INTERNATIONAL JOURNAL OF EMERGING RESEARCH & DEVELOPMENT

(Volume1, Issue3)
Available online at: www.ijernd.com

Analyzing and interpreting variations of public sentiments on social network

Gitesh Mudgal
gitesh.mudgal@ves.ac.in
Vivekanand Education Society Institute Of Technology, chembur, Mumbai, Maharashtra

Rohini Temkar
rohini.temkar@ves.ac.in
Vivekanand Education Society Institute Of Technology, Chembur, Mumbai, Maharashtra

ABSTRACT

Social networks have become one of the most popular means of communication on the internet, as a result of which internet users have grown rapidly. Millions of messages are regularly available on websites that offer web services such as Facebook, Tweeter, WhatsApp and LinkedIn. Millions of users share their personal opinions or views on various topics and discuss recent news on social websites, making it an important basis for tracking and analyzing the sentimental perception of the community. A social site is just as innovative as a small blog platform with more than one million unique weekly guests. On social sites, each user placed a message with the name tweeter or blog, which is visible to everyone. Such monitoring or analysis can provide important information for making decisions and evaluating opinions in different domains. In this work, we went a step further to interpret the mood swings. We have established that the emerging topics within the periods of mood variation are strongly related to the real reasons for the variations. We propose a model based on latent Dirichlet allocation, LDA for foreground and background (FB-LDA) to distill the foreground themes and filter the age-old background themes. These foreground themes help to interpret method swings in social networks. The feeling analysis, also known as Opinion Mining, plays a crucial role in determining the feelings in different web content. The analysis of opinions is very important to make decisions. Example, if you want to buy a new phone, a competent buyer of the web will always first assess the opinions to make a purchasing decision based on other experiences. The analysis of feelings extracts opinions, feelings, and emotions from the text and analyzes them. This information is very useful for governments, companies, and individuals. Although this content might be useful for analyzing most of the content generated by the user, it is difficult and time-consuming. Sentiment analysis is the automatic extraction of opinions, perspectives, and emotions from data sources via NLP.

Keywords – *Analysis, Sentiment Analysis, Sentiments on Social Media, Emerging subject mining, Big data*

1. INTRODUCTION

With the Extensive growth of user-generated messages on the internet, a Social site like Tweeter where a huge number of users used to share their opinion regarding some topic. Diagram indicates that web has a huge amount of data and social networks have part of that Big data. With the Extensive growth of user-generated messages on the internet, Social Media like Facebook where a large number of users used to share their opinion regarding some subject. Figure 1 shows that web has the Big amount of data and social media has part of that Big data. We can

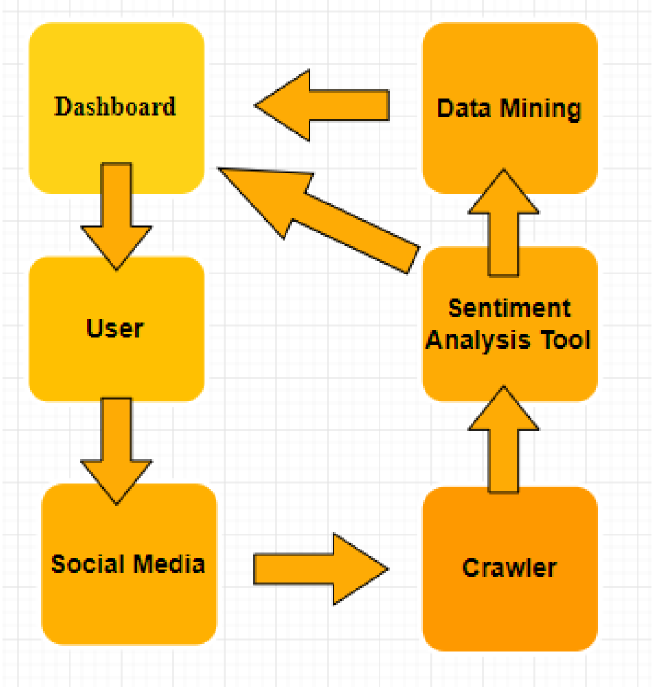

Fig. 1: Structure of Social Network Analysis

The analysis of sentiment on social websites has created a platform on which current public opinion can be uncovered in an economical and effective way, which is troublesome for decision-making in various areas. For example, a company could analyze the public sentiment in Tweeter to obtain users feedback towards its products or service; while a political leader can adjust his/her position with respect to the sentiment change of the public, opinion about movies can be most useful for a succession of movies Sentiment. The analysis is a technique for extracting sentiment associated with polarities of positivity, negativity, and neutrality. It is one of the types of natural language processing in which we can track the mood of the public about a particular entity. Sentiment analysis, which is also called opinion mining, is used for constructing analysis system to collect and examine opinions about the entities on tweets on Tweeter. Due to the tremendous of social media services, there is great opportunity to understand and analyze the sentiment of the public by analyzing its large-scale data and opinion-rich data. Sentiment analysis on Tweeter can be done by many approaches. Various types such as machine learning and lexicon-based approaches have been widely used for sentiment analysis on Tweeter like sites. Machine-learning approaches to sentiment analysis need to train the data. Searching for people opinions via surveys and polls have been an expensive and time-consuming task. The proliferation of Web 2.0 has changed the way people express their opinions and feelings. This so-called user-generated content posted in blogs, forums, product review sites and social networks is mostly publicly available and easy to obtain. The high value of this content arises from its subjective nature which, in aggregated form, indicates public opinion. It is difficult for humans to read and summarize all relevant documents in terms of the expressed sentiment. Thus, there is, a growing need for auto-analysis of this kind of data. This is a challenging task with foundations in natural language processing as well as text mining referred to as sentiment analysis. Many research studies in sentiment analysis are concerned with product reviews from websites like Tweeter is a most popular worldwide social media site, which provides a microblogging services and Social Media, enable its users to update their status in tweets, follow the people they are interested in (e.g. Ratan Tata) and retweet other‘s posts and even communicate with them directly. The public Sentimental analysis of Tweeter data has provided an economical and effective way to expose timely public sentiments, which is critical for decision making in various domains areas. For instance, a company can study the sentiments of the public in Tweets to obtain users ‘feedback towards its products. The tweeter is one of the most popular Social Media websites, which is drawing more and more attention from researchers from different disciplines. There are several streams of research investigating the role of Tweeter. The tweeter has attracted attention in both academia and industry for Research Area. Last research mainly focused on tracking public sentiment. There have been a large huge of research studies and industrial applications in the area of public sentiment tracking and modeling. Last research like O‘Connor [1] focused on tracking sentiment on Tweeter and studying its correlation with consumer confidence and presidential job approval polls. On Tweeter, any user can publish a message referred to a tweet, which is visible on the public display. Similar kinds of studies have been done for investigating the reflection of public sentiments on oil price indices and share markets. They reported that events in real indeed have a significant and immediate effect on the public sentiment on Tweeter. One valuable analysis is to find possible reasons behind sentiment variation, which can provide important information for decision-making. Eg. If negative public sentiment towards Barack Obama increases significantly, the White House Administration Office may be eager to know why people have changed their opinion and then react accordingly to reverse this trend. Another example is, Analyzing public opinion variation polling for Exit poll for any Election.

2. LITERATURE S URVEY

Several researchers conducted research on the analysis of social networks and sentiment analysis. SA is a word processing technique for deriving an opinion or intention based on the terms used in real-world language. The number of researchers focused on generating statistical inferences from social network data using sentiment analysis models. Bo Pang and Lillian Lee [2] gave an idea of the complete discussion of the emotional analysis. In doing so, they considered the contribution in positive words and incomplete words in order to appreciate the opinion. Today's users can easily get information, but they can also actively generate content. News, BBS, forums, and blogs are the main sources of public opinion information. The text of these sources can contain both facts and opinions that can be extracted using natural language processing mechanisms. Opinions are generally subjective expressions describing people's feelings or feelings towards entities and events. It is a sub-discipline of computational linguistics that focuses on extracting the opinions of people from the Internet. Social media technologies can take various forms, including social networks, blogs, blogs, magazines, Internet forums, social blogs, photos, blogs, comments, and social bookmarking. Microblog sites have become sources of various types of information. Because of the nature of microblogs where people publish real-time reports about their opinions on a variety of topics, they discuss current issues, negative feelings and positive positives for employees, events or products they use in their daily lives. The manufacturers of these products have begun to question these microblogs in order to get an idea of public opinion about their product. Public and private opinions on various topics are expressed and disseminated by various social networks. Sentiment analysis is used to determine the writer's attitude about a topic. The attitude can be your judgment, intentional emotional communication or the emotional state of the author in writing. A basic task in the analysis of feelings is the classification of the polarity of a given text in the word, phrase, document whether the opinion expressed in a word, phrase or document the set of characteristics has positive, negative or neutral. The classification of feelings is seen as an example, emotional states such as "happy", "angry", "sad" and "neutral" from. Sentiment analysis has become popular when evaluating consumer sentiment towards different brands. The way in which consumers express their opinions on social networks helps to judge that opinion [2]. When it comes to feelings, opinions or emotions, we are not interested in the subject of the text, but in the positive or negative opinion. People can express their views on social media such as blogs, microblogs, comments, discussion forums and social networking sites to people, events, products, services, news or free organizations. All these platforms are the source of a large amount of valuable information that we want to analyze. Several previous studies have estimated and used the aggregate textual feeling. The informal study by Lindsay focuses on the lexical induction in the construction of a data Sentiment classificatory owner of Facebook Publications (a means of web conversation / micro-blog that is very similar to Tweeter) and shows the correlations of several surveys that during A Part of the 2008 presidential election took place. No other research is known that we validate a text comment against traditional opinion polls, although a number of companies offer essentially a text analysis perception for this purpose (eg B. Nielsen BuzzMetrics). There is at least use time series of mood added text message or some other studies, including analysis of the behavior of actions based on the text of blogs, good and bad news (Gilbert and Karahalios 2010), news (belts and Shtrimberg 2004; Lavrenko et 2000) and Investors Message boards (Antweiler and Frank 2004; Das and Chen 2007). Dodds and Danforth (2009) use an emotional word counting technique for the purely exploratory analysis of multiple companies. Tweeter makes it a valuable platform to track and analyze public opinion. It provides information for decision-making in various areas related to the current problems of society. In this work, we interpret mood changes in different topics of society. A real issue within the period of mood variation refers to the true reasons behind the variations. Based on this observation, Latent Dirichlet Assignment (LDA FB) to distill (LDA) based model, foreground, and background foreground LDA subjects. It filters the long-standing background issues. The foreground themes can provide interpretations of mood swings. This proposed system selects the most representative tweet data for priority issues and developed a model called Candidate The reason and the other generative model called Background LDA (RCB-LDA) to be re-evaluated within the period of variation of its popularity. Models based on Deferred Dirichlet Allocation (LDA) to analyze Tweets in periods of significant variation and to identify possible causes of variation. This model can be referred to as L1 foreground and background (FB-LDA) mode 1 and can filter the background themes and extract the foreground topics from the tweets in the specified variation period by using a set of additional background twids that will generate just before the variation was. Candidate Reason and LDA Background (RCB-LDA). RCB-LDA first extracts representative foreground tweaks (obtained from FB-LDA) as a basic candidate. After that, you will associate each remaining Tweet in the variation period with a Basic Candidate and rank the candidates according to the number of Tweets assigned to them. There are many articles describing various classification techniques for the analysis of emotions. The classification of emotions can be formulated as a supervised problem with two class designations (positive and negative). In (Pang, Lee, and Vaithyanathan 2002), the authors use supervised learning methods, such as the naive Bayesian and Support Vector (SVM) machines, to classify film reviews into two classes. Most unsupervised feelings of classification attempts attempt to generate a general opinion or domain-dependent lexicon for opinion words or sentences. In (Riloff and Wiebe 2003) the authors have compiled references to subjectivity as part of their work. The clues were used in (Wiebe, Wilson and Cardie 2005) to detect semantic orientation. In this document, a start-up process was proposed in which high-precision classifiers use the familiar subjective vocabulary to separate subjective and objective sentences from a text collection without annotations. The Aspect Extraction method refers to the concept of determining goals of expression and their attributes that are mentioned in a document or sentence. Many information extraction techniques have been used so far. (91%)

3. PROPOSED METHODOLOGY

A. Common Architecture

Nowadays, almost all social networks are widely used to express opinions or emotions in public with the help of the Internet. And Tweeter was the magnet for several researchers in important areas. The sentiment analysis of Tweeter provides a quick and efficient way to analyze public opinion. The most important double contributions of this document are:

(1) Our research and knowledge are the best work of our study, which attempts to analyze and interpret mood changes in microblogging services such as Tweeter.

(2) Two new generative models are being developed to solve the basic problem of mining. The two proposed models are generic: they can be applied to other tasks, such as: For example, to find thematic differences between two sets of documents.

B. Suggested Architecture

In our work, we have proposed following three steps for sentiment tracking we extract tweets related to our interesting targets and preprocess the raw extracted for more Cleaned for sentiment analysis.

Second, we assign some label so-called sentiment label for every individual tweet by combining two state-of-the-art sentiment analysis tools.

Finally, depending upon the sentiment labels obtained for each tweet, we identify the sentiment variation for the corresponding targeted issues by using some descriptive statistics.

4. MODULES

A. Tweets Extraction and Preprocessing

Our First phase starts with extracting tweets lines related to the targeted issue, we go through the whole collected raw dataset and extract all the core lines tweets which contain the keywords of the targeted issues. Compared with regular text documents, tweets are generally somewhat informal and often written in an ad-hoc manner like it may contain short forms, some abbreviation. Sentiment analysis can tools applied to raw tweets but often achieve very poor performance in most cases. Hence there is need of preprocessing techniques on tweets are necessary for obtaining satisfactory results on sentiment analysis:

1) Slang words translation: The most common Tweets often contain a lot of slang words (e.g. lol, omg). These words are usually very important for sentiment analysis, but may not be included in root sentiment lexicons. Since the sentiment analysis is based on sentiment lexicon, therefore we are converting these all slang words into their standard forms using the Internet Slang Word Dictionary and then re-add them to the tweets.

2) Non-English tweets filtering: Since the sentiment analysis tools to be used only work for English texts, we remove all non-English tweets in advance as this non-English word doesn't have meaning for sentiment. A tweet could be treated as non-English tweet if more than 20 percent of its words (after slang words translation) does not appear in the GNU A spell English Dictionary.

3) URL removal: A lots of users may include various URLs in their tweets. These URLs may complicate our sentiment analysis process. So we decide to remove URLs from tweets. (100%)

B. Sentiment Label Assignment

For assigning sentiment labels for each tweet more confidently, we sort lexicons again to two state-of-the-art sentiment analysis tools. One is the SentiStrength tool [4]. This tool is based on the LIWC [1] sentiment lexicon. It works in the following way: first assign a sentiment score to each word in the text according to the sentiment lexicon; then choose the maximum positive score and the maximum negative score among those of all individual words in the text; compute the sum of the maximum positive sentimental score and the maximum negative sentimental score, denoted as Final sentimental Score; finally, use the sign of Final Score to indicate whether a tweet is positive, negative or it is neutral.

Lexicon based Techniques

In the unsupervised technique, classification is done by comparing the features of a given text against sentiment lexicons whose sentiment values are determined prior to their use. Sentiment lexicon contains lists of words and expressions used to express people's subjective feelings and opinions. For example, start with positive and negative word lexicons, analyses the document for which sentiment need to find. Then if the document has more positive word lexicons, it is positive, otherwise, it is negative. The lexicon based techniques for Sentiment analysis is unsupervised learning because it does not require prior training in order to classify the data.

The steps of the lexicon based techniques are below

1. Preprocess each raw tweet text (i.e. remove HTML tags, noisy characters)
2. Initialize the total text sentiment score: s _ 0
3. Tokenize text. For each token, check if tokens are present in a sentiment dictionary of the training set.
   If token is present in dictionary,
   If the token is positive, then s _ s + w.
   If the token is negative, then s _ s − w.
4. Look at aggregate text sentiment score s,
   If s > threshold, then classify the text as positive (b) If s < threshold, then classify the text as negative.

5. CONCLUSIONS

Overall, we conclude that social network based behavioral analysis parameters can increase the prediction accuracy. However, the presence of all the entities in unbiased and equal manner is necessary to provide accurate results. In this paper, we investigated the problem of analyzing public sentiment variations and finding the possible reasons causing these variations. we proposed two Latent Dirichlet Allocation (LDA), based models, Foreground and Background LDA (FB-LDA) and Reason Candidate and Background LDA (RCBLDA). These foreground topics can give potential interpretations of the sentiment variations. we have selected the descriptive tweets for foreground topics and develop another generative model called Reason Candidate and Background LDA (RCB-LDA) to rank them with respect to their ―popularity within the variation period. The FB-LDA model can filter out background topics and then extract foreground topics to reveal possible reasons. To give a more spontaneous, representation of the RCB-LDA model which can rank a set of reason candidates expressed in natural language to provide sentence-level reasons. The proposed models are general: they can be used to discover special topics or aspects in one text collection in comparison with another background text collection. Also, our proposed models evaluated on real Tweeter data

6. REFERENCES

[1] O'Connor, R. Balasubramanyan, B. R. Rutledge, and N. A. Smith, "From tweets to polls: Linking text sentiment to public opinion time series," in Proc. 4th Int. AAAI Conf. Weblogs Social Media, Washington, DC, USA, 2010.

[2] Bo Pang. Lillian Lee, "Seeing Stars: Exploiting class relationships for sentiment categorization with respect to rate scales", 2002

[3] J. Bollen, H. Mao, and X. Zeng, ―Tweeter mood predicts the share market," J. Computer. Science, vol. 2, no. 1, pp. 1–8, Mar. 2011.

[4] J. Bollen, H. Mao, and A. Pepe, "Modeling public mood and emotion: Tweeter sentiment and social-economic phenomena", AAAI Conf. Weblogs Social Media, Barcelona, Spain, 2011.