

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

Opinion Mining and Sentiment Analysis on Twitter

Gangawane A.A¹, H.B.Torvi²

Department of Computer science and Engineering, V.V.P. Institute of Engineering and Technology, Solapur, India

Assistant Professor, Department of Computer Science and Engineering, V.V.P. Institute of Engineering and Technology ,
Solapur, India.

ABSTRACT: The main objective of this system is to analyze the Sentiments on short sentence like abbreviations, improving sentiment word identification algorithm as well as Successful handling of bi polar sentiments. Twitter is the recent innovation and it provides lots of options to user to satisfying their communication needs around the globe. The motto of this work is to segment each and every tweet and also identify Rumor using segmentation methods. In this system we concentrate on a positive or negative sentiment on Twitter posts using a well-known machine learning method for text categorization. Bayesian Logistic Regression [BLR] classification model is used to train the dataset and classify as well as process the tweets according to that and provides results in more accurate manner. We utilized outer vocabularies to distinguish subjective or target tweets, included Unigram and Bigram highlights and utilized TF-IDF [Term Recurrence Inverse Document Frequency] to sift through the highlights. A similar approach can be utilized as a reason for anticipating future occasions.

KEYWORDS: Sentiment Analysis, Twitter, Segmentation, Classification, Bayesian Logistics Regression, TF-IDF.

I. INTRODUCTION

In the previous decade, new types of correspondence, for example, micro blogging and message informing have risen and turned out to be omnipresent. Twitter is a standout amongst the most widely recognized online web-based social networking and miniaturized scale blogging administrations which are an extremely well known strategy for communicating feelings and connecting with other individuals in the online world. Twitter messages give genuine crude information in the organization of short messages that express sentiments, thoughts and occasions caught at the time. Tweets and messages are short: a sentence or a feature as opposed to an archive.

While there is no restriction to the scope of data passed on by tweets and messages, regularly these short messages are utilized to impart insights and opinions that individuals have about what is happening in their general surroundings. The dialect utilized is exceptionally casual, with inventive spelling and accentuation, incorrect spellings, slang, new words, URLs, and shortenings, for example, RT for "re-tweet" and # hash labels, which are a kind of labeling for Twitter messages. Another part of web-based social networking information, for example, Twitter messages is that it incorporates rich organized data about the people required in the correspondence. For instance: Twitter keeps up data of who takes after whom and re-tweets and labels within tweets give talk information. When taken in accumulation tweets can give an impression of open assessment towards occasions, proposed framework give a positive or negative estimation on Twitter posts utilizing an outstanding machine learning strategies for content order, called Bayesian Logistic Regression [BLR] and Naive Bays Classification for giving positive or negative conclusion on tweets.

Twitter, a standout amongst the most widely recognized online web-based social networking and smaller scale blogging administrations, is an exceptionally well known technique for communicating conclusions and interfacing with other individuals in the online world. Twitter messages give genuine crude information in the arrangement of short

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

messages that express assessments, thoughts and occasions caught at the time. Tweets [Twitter Posts] are appropriate wellsprings of spilling information for feeling mining and assumption extremity discovery. Sentiments, assessments, feelings and theories regularly mirror the conditions of people; they comprise of obstinate information communicated in a dialect made out of subjective expressions.

In this framework, we analyze the adequacy of an ordinarily utilized content order technique called Bayesian Logistic Regression [BLR] Classification for giving positive or negative assumption on tweets. We utilize extricated Twitter feeling to search for relationships between's this estimation and major FIFA World Cup 2014 occasions as our contextual investigation.

Tweet Collection

The first module comprises of Data gathering for the creation of a training set and then collecting tweets from a particular event. Data gathering is made up of two steps using Twitter's Streaming API: the first is collecting the data to use as a training set to build the model. The second step is collecting tweets during any particular event and processing them by filtering some of the official hash tags. In addition, the Twitter usernames of concerned personalities related to that event are used to extract tweets relating to events. The data is in JSON format as a set of documents, one for each tweet.

Tweet Pre Processing

As a first step towards finding a tweet's sentiment and in order to obtain accurate sentiment classification, we needed to filter out noise and meaningless symbols that do not contribute to a tweet's sentiment from the original text. All of the following steps had to be performed sequentially for all the tweets, in order to use them for training a model.

- ✚ Tokenization
- ✚ Removing Stop-Words
- ✚ Conversion of Twitter Symbols into generic tags.
- ✚ Stemming

Tweet Classification

Labeling an opinionated text and categorizing it overall into a positive or negative 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. We need to use all the subjective tweets, including positive or negative sentiment. There are methods of extracting the useful words in order to detect the sentiment of tweets. All of the following steps had to be performed sequentially for all the tweets.

- ✚ Feature Extraction
- ✚ Feature Filtering

Machine Learning Approaches

Below are the machine learning techniques used in sentiment classification for text categorization, which includes Naviebayes and Static virtual machine. This is used for classifying sentiment as positive or negative.

(i) Naive Bayes

Naive Bayes is basically a probabilistic approach for the classification of tweets on the basis of their polarity in terms of either a positive opinion or a negative opinion.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

Advantages

- ✚ Work well on numeric and textual data
- ✚ Easy to implement and computation comparing with other algorithms

Disadvantages

- ✚ A disadvantage is it's conditional independence which is violated by real world data,
- ✚ And perform poorly when features are highly correlated.

(ii) Support Vector Machine

SVM is one of the most accurate classifier. As far as sentiment analysis is concerned an multiclass SVM with a one vs. one scheme is considered, which creates one binary classifier for every different pair of classes, and then uses an opinion mechanism to ultimately choose one class.

Advantages

- ✚ When learning text classifiers, one has to deal with many features, since SVM use over fitting protection, which does not necessarily, depends on number of features. They have the potential to handle these large feature spaces.
- ✚ SVM eliminates the need for feature selection making the application of text categorization easier

Disadvantages

- ✚ One drawback of SVM is that if the size of the feature vector is larger than the number of training samples, it tends to over fit on the training data, which lowers the accuracy on the testing data.
- ✚ Another drawback of SVM occurs with neutral opinion, which could pose an issue to our project due to the small size of our dataset.

II. SYSTEM ARCHITECTURE

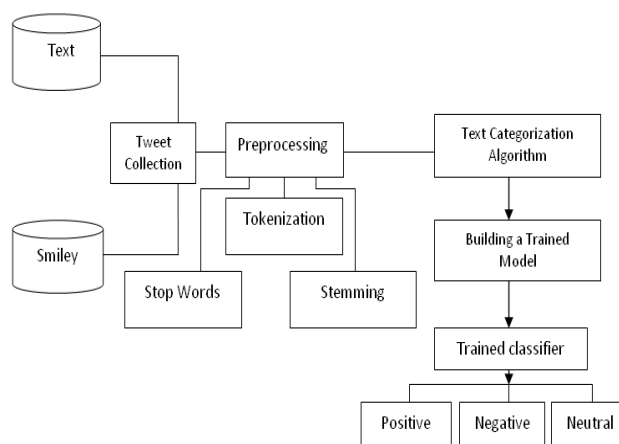


Fig. 1. Proposed System Architecture

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

III. LITERATURE SURVEY

[1] Chetan. A. Burandedescribes about the various techniques that can be used to pass information between the human and the computer. This deals with how to transfer signals to the computer through human interaction.

[2] G.R.S.Murthy tells about the advancement in computing and how neural networks can be used to recognize human hand gestures.

[3] HaticeGunes is a paper that describes facial expression detection and hand gesture recognition to interpret actions of a human.

[4] K.Brahmani presented about how a robotic arm can be controlled by electronic unit that recognizes the gestures made by the user.

[5] A. Syed, Z. Agasbal is one of the papers which used a flex sensor to control the robotic arm. The interface between these two was a microcontroller.

[6] JagdishLal Raheja. It is based on the control of the robotic hand. The hand is controlled by the images captured of the user's hand gestures.

[7] R. V. Dharaskar describes the control of robotic arm using both voice recognition as well as gesture recognition.

[8] Ron Oommen Thomas in his paperhas implemented the control of robotic arm with visual feedback.

[9] ShamsheerVerma, in his paperpresents an idea of controlling the robotic arm wirelessly by using hand gestures.

[10] Vicky Somkuwar, in his paper presents the robot can be controlled using an accelerometer which can detect the change in angle of the remote by reference to the gravitational center.

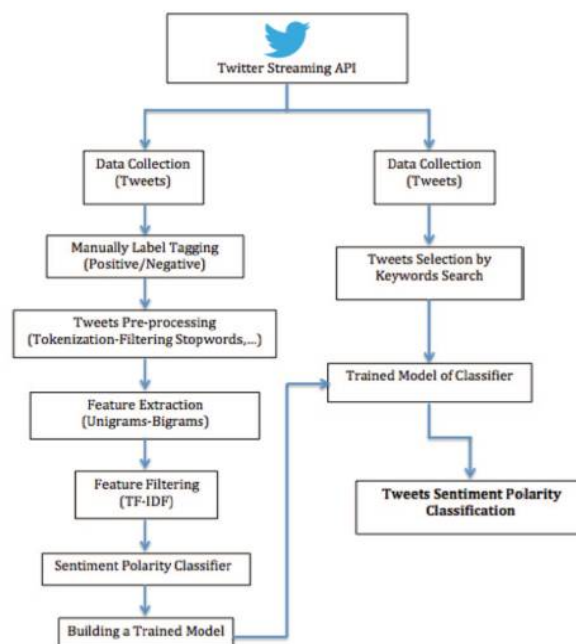


Fig.2. Building a trained model for polarity detection and applying it to some tweets

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

III. EXISTING APPROACH

Twitter is an online social networking service that enables users to send and read short characters messages called tweets Register users read and post tweets but unregistered users can only read them. A social networking service is a platform to build social networks among people who share similar interests, activities, backgrounds or real-life connections. A social network service consists of a representation of each user, his or her social links, and a variety of additional services such as career services.

Social network sites are one of the web-based services using this one can create a public profile, create a list of users with whom to share connections, and view and cross the connections in the system. A social networking service allows users to interact over the Internet, such as email and instant messaging. A social network site provides new information and communication tools like mobile connectivity for photo/video/sharing and blogging. Online community services are one of the social networking service, though it is a social networking service it is an individual-centered service but in case of online community services those are group-centered. Because of Social networking sites one can share ideas, pictures, posts, activities, events, and interests with people in their network.

Disadvantages

- ✚ A message can be sent tweet followers only.
- ✚ Fake messages can be possible.

IV. PROPOSED STUDY

Proposed system will identify the rumors in twitter. The proposed system will fetch the tweets from Twitter according to search criteria and processes the tweets, by applying text mining techniques and finally the system will performs the classification of tweets.

Advantages

- ✚ Possibility of fake messages will be removed.
- ✚ Admin will Monitor the user messages.

Datasets Used

Positive words Dataset for Tweet Analysis

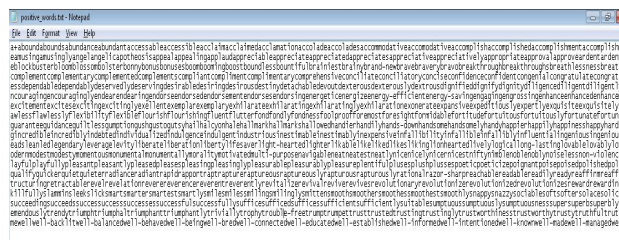


Fig.3. Dataset – Positive Words

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

Negative Words Dataset for Tweet Analysis

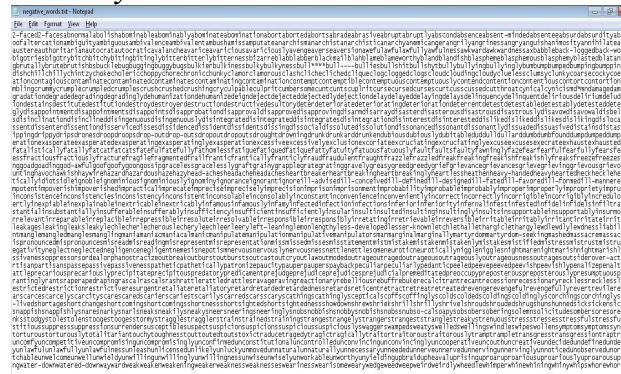


Fig.4. Dataset – Negative Words

VI. EXPERIMENTAL RESULTS

Data Collection

This module explains about how to collect tweeter datasets from multiple access Convert tweets into data sets in the form of XLS. When a tweet Message is collected by admin side, the messages are checked. Acquisitions are made up of purchase outstanding, but many companies use their own stock by exchanging it for the stock. Many of the times social media data collection is considered as tacit knowledge with the collection of tweets which are reduced to a single sentence without explanation to what it means, mechanisms or relative merit of the approach. This techniques for the capture of Twitter timeline data, inclusive of first person and third party methods for data capture from personal accounts, public accounts, and keyword searches. The acquiring Twitter data with a focus on individual timelines, and small to medium scale search sets. The main goal is to be able to obtain the twitter data , examine and convert it into knowledge which will require limited technical skills. This data collection assumes no prior programming knowledge. The explains how Twitter data can be retrieved from the these three sources:

- 1) personally controlled timelines,
- 2) third party timelines,
- 3) ongoing search results.

This explains how preliminary analysis can be used to perform ascertain content creation patterns, without any recourse to analysis of individual tweets.

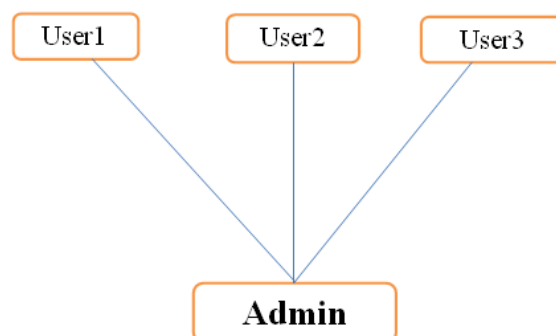


Fig.5. Data Collection

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

Data Preprocessing

Data pre-processing includes cleaning, normalization, transformation, feature extraction and selection, etc. This module, implement noise removal steps to eliminate stop words And using stemming analysis approach to detect stemming words.

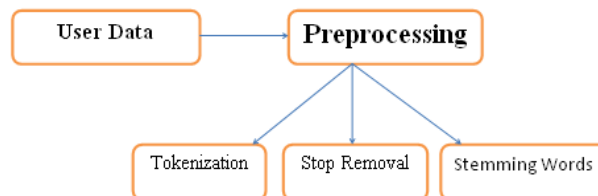


Fig.6. Data Preprocessing

Hybrid Segmentation

In stemming words Global and local contexts are constructed using key terms. Hybrid Segmentation uses both global and local contexts, and has the ability of learning from pseudo feedback.

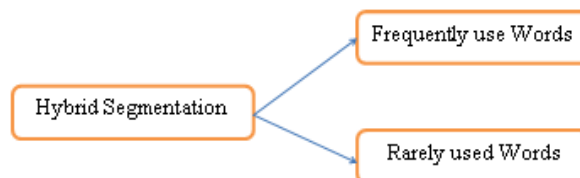


Fig.7. Hybrid Segmentation

Named Entity Recognition

These implements the POS tags of the constituent words of the segments first and then implement classification to classify each named entities.

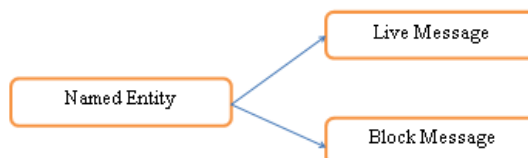


Fig.8. Named Entity Recognition

Rumor Identification

This module predict the rumors based on text matching classifier and then implement the text matching approach to classify the rumors by using three features such as content features, network features, blog features. Text matching classification approach is used to label the each tweets.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

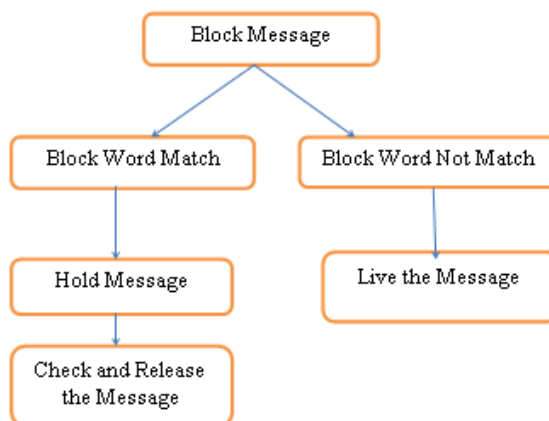


Fig.9. Rumor Identification

VII. CONCLUSION

Sentiment analysis and opinion mining has wide area of applications and also facing many research challenges. As the internet and internet related applications are used widely, opening mining and sentiment analysis become a most interesting research area. So that we require more innovative and effective techniques which should overcome the current challenges faced by opinion mining and sentiment analysis, Proposed system will try to overcome these challenges by using BLR.

REFERENCES

- [1] Chetan A. Burande, Raju M. Tugnayat, NitinK.Choudhary(2010), Advanced Recognition Techniques for Human Computer Interaction, 2nd International Conference on Computer and Automation Engineering (ICCAE), 2010, Singapore. Pages 480-483.
- [2] G.R.S. Murthy, R.S. Jadon, (2010) "Hand Gesture Recognition using Neural Networks", Advance Computing Conference (IACC), 2010 IEEE 2nd International Conference, Pages 19-20 Feb. 2010, Patiala.
- [3] HaticeGunes, Massimo Piccardi, Tony Jan(2004), "Face and Body Gesture Recognition for a Vision-Bases Multimodel Analyzer", Conferences in Research and Practice in Information Technology, Vol 36, 2004.
- [4] K. Brahmani, K. S. Roy, Mahaboob Ali, (2013), "Arm 7 Based Robotic Arm Control by Electronic Gesture Recognition Unit Using Mems", International Journal of Engineering Trends and Technology, Vol. 4 Issue 4
- [5] A. Syed, Z. Agasbal, T. Melligeri and B. Gudur, (2012) "Flex Sensor Based Robotic Arm Controller Using Micro Controller," Journal of Software Engineering and Applications, Vol. 5 No. 5, 2012, pp. 364-366.
- [6] Jagdish Lal Raheja, Radhey Shyam, Umesh Kumar, Bhanu Prasad (2010), "Real-Time Robotic Hand Control using Hand Gestures", Second International Conference on Machine Learning and Computing, 2010.
- [7] R. V. Dharaskar, S. A. Chhabria, Sandeep Ganorkar (2012), "Robotic Arm Control Using Gesture and Voice", International Journal of Computer Information Technology & Bioinformatics (IJCITB), Vol. 1, Issue 1, pp. 41-46.
- [8] Ron Oommen Thomas, K. Rajasekaran, (2014), "Remote Monitoring and Control of Robotic Arm with Visual Feedback using Raspberry Pi", International Journal of Computer Applications, Volume 92 – No.9, April 2014.
- [9] Shamsheer Verma, (2013), "Hand Gestures Remote Controlled Robotic Arm ", Advance in Electronic and Electrical Engineering (AEED), ISSN 2231-1297, Volume 3 Number 5 (2013), pp 601-606.
- [10] Simone Zimmermann, Sonja Rümelin, Andreas Butz, (2014), "I feel it in my fingers: haptic guidance on touch surfaces", TEI '14 Proceedings of the 8th International Conference on Tangible, Embedded and Embodied Interaction, Pages 9-12, ACM New York, NY, USA
- [11] Peiman Barnaghi, and John G. Breslin, "Opinion Mining and Sentiment Polarity on Twitter and Correlation Between Events and Sentiment". 2016 IEEE Second International Conference on Big Data Computing Service and Applications.
- [12] P. Barnaghi, P. Ghaffari, and J. G. Breslin, "Text Analysis and Sentiment Polarity on FIFA World Cup 2014 Tweets," in Conference ACM SIGKDD, 2015.
- [13] 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.
- [14] A. Bifet and E. Frank, "Sentiment knowledge discovery in twitter streaming data," in Discovery Science, 2010, pp. 1-15.
- [15] A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," in LREC, 2010, pp. 1320-1326
- [16] A. Genkin, D. D. Lewis, and D. Madigan, "Large-scale Bayesian logistic regression for text Categorization," Technometrics, vol. 49, pp. 291-304, 2007.