SENTIMENT ANALYSIS BASED ON TWITTER DATA ON VIOLENCE

NIHAL JUMHARE, RAJA RAJESWARI G, BALAJI JAYAKRISHNAN

Department of CSE, VIT University, Chennai, Tamil Nadu, India. Email: asha.s@vit.ac.in

ABSTRACT

Due to the large volume of opinion-rich web resources such as Twitter, Facebook, blogs, and news available in digital form, and much of the current research is focusing on the area of sentiment analysis using text analysis. People are getting attracted to develop a system that can extract opinions based on their response on social media sites. Algorithms can be developed so as to predict preferences of people to improve economic and marketing research. This paper presents a sentiment analysis on a recent scenario of Uri Attack.

Keywords: Rich web resources, Text analysis, Uri attack.

INTRODUCTION

Sentiment analysis is a technique for tracking the mood or sentiment of the public for a particular product or incident. Sentiment analysis is also called as opinion mining, and it involves in making a machine that collects and analyze the collected opinions through blogs, comments, reviews, or tweets. Sentiment analysis can be used in various ways for different analysis. For example, it helps in marketing to predict the success rate of the existing or new products and decides which versions of a product are a success and even identify which features are like or disliked.

CHALLENGES IN SENTIMENT ANALYSIS

The first is an opinion word that is considered to be positive in one situation may be considered to be negative in another situation. A second challenge is that people do not always convey their view in a similar way. The opinion may be varying from person to person. Most traditional text processing relies on the fact that small differences between two pieces of text do not change the meaning very much. In sentiment analysis, however, "the cake was good" is very different from "the cake was not that good." People can be differing in their opinions. Most of the reviews have both pros and cons, which can be managed by analyzing sentences one at a time. However, in the more informal medium such as Twitter or blogs, the many people are like to combine different opinions in the same sentence which is easy for a human to understand, but it is more difficult for a computer to parse. At times, even few people have difficulty in understanding what someone think based on a short piece of text due to the absence of context. For example, "that cake was as good as the previous cake" is entirely dependent on what the person expressing the opinion thought of the previous model.

The user's hunger is on for and dependence upon online advice and recommendations the data reveals is merely one reason behind the emergence of interest in new systems that deal directly with opinions as a first-class object. Sentiment analysis concentrates on attitudes whereas the traditional method of text mining mainly focuses on the analysis of facts. There are few main fields of research in sentiment analysis: Sentiment classification, feature-based sentiment classification, and opinion summarization. Sentiment classification deals with classifying entire document according to the opinions shown with respect to a certain object. While feature-based sentiment classification considers the opinions on the features of a certain object. Opinion summarization, task is different compared to traditional text summarization because only the features of the product are examined on which the customers have expressed their opinions through any social media. Opinion summarization does not summarize the reviews by selecting a subset or rewriting some of the original sentences from the reviews to capture the main points like in the classic text summarization method.

Languages that have been studied mostly for sentiment analysis are English and Chinese. Till date, there are very few researches that have conducted sentiment classification for other languages such as Arabic, Italian, and Thai. This analysis aims at focusing much of the work in English.

The beginning of sentiment analysis dates back to late 1990's, but becomes a major emerging subfield of information management discipline only from 2000, especially from 2004 onward, which this survey focuses because of the raise in social media text content.

The rest of the paper is organized as: Section 2 presents the related work, Section 3 presents proposed work, Section 4 presents data source, Section 5 presents sentiment classification, Section 6 presents software and tools used in the analysis, Section 7 presents evaluation and discussion of the results obtained, Section 8 presents conclusion of the analysis, and Section 9 presents the references taken for successfully accomplishing the work.

RELATED WORK

Sentiment analysis also knows as opinion minion is growing as a good field of research in which the opinions of people are taken from any means such as social media, news, blogs, and review websites. so as to predict the sentiment about any incident, movie, product, etc. A lot of work has been published supporting the sentiments in different areas.

Twitter has been the point of attraction to several researchers in important areas such as prediction of democratic electoral events, consumer brands, movie box office, stock market, and popularity of celebrities. Sentiment analysis over Twitter offers a fast and efficient way of monitoring the public sentiment. In this paper, the novel approach of exploiting the user influence factor to predict the outcome of an election result is introduced. A hybrid approach of extracting opinion using direct and indirect features of Twitter data based on support vector machines (SVM), Naive Bayes, maximum entropy, and artificial neural networks-based supervised classifiers is introduced. Principal component analysis with SVM in an attempt to perform dimensionality reduction is combined. This paper shows two different case studies of entirely different social scenarios, US Presidential Elections 2012 and Karnataka Assembly Elections 2013 [2].

In this paper, the utility of linguistic features for detecting the sentiment of Twitter messages is investigated. The usefulness of existing lexical resources as well as features that capture information about the informal and creative language used in micro-blogging is evaluated. A supervised approach is taken to the problem, but leverage existing hash tags in the Twitter data for building training data [3].

Measures of public opinion measured from polls with sentiment measured from text are connected. Several surveys on consumer confidence and political opinion over the 2008-2009 period is analyzed, and found whether they correlate to sentiment word frequencies in contemporaneous Twitter messages. While the results vary across datasets, in several cases the correlations are as high as 80% and capture important large-scale trends. The results highlight the potential of text streams as a substitute and supplement for traditional polling [4].

PROPOSED WORK

The work proposed to complete in this paper is the sentiment analysis of the recently incident Uri Attack through text analysis. R language is used along with RStudio as a tool to achieve the same and obtain the corresponding sentiment graph.

DATA SOURCE

User's opinion is a major criterion for the improvement of the quality of service of products and expressing views on a particular incident. A lot of social media such as Facebook, blogs and review sites are available. However, among them, a major source is Twitter which is, especially build to paint, a user's sentiment of any incident happening or for a response to a tweet shot by someone. In this analysis, more than 4500 tweets are being analyzed so as to get a sentiment of Uri attack.

SENTIMENT CLASSIFICATION

Much number of researches exists on sentiment analysis of user opinion data, which mainly judges the sentiment of the reviews given. There are mainly two types of techniques to perform the analysis: Machine learning and semantic orientation. Natural language processing is mainly done in sentiment orientation method. Hence, sentiment detection is basically a crossroad of the natural language processing and information retrieval.

The sentiment orientation approach is basically an "unsupervised learning" because it does not require a prior training to analyze the data. Instead, it measures how far the words tend toward positivity or negativity. The words are being mapped to the local set of words carrying the respective sentiment to map the tweet and get the respective sentiment of the tweet. Although machine learning approach gives more accuracy, the semantic orientation approach is light weighted method and gives the respective sentiments easily. Although may not be perfect, it gives an approximate sentiment of the event happened.

SOFTWARE AND TOOLS

The tool used here for the analysis of sentiment is RStudio and the language used to fetch tweets is R. Twitter API is used to fetch tweets. RStudio is a sophisticated tool having numerous packages so as to have a wide variety of work on cloud. It is also very useful for analyzing and plotting graphs; complex algorithms can be implemented with ease with this tool. Although the same analysis can be performed by various other languages like python, this is best in terms of representation of the received results.
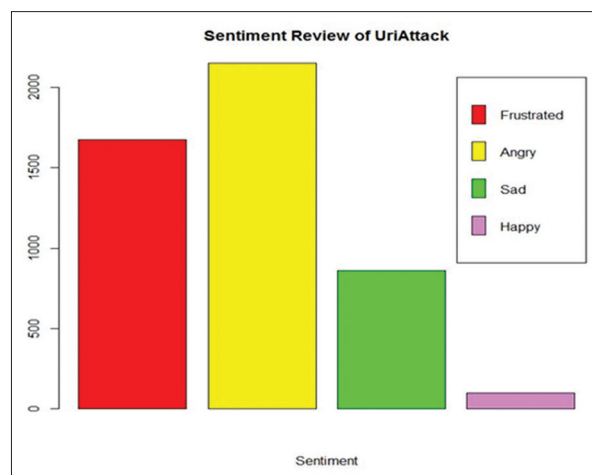

Fig. 1:

EVALUATION AND DISCUSSION

Text mining is a challenging task than other tasks as the challenges of text mining lies in the factual information related to real life and the ironic words used in writing the reviews. People may such as some happenings and some may not. Thus it consists of positive opinion. However, as far as happenings concerns people may show variety of emotions such as anger, sad, happy, and furious. This is what makes the sentiment analysis more complex than others.

In the sentiment, result provided below shows the result extracted from 4673 tweets and presented in form of graph. This basically shows the sentiment of the people on the recently happened "Uri Attack." Large number of people presented their sentiments among which a random sample is taken to come to a conclusion (Fig. 1).

CONCLUSION

Thus, the sentiment of the Uri Attack incident is analyzed and is shown in the form of graph. The large height of red and yellow bars shows that the people in the community are very frustrated and angry about the incident. Although a good height of green shows a good strength of people have soft heart as they are sad about the happening. Although the strength of frustrated and angry people is more, it will purely leads to an environment with full of war. It will lead to a negative impact on both the society and the world. Contrary the height of green bar should be increased so as to have peace in the world of terror and violence. There are a very few people showing positive sentiments on the serious issue. The number of such people should be decreased so as to maintain a healthy environment.

REFERENCES

1. Vinodhini G, Chandrasekaran RM. Sentiment analysis and opinion mining: A survey. Int J Adv Res Comput Sci Softw Eng 2012;2:282-92.
2. Anjaria M, Mahana R, Guddeti's R. Influence factor based opinion mining of twitter data using supervised learning. In: IEEE/ACM Conference on ASONAM; 2014.
3. Kouloumpis E. Theresa Wilson and Johanna Moore. Twitter Sentiment Analysis: The Good the Bad and the OMG! Proceedings of the 5th International AAAI Conference on Weblogs and Social Media; 2011.
4. O'Connor B, Balasubramanyan R, Routledge's BT. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. Proceedings of the 5th International AAAI Conference on Weblogs and Social Media; 2010.