Text Analysis and Sentiment Polarity on FIFA World Cup 2014 Tweets

Peiman Barnaghi[1,2]
peiman.barnaghi@insight-centre.org

Parsa Ghaffari[2]
parsa@aylien.com

John G. Breslin[1]
john.breslin@nuigalway.ie

[1]Insight Centre for Data Analytics, National University of Ireland Galway, Galway, Ireland

[2]Aylien Ltd., Dublin, Ireland

ABSTRACT

Social media has become one of the most popular communication tools for sharing opinions and everyday life-related events. Twitter as a micro-blogging service can be used to discover events and news in real time from anywhere in the world. As Twitter posts (tweets) are short and are being generated constantly, they are well-suited sources of streaming data for opinion mining and sentiment polarity detection. Tweets can reflect public sentiment when taken in aggregation, for example, during events such as the FIFA World Cup. In this paper, we examine the effectiveness of a machine learning method for providing positive or negative sentiment on tweets. We use the sentiments extracted from Twitter to look for a correlation between these sentiments and FIFA World Cup 2014 events of interest. Using this correlation, we determine changes in sentiment polarity based on some words in tweets and major events that occurred. We use Twitter's Streaming API for mining tweets and processing them by filtering using some of the official World Cup hashtags (e.g. "#WorldCup" and "#Brazil2014"). A training set of tweets is manually labeled as positive or negative and is used to test the performance of text categorization methods for sentiment analysis. We accomplish our work using one of the well-known machine learning methods for text categorization and determining sentiment, called Logistic Regression Classification (LRC). We attempted to perform sentiment polarity detection using our trained model to find correlations between tweets and events. This allows us to analyze the reaction of people towards unexpected events or unethical behavior during the tournament. Experimental results show that people have a very negative sentiment towards unethical behavior that has occurred.

Categories and Subject Descriptors

I.5.4 [Pattern Recognition]: Applications—Text Processing

General Terms

Algorithms.

Keywords

Sentiment Analysis, Opinion Mining, Stream Data Analysis, Sentiment Knowledge Discovery, Sentiment Classification.

Conference ACM SIGKDD'15, August 10-13, 2015, *Sydney, Australia*.
Workshop on Large-Scale Sports Analytics (LSSA)

1. INTRODUCTION

Social media is a very popular method for expressing opinions and interacting with other people in the online world. Twitter is one of the most common online social media and micro-blogging services. It enables users to connect with others and get updates on topics and events that interest them. Tweets (Twitter posts) provide real raw data in the format of short texts that express opinions, ideas and events captured in the moment. As tweets are short and generated constantly, they are well-suited streams of data for opinion mining and sentiment knowledge discovery. Opinions, evaluations, emotions and speculations often reflect the states of individuals; they consist of opinionated data expressed in a language composed of subjective expressions [1]. More than 400 million tweets are posted on Twitter per day. Each has a maximum of 140 characters but most tweets are around 30 characters long that would be over 12 billion characters generated per day on Twitter. A large collection of such tweets could be leveraged to provide a useful reflection of public sentiment towards some special events, for example, in the FIFA World Cup. After 64 football matches, 672 million tweets were sent related to World Cup 2014 since the tournament began [2].

Using polarity analysis, an idea of people's reactions to various events can be extracted, which can provide valuable results and also predictions [3]. This paper looks at some of the major talking points from the tournament, which were extracted from Twitter data during the World Cup 2014. The following event was one of the most controversial ones: Uruguayan Luis Suarez was accused of biting Italy's defender Giorgio Chiellini. A large volume of negative comments and feedback followed that event on social media platforms including Twitter. In this paper, we examine the effectiveness of a commonly used text categorization machine learning method called Logistic Regression Classification for providing positive or negative sentiment on tweets. We use extracted Twitter sentiment to look for correlations between this sentiment and major FIFA World Cup 2014 events.

2. Related Work

Sentiment polarity returns the overall opinion of a text or document for one single issue. Pang et al. [4] provided a broad overview of some of the machine learning techniques used in sentiment classification.

Opinions are classified as one of two opposing sentiment polarities (positive or negative), or may be labeled as neutral when there is a lack of opinion in the text or the opinion is located in between these two polarities. This kind of labeling can be used to summarize the content of opinionated texts and documents. A wide variety of features can be necessary for opinion and polarity recognition [5]. Sentiment analysis on Twitter data is not an easy task because a tweet may carry positive and negative feeling simultaneously[1].

Sinha et al. [6] used Twitter data as social media output to find a relationship with the National Football League (NFL) games. They used logistic regression classifier to predict game and betting outcome. In order to measure the performance of feature sets, they tried to find regularization coefficient based on all games from previous years weeks and select the best coefficient for each feature set and each test week. It used to compute the accuracy of predictions across all games in all test weeks.
It is essential to build a accurate classifier based on training dataset for sentiment analysis; so appropriate learning algorithms can be applied [7].

2.1 Tweet Sentiment Analysis Steps

Figure 1 shows the steps taken to build a model for sentiment analysis on Twitter data. The trained model is used for polarity detection on World Cup tweets and for finding some correlations between tweets and major events in the World Cup.

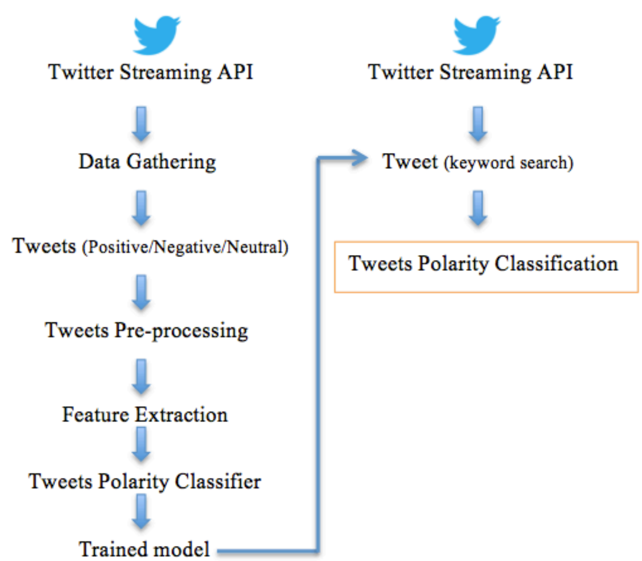

Figure 1. Building a trained model for polarity detection and applying it on some tweets.

2.2 Data Collection

Twitter's Streaming API was used for mining tweets. Data gathering was made up of two steps: the first one was collecting the data to use as a training set to build the model. This consisted of 4162 tweets manually labeled "positive" or "negative". The second step was collecting tweets during the World Cup tournament and processing them by filtering some of the official World Cup hashtags (e.g. "#WorldCup" and "#Brazil2014"), as well as team code hashtags (e.g. "#ARG" and "#GER"). In addition, the Twitter usernames of teams and players were used to extract some more tweets related to events that occurred during the tournament. The data was in JSON format as a set of documents, one for each tweet [8].

2.3 Tweet Text Pre-processing

As a first step towards finding a tweet's sentiment and in order to obtain accurate sentiment classification, we need to filter some noise and meaningless symbols from the original text of tweets that do not contribute to a tweet's sentiment. This was done by splitting up the text using spaces and constructing a bag of words, which is called tokenization. Each word can be used as a feature to train the classifier, but we needed to keep some prominent ones and remove some useless and meaningless words or symbols. Tweets may also include symbols; for example the word following the "@" symbol is a username and "#" is used to mark topics or keywords in a tweet. All usernames and URLs were converted to generic tags (e.g. all @usernames tagged as "username"), and some of mentions can be used to improve the performance of the sentiment classifier [9].

2.4 Sentiment Classification

Labeling an opinionated text and categorizing it overall into a positive, negative or neutral class is called sentiment polarity classification. The neutral label is used for more objective items that have a lack of opinion in the text, or where there is a mixture of positive and negative opinions therein [5].

2.4.1 Feature extraction

Selecting a useful list of words as features of a text and removing a large number of words that do not contribute to the text's sentiment is defined as feature extraction. It helps us to filter noise from the text and obtain a more accurate sentiment for a tweet. In this paper, we use an N-grams feature for sentiment classification of World Cup tweets.

2.4.1.1 Unigrams feature

Unigrams are defined as looking at one word at a time in a text, which can be extended to an N-gram in order to exploit the ordering of words.

2.4.1.2 N-grams feature

An N-grams feature is defined as taking a set of sequential words in a text; for example if N=2, it means looking at a pair of sequential words at a time, which is called a bigram.

Related work based on unigrams shows that, the nature of the dataset has an impact on classification performance. Pang et al. [4] show that unigrams yield better performance on movie reviews for sentiment polarity classification. As tweets are very short texts with at maximum a length of 140 characters and most tweets are around 30 characters long, the N-grams feature with N=1 to 2 is used, which uses a reasonable list of sequential words for sentiment classification.

2.4.2 Feature filtering

The large number of features (different words in corpora) forces us to use methods for selecting the top features to use for training the classifier. Term Frequency-Inverse Document Frequency (TF-IDF) is a numerical statistic method to filter the features by weighting and scoring each of the unigrams and N-grams using the frequency of words in the text [4].

3. Machine Learning Method

One of the most well-known and frequently used classifying algorithms for text analysis and categorization has been described below which is used to train a model based on Twitter data. The tweet polarity classifier is trained based on N-gram features (N=1 to 2) using WEKA[1] as a data mining and machine learning framework. Cross validation as a repeated holdout method is used on the dataset by splitting it into 10 sections. This method selects 90% for the training set, and 10% for the testing set, repeating it on 10 different sections of dataset. Finally, the result is averaged over the rotated divided sections. The goal of using this method is testing the model in the training phase [10].

3.1 Bayesian Logistic Regression

The Bayesian Logistic Regression (BLR) models simultaneously select features and provide shrinkage for performing text categorization. It uses a Laplace prior to avoid over-fitting and produces sparse predictive models for text data [11]. The Logistic Regression estimation of P(c|f) has the parametric form:

$$P(c|f) = \frac{1}{z(f)} exp((\sum_{i} \lambda_{i,c} F_{i,c}(f,c))$$

Where z(f) is a normalization function, λ is a vector of weight parameters for the feature set [12], and $F_{i,c}$ is a binary function that takes as inputs a feature and a class label. It is defined as:

$$F_{i,c}(f,c') = \begin{cases} 1, & n(f) > 0 \text{ and } c' = c \\ 0, & \text{otherwise} \end{cases}$$

This binary function is trigged when a certain feature (unigram, bigram, etc.) exists.

3.2 Trained Classifier Evaluation

The first phase of this work is an evaluation of how the BLR classifier (as a well-known text classifier based on N-gram feature sets) affects the performance of a simple two-class (positive/negative) sentiment analyzer. The following table displays the corresponding values for each experiment.

Table 1. Tweet polarity classifiers based on an N-grams feature.

ML method	Correctly classified instances %		Precision		Recall		f Measure	
	Uni	Bi	Uni	Bi	Uni	Bi	Uni	Bi
BLR (Positive)	72.17	66.21	71.7	63.3	72.9	76.2	72.3	69.2
BLR (Negative)			72.7	70.5	71.5	56.3	72.1	62.6

[1]Weka is a collection of machine learning algorithms for data mining tasks. (http://www.cs.waikato.ac.nz/ml/weka/)

4. Tweets and World Cup 2014 Events

For 64 football matches during the World Cup 2014, 672 million tweets were posted related to this event since the tournament began [2]. Patterns can be extracted based on the number of tweets over time during matches and also when a major event has occurred. For example, there is a correlation between the number of tweets about a team after elimination or qualification for the next round, or some other unexpected result.

4.1 Sentiment Analysis

In this paper, sentiment analysis was carried out using our trained model for some of the major events that occurred during the tournament. We analyzed only English tweets from the 30 million gathered tweets because a language analysis of the World Cup tweets showed that the majority of tweets (51.56%) were in English (shown in Figure 2). The positive, negative or neutral polarity values of these tweets were used to see what these values are for different entities and how they change over time, as a result of various events.

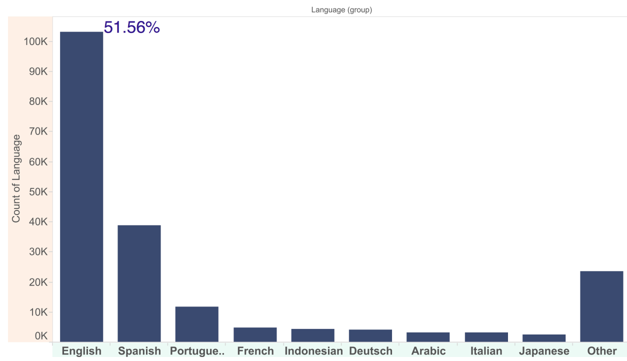

Figure 2. Most-used languages in tweets during the FIFA World Cup 2014.

4.2 Sentiment Over Time

Tweets can provide a reflection of public sentiment when taken in aggregation during special events such as the FIFA World Cup. Different events concerning players or a team affect how people think and talk about them. For example, Figure 3 shows an average sentiment polarity breakdown for the top 10 players in World Cup 2014.

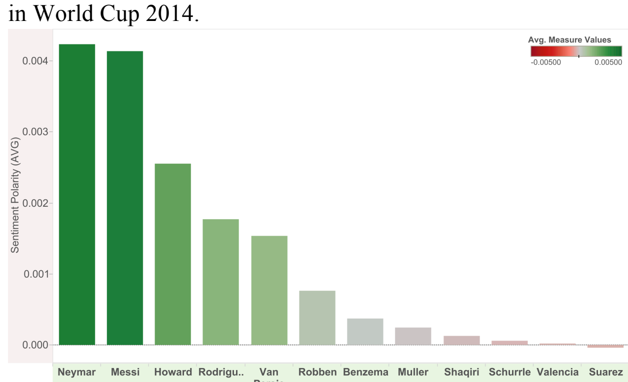

Figure 3. Average polarity of the top 10 players in World Cup 2014 (the first one with the top positive polarity is "Neymar" and the tenth one with negative polarity is "Suarez").

Polarity analysis can be helpful to get an idea of people's reactions to various events, which can provide valuable insights. Trends of tweet polarity around talking points from the tournament are shown in the following sections.

4.2.1 A major event during the World Cup

We tried to extract all related tweets for a major event and to find any correlations between the tweets and the event that occurred. During the FIFA World Cup 2014, on June 24th, Uruguayan Luis Suarez was accused of biting Italy's defender Giorgio Chiellini. The event was followed by a large volume of negative tweets on Twitter. Using the trained model, sentiment classification was performed on all tweets that mentioned the player's name. The sentiment classification output (Figure 4) shows that the trend of tweet polarity is divided into three different parts of sentiment for the aforementioned player.

The first part consists of polarity values of all tweets before the incident of biting the defender on Italy's team. There is a fluctuation of sentiment polarity rates based on player performance and match results. Almost all of these sentiments are positive with different strength (such as strongly positive) or else neutral. The second part of the sentiment polarity shows the beginning of a negative trend after the incident. Almost all tweets are negative with different rates. The third part of the sentiment polarity starts when Suarez issued an apology on June 30th, which seems to have been satisfactory for the Twitter community and a positive trend starts growing and reaches a peak level of positive polarity when he signed his new contract with Barcelona FC.

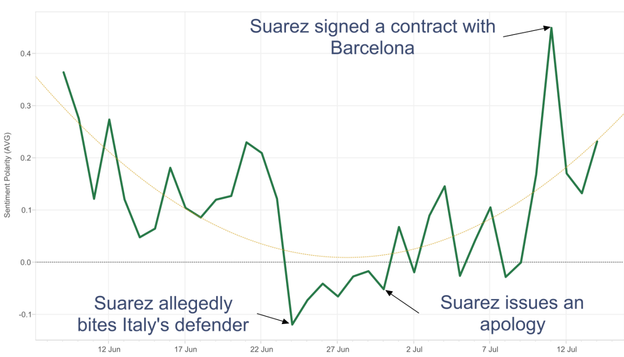

Figure 4. Sentiment polarity of tweets about Luis Suarez during World Cup 2014.

5. Future Work

Future work will be focused on the polarity classification of scalable topic-level streaming feeds, with classification of the streaming feeds' sentiment towards a given topic (and not just a keyword). The next step can be defined as trend detection toward a topic on a set of streaming feeds to determine the polarity of them towards a target topic.

6. Conclusions

Analyzing Twitter posts allows the extraction of detailed insights into opinions and trends around sporting events such as the FIFA World Cup, players, teams, etc. and how they change over time during a critical event or after unethical behavior. In this paper, a sentiment classification model is trained based on Twitter data using text features. We extracted sentiment polarity for one major event that occurred during the World Cup using our trained model. The experimental results show the positive and negative reaction of people towards such events and also changes in the trends based on some changes within those events. This kind of sentiment analysis helps us to use Twitter data for extracting patterns based on opinionated texts. In addition, teams, players, etc. can receive an overall sentiment in relation to their performance and behaviors, that could be used to help to improve the quality of matches by highlighting controversial ethical issues as well.

7. ACKNOWLEDGMENTS

This project has emanated from research conducted with the financial support of Irish Research Council (IRC) under Grant Number EBPPG/2014/30 and Aylien Ltd. as Enterprise Partner. This publication has emanated from research supported in part by a research grant from Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289 (Insight).

8. REFERENCES

[1] A. Bifet and E. Frank, "Sentiment knowledge discovery in twitter streaming data," in *Discovery Science*, 2010, pp. 1-15.

[2] T. Blog, "Insights into the #WorldCup conversation on Twitter," in *Twitter Blog*, ed, 2014.

[3] S. Asur and B. A. Huberman, "Predicting the future with social media," in *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, 2010, pp. 492-499.

[4] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, 2002, pp. 79-86.

[5] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and trends in information retrieval,* vol. 2, pp. 1-135, 2008.

[6] S. Sinha, C. Dyer, K. Gimpel, and N. A. Smith, "Predicting the NFL using Twitter," *arXiv preprint arXiv:1310.6998,* 2013.

[7] P. Priyanthan, B. Gokulakrishnan, T. Ragavan, N. Prasath, and A. S. Perera, "Opinion mining and sentiment analysis on a twitter data stream," *ICTer 2012,* 2012.

[8] D. Terrana, A. Augello, and G. Pilato, "Automatic Unsupervised Polarity Detection on a Twitter Data Stream," in *Semantic Computing (ICSC), 2014 IEEE International Conference on*, 2014, pp. 128-134.

[9] L. Zhang, "Sentiment analysis on Twitter with stock price and significant keyword correlation," 2013.

[10] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD explorations newsletter,* vol. 11, pp. 10-18, 2009.

[11] A. Genkin, D. D. Lewis, and D. Madigan, "Large-scale Bayesian logistic regression for text categorization," *Technometrics,* vol. 49, pp. 291-304, 2007.

[12] H. Daumé III, "Notes on CG and LM-BFGS optimization of logistic regression", vol. 198, p. 282, 2004.