A Comparison of SVM Versus Naive-Bayes Techniques for Sentiment Analysis in Tweets: A Case Study with the 2013 FIFA Confederations Cup

André L. F. Alves
Federal University of Campina Grande, State University of Paraíba
Campina Grande, Brazil
andre@uepb.edu.br

Cláudio de S. Baptista
Federal University of Campina Grande
Campina Grande, Brazil
baptista@dsc.ufcg.edu.br

Anderson A. Firmino
Federal University of Campina Grande
Campina Grande, Brazil
anderson.firmino@ccc.ufcg.edu.br

Maxwell G. de Oliveira
Federal University of Campina Grande
Campina Grande, Brazil
maxwell@ufcg.edu.br

Anselmo C. de Paiva
Federal University of Maranhão
Maranhão, Brazil
paiva@deinf.ufma.br

ABSTRACT

The widespread of social communication media on the Web has made available a large volume of opinionated textual data stored in digital format. These media constitute a rich source for sentiment analysis and understanding of the opinions spontaneously expressed. Traditional techniques for sentiment analysis are based on POS Tagger. Considering the Portuguese language, the use of POS Tagging ends up being too costly, due to the complex grammatical structure of this language. Faced with this problem, a case study is carried out in order to compare two techniques for sentiment analysis: a SVM versus Naive-Bayes classifiers. Our study focused on tweets written in Portuguese during the 2013 FIFA Confederations Cup, although our technique could be applied to any other language. The achieved results indicated that the SVM technique surpassed the Naive-Bayes one, concerning performance issues.

Categories and Subject Descriptors

H.2.8 [**DATABASE MANAGEMENT**]: Database Applications—*Data Mining*; H.3 [**Information Storage and Retrieval**]: Miscellaneous; I.7 [**Document and Text Processing**]: Miscellaneous

Keywords

Analysis of Sentiment; Support Vector Machine (SVM); Naive-Bayes; Natural Language Processing (NLP)

WebMedia'14, November 18–21, 2014, João Pessoa, Brazil.

http://dx.doi.org/10.1145/2664551.2664561.

1. INTRODUCTION

The Web 2.0 has lead to a widespread of non-structured information by means of blogs, discussion forums, online product evaluation sites, microblogs and several social networks. This fact has brought out new challenges and opportunities in the information retrieval area [6].

It is important to notice that the opinions about several themes expressed by Web users are made in a spontaneous manner and in real time [4]. In this context, sentiment analysis has emerged providing the possibility of capturing opinions of the general public, in an automated way, about some theme. This research field has an increasingly interest both to the scientific community and to business. It is an open issue with many research challenges and very useful for decision making, due to the benefits of understanding the feelings of people instantly and automatically.

With this new way of using the Web, users do not simply browse it, they actively contribute to its contents through applications, helping to build a collective intelligence [20]. This intelligence has spread to several domains, specially those related to daily life, such as commerce, tourism, education and health, causing the social Web to expand exponentially [1]. Understanding what people are thinking or their opinions is fundamental for decision making, mainly in the context where people express their comments voluntarily, aiming to cooperate with each other.

According to Liu [16], sentiment analysis, also known in the literature as opinion mining, is the study field that analyzes people's sentiments, evaluations, attitudes and emotions in the favor of the entities such as products, services, organizations, individuals, issues, events, topics and their attributes. Sentiment analysis is a recent research field, that uses advanced techniques for text mining, machine learning, information retrieval and natural language processing (NLP) to process large amounts of non-structured content generated by users, mainly in social media [26]. This way, the objective of sentiment analysis is to extract the opinion and the subjective knowledge from online texts and then formalizing this discovered knowledge and analyzing it for specific purposes [16].

The analysis of opinionated comments expressed in social media requires too much effort to be carried out manually, mainly due to the volume of data. Hence, we seek for a summarization of the opinions. A common way of accomplishing the summarization is by means of the classification of the opinion of an object into categories: positive, negative and neutral. This kind of classification is referred to in the literature as sentiment polarity or polarity classification [16].

Twitter is a rich source to understand the people's opinions about many aspects of the daily life [14]. Performing sentiment analysis on tweets is not a trivial task due to the textual informality of the users. Algorithms that carry out sentiment analysis on tweets usually use NLP techniques, such as part-of-speech (POS) tagging. POS Tagging is used to detect subjective messages by identifying the grammatical classes of the words used in the text. POS Tagging tweets are not an easy task too, if we consider the plenty of abbreviations used in the text, due to the character limitation of the messages, repeated letters in words to emphasize terms or even the absence of consonants. In texts written in Portuguese, these problems get worse due to the grammatical complexity, inherent of that language.

In this context, we implemented and compared two approaches for sentiment analysis. We carried out a case study using tweets written in Portuguese related to the FIFA's Confederations Cup, held in Brazil in 2013. Both approaches address machine learning techniques for text classification, aiming to replace POS Taggers in the identification of opinionated tweets.The first approach is based on Naive-Bayes classifiers[30] and the last uses SVM classifiers. Each one works with two classifiers: The first classifier is used for detecting whether a tweet presents opinionated content, and the second one is used for classifying the subjective polarity of the message as either positive or negative.

The main contributions of this paper are: the implementation and comparison of sentiment classification techniques for texts without the use of POS Taggers; the presentation of a case study over sentiment analysis performed for tweets written in Portuguese; and a temporal analysis of sentiment during the FIFA's Confederation Cup, held in Brazil in 2013.

The remainder of this paper is structured as follows. Section 2 highlights related work. Section 3 addresses the problem definition and challenges in the natural language processing in microblogs. Section 4 focuses on the case study. Section 5 presents and discusses the results obtained. Finally, section 6 concludes the paper and proposes further work to be undertaken.

2. RELATED WORK

Since the beginning of the year 2000, sentiment analysis has been one of the most active research area in the field of Natural Language Processing - NLP [16],[8]. Sentiment analysis has been used in many applications with several purposes: stock exchange companies, enabling the identification of the mood of the market based on specialists' opinions [13, 19]; in analysis of consumers' reviews of products or services [6],[11]; analysis of places or tourism regions by means of the tourists' comments [3]; analysis of politicians [2] or subjects related to politics [7]; monitoring real time disease outbreaks in the regions of a country through sentiment analysis of messages posted on social networks[28].

Activities related to sentiment analysis comprise the detection of subjective or opinionated content, classification of the content polarity and summarization of the general sentiment of the evaluated entities. The sentiment detection in a text occurs in different levels: document level, sentence level and entity or aspects level. Several methods have already been proposed to classify the sentiment polarity of a text and the main approaches used are based on machine learning techniques, semantic analysis techniques, statistical techniques and techniques based on lexical analysis or thesaurus. A comparison between these techniques can be found in Sharma & Dey [26] and an overview of sentiment analysis can be found in Pang & Lee [22] and Liu [16].

The approaches to sentiment analysis that use machine learning implement classification algorithms such as Naive-Bayes, Support Vector Machine (SVM), Maximum Entropy, Decision Trees (C4.5), KNN (K-nearest neighbor), Condition Random Field (CRF), etc. One of the main limitations in the use of supervised learning is the need for labeled data for training and tests. In order to help in the task of collecting the labeled data in an automated way, many works proposed the use of emoticons - characters that transmit emotions. In Li & Li [15], 87% of the tweets containing emoticons have the same sentiments represented in the text. Studies that employ emoticons to train the classifiers have presented excellent accuracy results (above 80%). The works by Go et al. [9], Pak & Paroubek [21] and Read [24] report good results on sentiment classification using the Naive-Bayes classifier.

The work by Pak & Paroubek [21] uses the emoticons strategy to build the dataset to train a Naive-Bayes classifier and categorize tweets as either positive or negative based on N-grams and in the grammatical classification of the words of the text by means of POS Tagger. One of the main problems in using only the emoticons in the collection of the data to train the classifiers is related to the recall metric, since emoticons are present in at least 10% of the tweets [10]. Our work differs from Pak & Paroubek [21] because it does not need POS Tagger to identify an opinionated (subjective) content and in the use of a set of manually annotated data to train the classifier, thus incrementing the database obtained by means of the emoticons.

There are a few works in the literature that perform sentiment analysis using a corpus in Portuguese. The works by Chaves et al. [5], Sarmento et al. [25] and Tumitan & Becker [29] using lexical analysis techniques based on thesauri and only the work by Nascimento et al. [18] uses machine learning techniques.

In Chaves et al. [5] it is presented an algorithm called PIRPO (Polarity Recognizer in Portuguese) that uses a lexical analysis approach for classification of sentiments in comments in Portuguese. The algorithm uses ontologies and a list of polarized adjectives (positive, negative and neutral) that express sentiments to define the semantic orientation of the analyzed texts. The results achieved with PIRPO indicate a mean of the F-Measure of only 0.32 at the recognition of polarity. In Tumitan & Becker [29], the authors analyze the opinions in comments about politicians made in newspapers, and study the correlation of sentiments expressed with the vote intention surveys. The polarity identification algorithm uses a word thesaurus (SentiLex-PT) that contains the polarity for every word (positive, negative, neutral), no matter the context.

Nascimento et al. [18] use sentiment classifiers to evaluate the reactions of people on twitter concerning the news on the media. The results achieved in terms of accuracy vary from 70% to 80% according to the kind of news and the classifier used.

In Yu & Hatzivassiloglou [30], it was used two classifiers. The first one is used to classify if a tweet is informative or opinative. The second one classifies the tweet polarity, when it is opinative. A Naive-Bayes classifier was used, in both steps.

Faced with this scenario, our approach for sentiment analysis is similar to Yu et al., using two classification steps. However we propose the use of a SVM classifier and compare it with a Naive Bayes one. This is done for both tasks: opinative tweet identification and polarity analysis. The use of two classifiers eliminates the need for using POS Tagger (Part-of-Speech) in the identification of an opinionated content. So, the first classifier detects whether a content is subjective or objective; and the second classifier identifies the polarity (positive or negative) of the content previously detected as opinionated.

3. PROBLEM DEFINITION

Microblogging is a communication tool very popular among Internet users [21]. The messages shared by microblogging users are not just about their personal lives, but also about opinions and information about products, people, facts and events in general [17]. The sites providing microblogging services, such as Twitter, become rich sources for mining user opinions. Twitter has more than 200 million active users who write more than 400 millions messages a day[1]. Since it is a rich source of real-time information, many entities (companies, politicians, government, etc.) have demonstrated interest in knowing the opinions of people about services and products. The importance of Twitter for opinion mining has already been reported in other works [14].

A tweet is a short message on Twitter, with a maximum of 140 characters. Table 1 presents examples of some messages posted on Twitter about the 2013 FIFA Confederations Cup.

Since the text is essentially informal, many challenges must be taken into account in order to perform the sentiment analysis on tweets: grammatical errors, slang, repeated characters, etc. So, it is necessary to deal with the text of a tweet in a specific manner. The literature presents some proposals for dealing with such information, namely:

- Filtering: removal of URLs, Twitter user names (starting with @) and Twitter special words (”RT”, ”via”, ...);
- Removal of stopwords;
- Use of synonyms for the decomposed terms;
- Part of speech tagging usage (POS tagging);
- Recognition/Extraction of entities;
- Stemming: method for reducing a term to its radical, removing the endings, affixes, and thematic vowels; and
- Treatment of the composite terms containing HashTags. The terms are normally separated according to the capitalization of the letters. For example, “#VeryGood” becomes “Very Good” - a blank space is added between the words.

[1]https://blog.twitter.com/2013/celebrating-twitter7

Sentiment analysis of tweets may be handled as a Natural Language Processing task or, more specifically, as a text categorization task. Text categorization is the task defined as assigning predefined categories to text documents, where documents can be news stories, technical reports, web pages, tweets, etc.These categories are most often subjects or topics, but may also be based on style (genres), pertinence, among others.

More formally, the text categorization task means finding a function that approximates the classification function $F : T \to C$, $f(t_i) = c_j$. This function describes how texts are associated to the classes, and also assigns a text $t_i \in T$ to its category $c_j \in C$, where T is a domain of texts and $C = \{c_1, ..., c_n\}$ is a set of n predefined categories.

In general, a text classification task starts with a training set $T = (t_1, ..., t_n)$ of texts that are already labeled with a category $c_j \in C$ (e.g. objective, subjective). The task is then to determine a classification model (function F) which is able to assign the correct class to a new text t_i of the domain T.

To measure the performance of a classification model, a fraction of the labeled texts is set aside and not used for training. They are used to apply the proposed classification model and compare the estimated labels with true labels.

To build the classification model in this case study we use Support Vector Machine (SVM) (Vapnik, 2000). SVM is a machine learning method based on minimization of structural risk involved in the creation of the high dimension hyperplan for class separation. At the same time, it generates a classification model with the support vectors and analysis individuals that better play that role (positive or negative). SVM also guarantees that the classifying function has the smallest Vapnik-Chervonenkis (VC) dimension which also guarantees the greater generality of the classifier. To perform the class separation, SVM uses spatial transformation functions, called kernel functions. A important kernel function family is the radial basis function, often used in pattern recognition problems and also used in this work. The radial basis function is defined by:

$$R(x_i, y_i) = \epsilon^{-\gamma(x_i - x_j)^2} \quad (1)$$

where γ is a chosen parameter.

The cost parameters (C), associated with the penalties of each class and also the parameter of the radial kernel function are estimated through a grid search algorithm that is responsible for finding the set of parameters that optimize the solution. The search algorithm uses the training base (that does not have samples used in test) to approximate the best parameters. These parameters are chosen analyzing the accuracy after a cross-validation procedure.To evaluate the generality of the classifier, each result also presents the average number of support vectors used. The lesser this number in relation to the number of individuals in the training base, more generic is the classifier.

The Naive-Bayes classifer is based on the modeling of the uncertainty by means of probabilities, considering the inputs independently. However, the Naive-Bayes classifier

Table 1: Example of tweets sent by users

User	Date and Time	Original Message	Translated Message
@BH2014Copa	2013-04-12 13:16:23	Bom dia, faltam 63 dias para a Copa das Confederações da FIFA. BH está ansiosa!	Good morning, we're 63 days away from FIFA's Confederation's Cup. BH is looking forward to it!
@apenas_be	2013-04-12 14:05:48	Quero estar no Rio na copa das confederações #vaiSerMuitaOndaaa	I want to be in Rio for the Confederation's Cup #itsGonnaBeAwesome

presents optimum results even problem classes which have highly dependent attributes [22, 9, 21, 24].This occurs due to the fact that the conditional independence of attributes is not a necessary condition for Naive Bayes optimality. The Naive-Bayes algorithm basically brings the same mathematical foundations of the Bayes Theorem. Applying this theorem to the context of tweets classifiers, we have:

$$P(c|t) = \frac{P(c)P(t|c)}{P(t)} \tag{2}$$

where:
P(c) is the occurrence probability of the category;
P(t) is the occurrence probability of the tweet;
P(t|c) is the occurrence probability of the tweet, given that the category occurred;
P(c|t) is the probability of tweet, given that it occurred, belonging to the category.

The term is computed taking into consideration the conditional probability of occurrence of each word that forms the tweet, since the category have occurred. This term could be written as:

$$P(t|c) = \prod_{1<=k<=n} P(t_k|c) \tag{3}$$

where:
$P(t_k|c)$ is the probability of the term k occurring, given that the category occurred;
n is the tweet length.

We also analyze the case study results through the standard Information Retrieval (IR) metrics, which are Precision, Recall and F-Measure. The fraction of correctly classified documents in relation to the total number of documents is called accuracy, and is a basic performance measure.

4. CASE STUDY DESIGN

In order to carry out this study, we needed to create a sentiment classifier whose classification process was done in two steps, as shown in the diagram of Figure 1. We present in this section the overall design of our developed case study. To describe this case study, this section is subdivided into three subsections: the selection of the corpus, which describes the used dataset; the sentiment polarity classification of tweets, which addresses the usage of our technique; and the evaluation, which focuses on validation.

4.1 Selection of the Corpus

We collected approximately 300,000 tweets in the Portuguese language concerning the theme of FIFA's Confederations Cup, which took place in Brazil, in 2013. The Twitter's REST search API was used for collecting tweets by

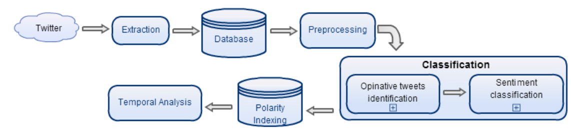

Figure 1: Process for classification of tweets' Sentiment Polarity

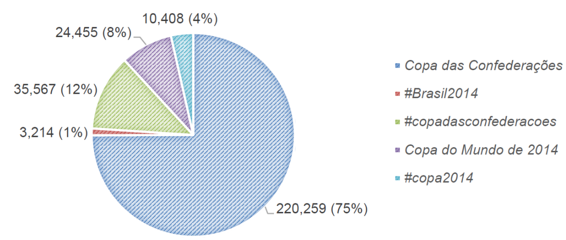

Figure 2: Number of tweets obtained through query terms

GET requests. We created a crawler which collected automatically, every day between April and August 2013, tweets containing at least one of the following terms: #copa2014, #Brasil2014, #Brasil2014, Copa do Mundo de 2014 (World Cup 2014), Copa das Confederações (Confederations Cup) and #copadasconfederacoes. After the data be collected and stored in a database, the tweets were submitted to a preprocessing, that includes the removal of stopwords, special terms (RT, via, etc.), removal of user names; and hashtags treatment (separation of composite terms, according to the capitalization of letters).

Figure 2 presents a graphic containing the number of tweets obtained by the search terms. It can be clearly noticed that the term "Copa das Confederações" (Confederations Cup) used in the queries had the largest number of tweets.

The data was collected between April 12th and August 12th in the year of 2013, approximately two months before the beginning and two months after the end of the competition, which occurred between July 15th and 30th, in 2013. That period of time is important for a temporal analysis, enabling the perception of possible sentiment - opinion - of the Brazilian people with respect the theme of the cup in Brazil. Figure 3 illustrates the number of tweets posted every day during the data gathering.

It is possible to see (Figure 3) that by the time of the competition the number of tweets sent was higher, as expected. It is also possible to notice that, in May 14th of 2013, there was a unusual number of tweets. About 17,000 tweets were collected in that date. In that date, the list of

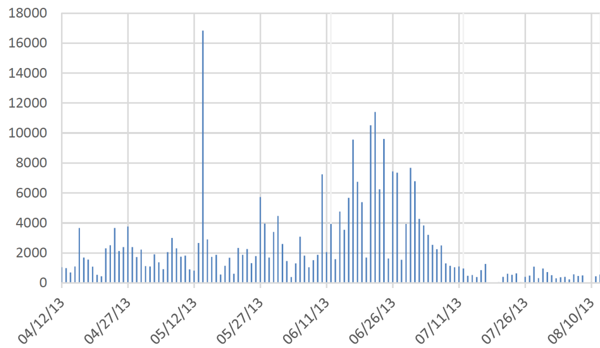

Figure 3: Number of collected tweets by day

players summoned by the Brazilian football team was publicized. Supposedly, most of those tweets were opinionated and must reflect the popular opinion about the selected players to form the team for the competition.

4.2 Sentiment Polarity Classification of the Tweets

The use of classifiers such as SVM and Naive Bayes, has already been discussed in the literature about sentiment analysis combined with other techniques such as Pos Taggers [26, 22, 24, 9, 21, 30]. However, in this work we implemented two approaches for sentiment classification of tweets and compared the achieved results. The process of sentiment classifying for both approaches is performed in two classification steps: in the first step only tweets with subjective text are classified; and in the last step each tweet is classified with a single sentiment either positive or negative. The SVM and Naive Bayes techniques have been used in the literature for traditional sentiment classification processes, in which several authors argue that SVM performs better in the text classification [26, 22, 24, 12]. Thus, this study aims analyzing the behaviour of such approaches when the process of sentiment classifying is performed in two steps, as mentioned earlier.

Support Vector Machine (SVM) has showed a good generalization performance and can easily learn the exact parameters for the global optimum. SVM seeks a hyperplane represented by vectors that splits the positive and negative or subjective and objective training vectors of documents (tweets) with maximum margin. The problem of finding this hyperplane can be translated into a constrained optimization problem. SVM algorithm classifies opinionated text vectors by separating it into positive and negative classes with a hyperplane, which can be further extended to non-linear decision boundaries using various kernels [27]. In this work we utilized the bag of words representation for each tweet as feature of the SVM. This study has used a variant of SVM for fast training using Sequential Minimal Optimization (SMO). For more details about SMO see [23].

In the approach of construction of a polarity classifier through supervised machine learning techniques, it was necessary to have labeled (classified) data to train the classifier. So, we adopted two approaches to obtain the tweets with sentiments labeled:

Emoticons: using the approach from Pak & Paroubek [21], which assumes that all words in the message have characters that express emotions, e.g. Happy emoticons - ":-)", ":)", "=)", ":D", etc - and Sad emoticons - ":-(", ":(", "=(", ";(", etc -, are also associated to the emotion of the character (emoticon). So, if a tweet presents a happy emoticon (":-)"), for instance, its polarity is considered positive.

Manual Labeling: 1,500 tweets from the collected one were randomly chosen and separated for manual labeling of the sentiment polarity. We asked 10 volunteers to give their opinions about the sentiments present in the tweets, so that, using the majority vote of the discrepant opinions, only those that presented a dominant sentiment would be considered valid for labeling.

Both methods for obtainment of sentiments labeling were used in the comparison and combination of results of the classifiers. Considering the randomness in the choice of tweets to be labeled manually, we found that only 12 tweets had emoticons and that the process of labeling, at least one of the users confirmed the sentiments expressed by emoticons. Table 2 presents the number of tweets with sentiments labeled using the two approaches. We considered neutral tweets those who do not express any opinion.

Table 2: Number of Tweets Labelled (Training and Testing Sets)

Approach	Positive	Negative	Neutral	Total
Emoticons	1,468	492	-	1,960
Manual Labeling	326	321	463	1,110

Using distinct training sets, two binary classifiers were built: a classifier to check whether a tweet is subjective, that is, presents an opinion; and a polarity classifier to distinguish the sentiment as positive or negative. Once the sentiment classifier is trained, all tweets were analyzed and indexed with the opinion polarity obtained by the classifier.

Finally, the collected sentiments were summarized by means of a temporal analysis, which enabled us to follow the general orientation of the sentiments expressed by the Brazilian people with respect to the subject of the Confederations Cup. Besides, a word frequency counter was used to detect the main terms cited in the tweets.

4.3 Evaluation

In order to validate the sentiment polarity classifier, we used 10-fold cross validation technique with all of the labeled tweets. The metrics accuracy, precision, recall and F-Measure were used to evaluate the results. These metrics are defined in equations 4, 5, 6, and 7, respectively. In these equations, TP indicates true positive, which is defined as the number of tweet-opinion pairs that the system identifies correctly as positive, TN indicates true negative, which is defined as the number of tweet-opinion pairs that the system identifies correctly as negative, FP indicates false positive which is defined as the number of feature-opinion pairs that are identified falsely by the system, and FN indicates false negatives which is the number of feature-opinion pairs that the system fails to identify.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

Table 3: Comparison of Developed Classifiers (SVM and Naive-Bayes)

Classifier / Technique	**Dataset (Training and Testing)**	**Accuracy**	**Class**	**Precision**	**Recall**	**F-Measure**
Subjective tweet classification + Polarity Classifier / SVM	Emoticons + Manual Labeling	0.800	**Positive**	0.839	0.873	0.856
			Negative	0.715	0.657	0.685
			Weighted Average	0.799	0.802	0.800
Subjective tweet classification + Polarity Classifier / Naive-Bayes	Emoticons + Manual Labeling	0.777	**Positive**	0.91	0.742	0.817
			Negative	0.616	0.849	0.714
			Weighted Average	0.813	0.777	0.783

Table 4: Comparison of training classifier datasets (SVM)

Classifier	**Dataset (Training and Testing)**	**Accuracy**	**Class**	**Precision**	**Recall**	**F-Measure**
Subjective tweet classification + Polarity Classifier	Emoticons	0.870	**Positive**	0.953	0.847	0.897
			Negative	0.748	0.916	0.824
			Weighted Average	0.885	0.870	0.873
	Manual Labeling	0.656	**Positive**	0.762	0.716	0.738
			Negative	0.469	0.529	0.497
			Weighted Average	0.668	0.656	0.661

Table 5: Comparison of training classifier datasets (Naive-Bayes)

Classifier	**Dataset (Training and Testing)**	**Accuracy**	**Class**	**Precision**	**Recall**	**F-Measure**
Subjective tweet classification + Polarity Classifier	Emoticons	0.727	**Positive**	0.820	0.765	0.791
			Negative	0.569	0.649	0.606
			Weighted Average	0.739	0.729	0.733
	Manual Labeling	0.650	**Positive**	0.805	0.636	0.710
			Negative	0.472	0.678	0.556
			Weighted Average	0.697	0.650	0.672

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F - Measure = \frac{2 \times PrecisionRecall}{Precision + Recal} \tag{7}$$

5. RESULTS AND DISCUSSION

As it can be observed from Table 3, which presents the results achieved with the developed classifiers, the best results are obtained by the classifier who implements the SVM technique. Tables 4 and 5 present a comparison of the datasets used to train the classifiers. The results of the Naive-Bayes and SVM subjectivity classifiers shown an accuracy of 82% and 84%; and a F-measure of 0.819 and 0.821, respectively.

One of the main disadvantages in the approaches that use machine learning for sentiment classification is the construction of a training set. So, we used the proposal of Pak & Paroubek [21] for automatic collection of data through emoticons to train the sentiment classifiers. The results achieved with the classifier by the training carried out with the data resulting from the automatic labeling were compared with the classifier trained with the data manually labeled. It can be noticed, in Tables 4 and 5, that both classifiers that use the dataset obtained automatically present the best results.

Having built and validated the sentiments classifier and having used all the labeled tweets, the next step was to obtain of the general semantic orientation of the sentiments expressed by the Brazilian people regarding the 2013 FIFA Confederations Cup. Figure 4 presents the result of the sentiments classifier applied to all of the collected tweets. In that graphic, the number of sentiments classified with negative polarity was plotted in the negative semi-axis to avoid superpositions with the number of sentiments classified with positive polarity.

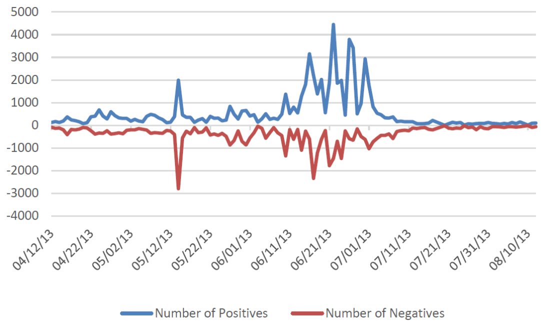

Figure 4: Number of positive and negative tweets

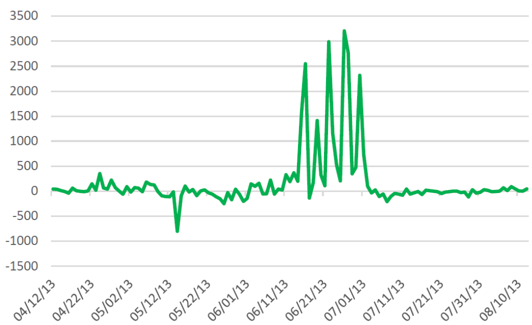

Figure 5: Semantic orientation of the collected tweets' sentiment

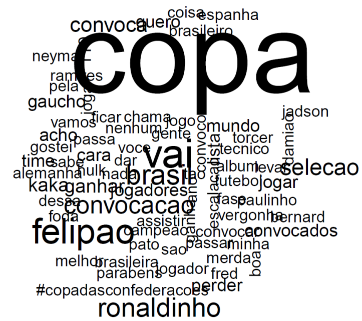

Figure 6: Negative polarity tweets in 05/14/2013

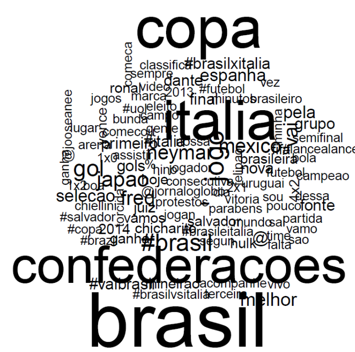

Figure 7: Positive polarity tweets during the competition

One way of obtaining the general semantic orientation of the sentiment expressed in the tweets is by subtracting the number of tweets with negative sentiments from the number of tweets with positive sentiments. Figure 5 presents the summary of the semantic orientation of the collected tweets.

The tool developed in the present work enables the decision maker to analyze the words more frequently used in a selected time interval, thus helping to identify possible dissatisfactions expressed in the messages or even complements about the evaluated objects. Figure 6 presents a cloud of words with the most frequent terms used in the tweets of 2013-05-14. Figure 7 presents a cloud of words obtained during the games played by the Brazilian team, in which the semantic orientation of the tweets was positive.

6. CONCLUSION

We have implemented and compared two approaches for sentiment analysis in case study about tweets written in Portuguese related to the FIFA's Confederations Cup, occurred in Brazil in 2013. The first approach uses two Naive-Bayes classifiers and the second one uses two SVM classifiers. The developed approaches replaced POS Tagger for identification of opinative tweets. The results obtained by the SVM sentiments classifier indicated an F-Measure of 0.873 and an accuracy of 80.0% for detection of sentiment polarity. The Naive-Bayes sentiment classifier presented a F-measure of 0.791 and accuracy of 72.7%. The results presented by both classification approaches are considered satisfactorily for Portuguese tweets, specially if we consider that the polarity of subjective content is not always consensual. The use of a SVM classifier increases the accuracy in 8%, that is a good result [12]. For example, in annotations made by humans, consensus is hardly above 75% [22]. Other sentiment analysis studies applied to the English language obtained, at the best scenarios, an accuracy of around 95% for detection of sentiment polarity [27]. The results achieved, however, are not enough to conclude whether it is always better to use SVM than Naive-Bayes classifiers. It would be necessary novel studies with different datasets aiming to reach a generalization.

Also, we performed a temporal analysis on the data in this work, aiming to identify the semantic orientation of the sentiments expressed by means of the tweets, as well as the identification of the more cited terms in the opinionated messages.

One of the main limitations identified is concerned with the identification of the entity referred to by the opinion detected in the tweet. Though the collected tweets are related to the 2013 FIFA Confederations Cup, the opinions expressed in the messages may refer to other entities. In this sense, we believe that the application of Named Entities Recognition techniques might minimize this problem. Also, it will be interesting to investigate relationships across Twitter entities aiming to improve accuracy for detection of sentiment polarity .

A future work will explore temporal series to help in the prediction of sentiments according to the detected tendency. Regarding the generation of the clouds of words through the most frequent terms, we believe that it is possible to apply an approach which is similar to that used by Hu & Liu [11], in which the authors perform the automatic summarization of opinion about products and services reviews, including the aspects (features) of the observed entities. So, it will be possible to identify which aspects regarding the FIFA Confederations Cup were considered positive or negative.

7. REFERENCES

[1] Appelquist, D., Brickley, D., Carvahlo, M., Iannella, R., Passant, A., and Perey, C. A Standards-based, Open and Privacy-aware Social Web, 2010.

[2] Awadallah, R., Ramanath, M., and Weikum, G. PolariCQ. In *Proceedings of the 21st ACM international CIKM '12* (New York, New York, USA, 2012), ACM Press, p. 1945.

[3] Bjø rkelund, E., Burnett, T. H., and Nø rvåg, K. A study of opinion mining and visualization of hotel reviews. In *Proceedings of the 14th International Conference on IIWAS* (New York, New York, USA, 2012), ACM Press, p. 229.

[4] Calais Guerra, P. H., Veloso, A., Meira, Jr., W., and Almeida, V. From bias to opinion: A transfer-learning approach to real-time sentiment analysis. In *KDD '11* (New York, NY, USA, 2011), ACM, pp. 150–158.

[5] Chaves, M., de Freitas, L., Souza, M., and Vieira, R. Pirpo: An algorithm to deal with polarity in portuguese online reviews from the accommodation sector. In *Natural Language Processing and Information Systems*, G. Bouma, A. Ittoo, E. Métais, and H. Wortmann, Eds., vol. 7337 of *LNCS*. Springer Berlin Heidelberg, 2012, pp. 296–301.

[6] Eirinaki, M., Pisal, S., and Singh, J. Feature-based opinion mining and ranking. *Journal of Computer and System Sciences 78*, 4 (July 2012), 1175–1184.

[7] Fang, Y., Si, L., Somasundaram, N., and Yu, Z. Mining contrastive opinions on political texts using cross-perspective topic model. In *WSDM'12* (New York, New York, USA, 2012), ACM Press, p. 63.

[8] Feldman, R. Techniques and applications for sentiment analysis. *Communications of the ACM 56*, 4 (Apr. 2013), 82.

[9] Go, A., Huang, L., and Bhayani, R. Twitter sentiment classification using distant supervision. Tech. rep., Stanford, 2009.

[10] Gonçalves, P., Araújo, M., Benevenuto, F., and Cha, M. Comparing and combining sentiment analysis methods. In *COSN '13* (New York, 2013), ACM Press, pp. 27–38.

[11] Hu, M., and Liu, B. Mining and summarizing customer reviews. In *KDD '04* (New York, New York, USA, 2004), ACM, ACM Press, p. 168.

[12] Joachims, T. Text categorization with suport vector machines: Learning with many relevant features. In *Proceedings of the 10th ECML* (London, UK, UK, 1998), Springer-Verlag, pp. 137–142.

[13] Koppel, M., and Shtrimberg, I. Good News or Bad News ? Let the Market Decide. In *AAAI Spring Symposium on Exploring Attitude and Affect in Tex* (2004), pp. 86–88.

[14] Kwak, H., Lee, C., Park, H., and Moon, S. What is Twitter, a social network or a news media? In *WWW '10* (New York, 2010), ACM Press, p. 591.

[15] Li, Y.-M., and Li, T.-Y. Deriving Marketing Intelligence over Microblogs. In *44th HICSS* (Jan. 2011), IEEE, pp. 1–10.

[16] Liu, B. Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies 5*, 1 (May 2012), 1–167.

[17] Naaman, M., Boase, J., and Lai, C.-H. Is it really about me? In *CSCW '10* (New York, New York, USA, 2010), ACM Press, p. 189.

[18] Nascimento, P., Aguas, R., Lima, D. D., Kong, X., and Osiek, B. Análise de sentimento de tweets com foco em notícias. In *I BRASNAM* (2009).

[19] O'Hare, N., Davy, M., Bermingham, A., Ferguson, P., Sheridan, P., Gurrin, C., and Smeaton, A. F. Topic-dependent sentiment analysis of financial blogs. In *Proceeding of the TSA '09* (New York, New York, USA, 2009), ACM Press, p. 9.

[20] O'Reilly, T. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *Communications & Strategies 1*, 65 (2007), 17–37.

[21] Pak, A., and Paroubek, P. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In *LREC'10* (2010), pp. 1320–1326.

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

[23] Platt, J. C. Advances in kernel methods. In *Fast Training of Support Vector Machines Using Sequential Minimal Optimization*, B. Schölkopf, C. J. C. Burges, and A. J. Smola, Eds. MIT Press, Cambridge, MA, USA, 1999, pp. 185–208.

[24] Read, J. Using Emoticons to reduce Dependency in Machine Learning Techniques for Sentiment Classification. In *ACL 05* (2005), pp. 43–48.

[25] Sarmento, L., Carvalho, P., Silva, M. J., and de Oliveira, E. Automatic creation of a reference corpus for political opinion mining in user-generated content. In *Proceedings of the TSA '09* (New York, 2009), ACM Press, p. 29.

[26] Sharma, A., and Dey, S. A comparative study of feature selection and machine learning techniques for sentiment analysis. In *RACS '12* (New York, New York, USA, 2012), ACM Press, p. 1.

[27] Sharma, A., and Dey, S. A boosted svm based sentiment analysis approach for online opinionated text. In *RACS '13* (New York, NY, USA, 2013), ACM, pp. 28–34.

[28] Silva, I. S., Gomide, J., Barbosa, G. A. R., Santos, W., Veloso, A., Meira Jr, W., and Ferreira, R. Observatório da dengue: Surveillance based on twitter sentiment stream analysis. In *SBBD 2011* (2011), pp. 49–54.

[29] Tumitan, D., and Becker, K. Tracking Sentiment Evolution on User-Generated Content : A Case Study on the Brazilian Political Scene. In *SBBD 2013* (2013), pp. 1–6.

[30] Yu, H., and Hatzivassiloglou, V. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *EMNLP '03* (Stroudsburg, PA, USA, 2003), pp. 129–136.