International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Elucidating the Public Sentiment Fluctuations on Twitter

Bhagyashree Kolte

Computer engineering Department, Pune University, Pune, Maharashtra, India

Abstract: ***Now a day's numbers of people share their opinion on different aspect of life every day. So, twitter is used as microblogging platform for communicate with each other and also tracking and evaluating public emotions. It facilitates the details for making decisions in different domain. It is an attraction for organization and in academics field. Antecedently, research concentrated modelling and tracking public emotions. Here we proceed for the next step that is elucidating sentiment fluctuations. We have to take recent topic within the sentiment fluctuation periods which are strongly linked to the reason behind fluctuations. Grounded on this, we propose a Latent Dirichlet Allocation (LDA) base model. Foreground and Background LDA (FB-LDA), take the recent (foreground) topics and separate out long-lasting (background) topics. These recent topics give the possible elucidation of sentiment fluctuations. For better readability, we choose the illustrative tweets for recent topics and trained another productive model addressed Reason Candidate and Background LDA (RCB LDA).For ranking this model according to their popularity within the fluctuation period. This method can detect recent topics and rank reason candidates.***

Keywords*:* Emotion evaluation, Public emotions, recent topic, Latent Dirichlet Allocation.

1. Introduction

Emotion analysis is also called opinion mining, is the great extraction of people sentiment, feelings and what exactly he/she thinking that is known by given tweets. There is a fantabulous survey on this subject (Pang, 2008). Emotion can be evaluated at different levels: document, section, paragraph, and sentence, from all of these document levels is very usual. There are studies on general sentiment from standard documents and related to twitter sentiment (Go et al., 2009; Pak & Paroubek, 2010). And evaluate the polarity of company's particular product reviews and cinema reviews at the documentation level. There are two common methods in sentiment analysis that are statistical natural language processing and machine learning.(1)With natural language processing, the opinion polarity of a sentence or a document is specified by a set of indicative opinion words and an opinion lexicon, that press out positive or negative sentiment such as "good" or "bad". (2)The machine learning approach is used to form a sentiment classifier that is free-based on manually labeled training data that is used to forecast the class of sentiment. Experts getting huge amount of training data that is not manageable and also judgment of the sentiment in tweets is not exact like an automated approach. To defeat these difficulties, combination of these two techniques (Lu &Tsou, 2010; Tan et al., 2008).

For collecting sentiment more efficiently, Machine learning techniques were built. Thus, sentiment classification was more challenging (Pang, 2008), because they have not predict on sentiment as on traditional topic based classification. For improvement of the machine learning approach's performance Novel technique is used. It is used to finding minimum cuts in graphs for taking related information from the document (Pang, 2008). Thus taking out the unrelated text by keeping the related portion (Pang et al., 2002).There is character limitation in Twitter tweet that is 140 characters. So, Twitters have used truncated and jargon expressions to overwhelm this limit. Here they have made different flavor language. It also has featured for handling variety of topics unlike other blogging sites (one or few topics) and contain shorter and more mislaid than others. Therefore, it is not straight to get the sentiment of tweets. Citizenry generally use Twitter for everyday conversation, opinion sharing, and for getting news (Java et al., 2007).Previous study said that 19 percent of texts refer a several brand, and 20% of them have a sentiment (Jansen et al., 2009).In general, these massages could be separated into two groups: about Twitter users they and information sharing. In both groups, text having information about the temper (mood) of their writers (MorNaaman&Boase, 2010). The human mood have six dimensions (tension, depression, anger, vigor, fatigue, confusion). This mood patterns took from Twitter data which are related to actual offline events, Like, changes in the stock market and the petrol or diesel price, and the regarding election (Bolle et al., 2011).

For discovering the association between public opinions from polls and the sentiment from Twitter messages, positive and negative words were defined by a particular lexicon, a group of words having 1600 positive words and 1200 negative words (O'Connor et al., 2010). A message was specified as positive if it carried any positive word, and negative if it carried any negative word. We can say that messages are both positive and negative. Twitter data was extremely correlated with the polls. By using Twitter data, the fundamental components in the applications is Sentiment detection of tweets. Some sentiment tools developed for Twitter data like Twitter Sentiment, TweetFeel, Viralheat and Twendz, but still these are not giving accuracy because of unique characteristics of tweets.

When students twits on twitter regarding their final exams, at this time we can assemble real-time sentiment of students flow of text messages which are related to final exam, that are analyzed. Our concern is to find the variations in sentiment about final exams from this particular group of Twitter user that is by week, day, hour, thus, at different levels of temporal granularity, the investigation could expose the sentiment. It is possible to evaluate sentiment for this topic with the use of Twitter Stream API. Twitter provides a place for student to show their opinion regarding final exams. So students have feeling more confidence for the coming final exams, because they have prepared very well, otherwise there is a possibility of feeling uneasy, afraid, and anxious, scared. And they can show the feelings to the instructors and grades from these exams.

2. Twitter Data

Twitter is used as micro blogging platform for communicate with each other by posting short messages. This vast amount of messages, generated by user having real subjective information for computational analysis. To get exact public opinion for several subject and result, Twitter data is used as suggested by recent research.

By using Twitter Stream API, we got 260,749 tweets from student regarding final exams on Twitter in two sequential week (Oct 17-Oct 30, 2011).The intention was to look into the actual-time twitter sentiment on final exam are differentiate by hour, day, and week. For these two weeks, an opinion lexicon (vocabulary) increased by sentiment predictor for this particular domain. At different levels of temporal granularity, the analysis discovered the fluctuation of sentiment. The average sentiment of the first week (Oct 17-23) was more negative and the average sentiment of the second week is less negative (Oct 24-30).the overall trend curves of sentiment increased from Monday to Sunday. However, for each and every weekdays there was a maximum sentiment within a period around 9:00 am to 5:00 pm EST. For each and every weekend, maximum sentiment within a period around 5:00 am to 8:00 am and after that it will be decreased. Moreover, the observed some consistent group behavior of Twitter users based on seemingly random behavior of each individual. Each day the minimum number of tweets took –place around 5:00 am to 6:00 am and the maximum number of tweets took –place around 1:00 pm exclude Sunday. We can say that who have maximum positive sentiment, they have more friends and followers compare to negative sentiment. And for general user it is not reliable that user shared same information was given to similar number of friends and followers.

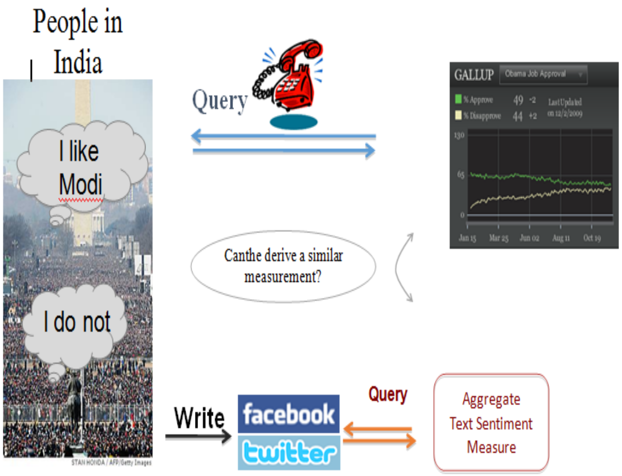

Figure 1: Measuring public opinion through social media

Twitter is a service for information network and communication, which makes over a 200 million tweets per day. It provides three APIs to access its corpus of data and support developers to establish applications using Tweets.

1) The Search API is used for Twitter content's query.
2) 2)The REST API provide the access for some primitives like, including timelines, status updates, and user information.
3) The Streaming API is the real-time sample of the Twitter Firehouse. We can extract tweets by user ids, keywords, random sampling, geographic location, etc.

This API is top for constructing data mining applications. Using Twitter Stream API (https://dev.twitter.com/docs/streaming-api) and Twitter4J (http://twitter4j.org), the collected a corpus of 260,749 tweets on final exams during a period of two sequential weeks, from Oct 17 to Oct 30, 2011. As shown in fig.1, for getting preliminary view of the tweet data, average tweet count and tweet. Lengths were calculated during these two weeks. Surprisingly, some group patterns of these student Twitters were observed from the random behavior of each case-by-case.

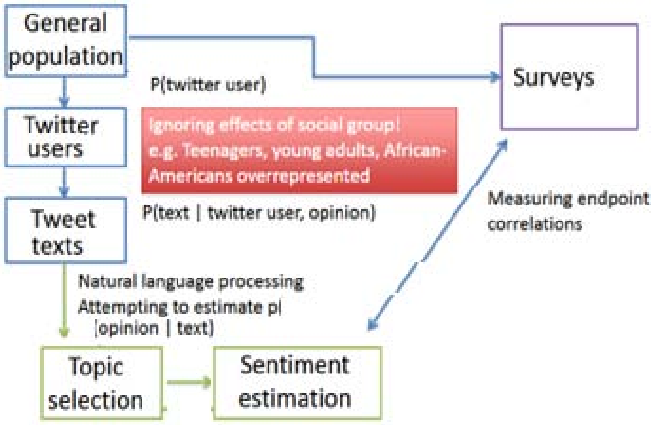

Figure 2: Block Diagram of Surveying data from twitter

3. Sentiment Predictor

In the present survey, an opinion lexicon (Hu & Bing, 2004) of around 6800 words to make the sentiment predictor is employed. Web derived lexicon like taken from this paper (Hu & Bing, 2004) could improve Lexi-con-based sentiment evaluation (et al., 2010). Considering the nature of final exams, the augmented opinion lexicon from (Hu & Bing, 2004) with some field related words such as 'bombed', and 'aced', and removed some negative words such as 'criminal', 'fall', and 'break' from this lexicon since 'criminal' can be part of the name of an exam like 'criminal justice final", 'fall' could mean fall year, and 'break' could mean a college break that students look ahead. Promoted by the results in (O'Connor et al., 2010), adopted their approach in the study to count instances of positive and negative words and emoticons, at the time of analyzing the sentiment of a tweet on final using an opinion lexicon. Looking at the characteristics of tweets, a weight +1 was assigned to a positive word, weight –1was assigned to a negative word and +5was assigned to a positive emoticon, and –5 was assigned to a negative emoticon, from that time emoticons are key not expressed in spoken words sentiment indicators in tweets. Furthermore,–5 to each suggestive word commonly used toward final exams was assigned. An opinion text may be merged with a negation word, like "no" or "not", was assigned to its inverse weight. Each tweet was decomposed into 'N' unequaled tokens (words and emotions) and opinion score was defined.

4. Conclusions

One trouble is to evaluating public opinion fluctuations and obtaining the reasons behind these fluctuations is investigated in this paper. To resolve this trouble, we propose a Latent Dirichlet Allocation (LDA) base model. Foreground and Background LDA (FB-LDA), take the recent (foreground) topics and separate out long-lasting (background) topics. These recent topics give the possible elucidation of sentiment fluctuations. For better readability, we choose the illustrative tweets for recent topics and trained another productive model addressed Reason Candidate and Background LDA (RCB LDA).For ranking this model according to their popularity within the fluctuation period. This method can detect recent topics and rank reason candidates.

References

[1] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2, 1-135
[2] Pak, A., &Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. Proceedings of The International Conference on Language Resources and Evaluation Conference, Malta, 17-23 May 2010, 1320-1326.
[3] Turney, P. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. Proceedings of the Association for Computational Linguistics, Philadelphia, 6-12 July 2002, 417-424
[4] Tan, S. B., Wang, Y. F., & Cheng, X. Q. (2008). Combining learn-based and lexicon-based techniques for sentiment detection without using labeled examples. Proceedings of the 31st ACM SIGIR Conference on Research and Development in Information Retrieval, Singapore, 20-24 July 2008, 739- 740
[5] Lu, B., &Tsou, B. K. (2010). Combining a large sentiment lexicon and machine learning for subjectivity classification. Proceedings of the International Conference of Machine Learn ing and Cybernetics, Qingdao, 11-14 July 2010, 3311-3316.
[6] MorNaaman, C.-H. L., &Boase, J. (2010). Is it all about me? User content in social awareness streams. Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work, Savannah, 6-10 February 2010.
[7] O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. Proceedings of the International Artificial Intelligence Conference on Weblogs and Social Media, Atlanta, 11-15 July 2010, 122-129.
[8] Java, A., Song, X., Finin, T., & Tseng B. (2007). Why we twitter: Understanding microblogging usage and communities. Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web mining and Social Network Analysis, San Jose, 12-15 August 2007, 56-65.
[9] Jansen, B. J., Zhang, M., Sobel, K., &Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. Journal of the American Society for Information Science and Technology, 60, 2169-2188. [1].