analysis of Meaning Cloud | | | Bitext's API | | |
|---|--|--------------|----------|---------|-------------------------------------|----------|---------|--------------|----------|---------|
| | | Result | | Success | Result | | Success | Result | | Success |
| Type sentence | Example sentence | Expected | Obtained | Success | Expected | Obtained | Success | Expected | Obtained | Success |
| Simple positive | Ayer fue un excelente día | Positive | Neutral | No | Positive | Positive | Yes | Positive | Positive | Yes |
| Positive with NIA | No estoy enojado | Positive | Neutral | No | Positive | Positive | Yes | Positive | Positive | Yes |
| Simple negative | Odio ir a la escuela | Negative | Neutral | No | Negative | Negative | Yes | Negative | Negative | Yes |
| Negative with NIA | No me siento bien | Negative | Negative | Yes | Negative | Negative | Yes | Negative | Negative | Yes |
| Simple neutral | Prueba 1 | Neutral | Neutral | Yes | Neutral | Neutral | Yes | Neutral | Neutral | Yes |
| Neutral complex | No estoy triste pero tampoco estoy feliz | Neutral | Neutral | Yes | Neutral | Neutral | Yes | Neutral | Neutral | Yes |
| Percentage of success for 100 sentences: | | 50% | | | 100% | | | 92% | | |

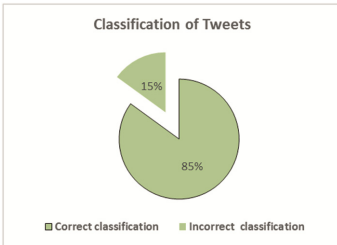


Fig. 1. Tweets classification for sentences

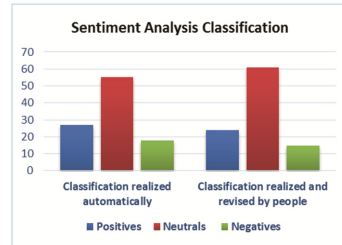


Fig. 2. Automatic and manual classification

4 Conclusions

Tweets storage increase at a rate of about 7000 tweets per second [5], and because of this, a strategy for processing and storing information was defined [4], in this work were analyzed only six APIs and resources for sentiment analysis, although. There are many API tools developed for realizing sentiment analysis mainly in English language, the results obtained with the API selected “Sentiment Analysis of Meaning Cloud” were very good, 85% of the tweets analyzed were correctly classified, and the detected errors were mainly caused by slang use, sarcasms and regionalisms in Spanish. The Storage realized in a NoSQL database allowed us to improve the data storage for future analysis, and improves the speed in updates and consults.

References

1. Flores, J., Moran, J., Rodriguez, V.: Las Redes Sociales. USMP (2009). <http://www.usmp.edu.pe/publicaciones/boletin/fia/info69/sociales.pdf>
2. Análisis de sentimiento: capturando la emoción. 27 de mayo de 2015, de www.daedalus.es Sitio web (2012). <http://www.daedalus.es/blog/es/analisis-de-sentimiento-capturando-la-emocion>
3. Valverde J.C.: Sistema de extracción de entidades y análisis de opiniones en contenidos Web generados por usuarios. Trabajo de fin de grado. Universidad Autónoma de Madrid. Madrid, España (2013)
4. Gonzalez, D., Mejia, D., Mendoza, M., Enciso, A.: Desarrollo de una aplicación para la extracción de información de la red social Twitter y análisis de sentimientos de Tweets, Revista de Tecnologías de la Información, vol. 2, no. 2, ECFORFAN (2015)
5. Internetlivestats (2016). <http://www.internetlivestats.com>