Volume 5, Issue 7, July 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

# A Novel Approach for Analyzing the Public Sentiment Variations on Twitter Using Multi-Core Programming

**Gayatri R Bhongade**
PG Student, M. techCSE
AGPCET, Nagpur, India

**Prof. Pragati Patil**
Professor, Department of CSE
AGPCET, Nagpur, India

***Abstract— Now a days, social contacts are vital to find relevant content. We need to connect with people with similar interests because they provide content that matters. Every day is more clear that in the future of document recommendations will be necessary to cross the traditional data with the data obtained from twitter. For instance, in order to provide the best content available we can use sentiment analysis techniques to prioritize content with good reviews. The aim of this project is to offer a better sentiment recognition strategy. analyzing short messages about brands in Twitter trying to classify them between positive and negative using Sentiwordnet. Based on this observation, we propose a Latent Dirichlet Allocation (LDA) based model, Foreground and Background LDA (FB-LDA), to distill foreground topics and filter out longstanding background topics. We examine sentiment analysis on Twitter data. The contributions of this paper is : (1) we use Parts Of Speech (POS)-specific prior polarity features and used the multi-core programming to perform parallel processing in multi-core way to get the output in minimum time.***

***Keywords— Twitter, public sentiment, emerging topic mining, sentiment analysis, latent Dirichlet allocation.***

## I. INTRODUCTION

WITH the explosive growth of user generated messages, Twitter has become a social web site wherever millions of users will exchange their opinion. Sentiment Analysis on Twitter knowledge has provided cost-effective and effective thanks to expose belief timely that is essential for deciding in numerous domains. as an example, accompany will study the general public sentiment in tweets to get users' feedback towards its products; whereas a politician can adjust his/her position with relation to the sentiment amendment of the general public. There are an outsized range of analysis studies and industrial applications within the space of public sentiment trail in and modeling. Previous analysis like O'Connor et al. [14] centered on trailing public sentiment on Twitter and studying its correlation with shopper confidence and presidential job approval polls. Similar studies are done for work the reflection of public sentiment on stock markets [8] and oil value indices [7]. They rumored that event in reality so have a major and immediate effect on the general public sentiment on Twitter. However, none of these studies performed additional analysis to mine helpful insights behind vital sentiment variation, called public sentiment variation. One valuable analysis is to search out possible reasons behind sentiment variation, which might give important decision-making info. as an example, if negative sentiment towards Barak Obama will increase considerably, the White House Administration workplace is also eager to understand why individuals have change their opinion and then react consequently to reverse this trend. Another example is, if public sentiment changes greatly on some products, the related companies may want to know why their products receive such feedback.

## II. LITERATURE SURVEY

Tan S et al. [1] investigated the problem of analyzing public sentiment variations and finding the possible reasons causing this variation. To solve the problem, they proposed Latent Dirichlet Allocation (LDA) based models, Foreground and Background LDA (FB-LDA)

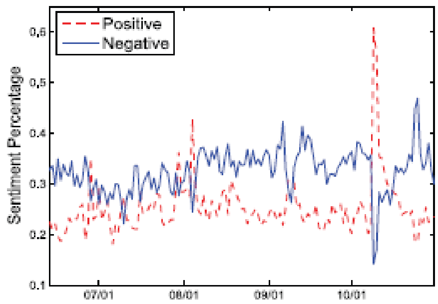


Fig 1. Sentiment Variation track of "obama"

Chen X et al. [2] developed a workflow to integrate both qualitative analysis and large-scale data mining techniques and focused on engineering students' Twitter posts to understand issues and problems in their educational experiences.

 

Gokulakrishnan B et al. [3] analyzed the problem of sentiment analysis and opinion classification of Twitter micro blog data, which, as discussed, is significantly different from other sentiment classification problem on structured and detailed messages. Becker H et al.[4] exploited the rich context" associated with social media con tent, including user-provided annotations (e.g., title, tags) and automatically generated information (e.g., content creation time) Using this rich context, which includes both textual and non-textual features, they define appropriate document similarity metrics to enable online clustering of media to events. Blei et al. [6] proposed Latent Dirichlet Allocation (LDA), a generative probabilistic model for collections of discrete data such as text corpora. They presented efficient approximate inference techniques based on variation methods and an Expectation–maximization algorithm for empirical Bayes parameter estimation. Bollen J et al. [7] proposed system based on a sentiment analysis of all public tweets broadcasted by Twitter users between August 1 and December 20, 2008. Compared results to punctuations recorded by stock market and crude oil price indices and major events in media and popular culture. Zenge et al. [8] proposed behavioural economics tells us that emotions can profoundly affect individual behaviour and decision-making. They analyzed the text content of daily Twitter feeds by two mood tracking tools, namely Opinion Finder that measures positive vs. negative mood and Google-Profile of Mood States (GPOMS) that measures mood in terms of 6 dimensions (Calm, Alert, Sure, Vital, Kind, and Happy). Chakrabarti n Punera [9] proposed some highly structured and recurring events, such as sports, it is better to use more sophisticated techniques to summarize the relevant tweets. Formalize the problem of summarizing event-tweets and give a solution based on learning the underlying hidden state representation of the event via Hidden Markov Models. Haung et al. [10] described the pre-processing steps needed in order to achieve high accuracy. The main contribution of this paper is the idea of using tweets with emoticons for distant supervised learning. Griffiths et al. [11] showed that the extracted topics capture meaningful structure in the data, consistent with the class designations provided by the authors of the articles, and outline further applications of this analysis, including identifying ‘‘hot topics’’ by examining temporal dynamics and tagging abstracts to illustrate semantic content. Hall et al. [12] proposed unsupervised topic modelling to the ACL Anthology to analyze historical trends in the field of Computational Linguistics from 1978 to 2006. They induced topic clusters using Latent Dirichlet Allocation, and examined the strength of each topic over time. Their methods find trends in the field including the rise of probabilistic methods starting in 1988, a steady increase in applications, and a sharp decline of research in semantics and understanding between 1978 and 2001, possibly rising again after 2001. Heinrich et al. [13] presented parameter estimation methods common with discrete probability distributions, which is of particular interest in text modelling. Starting with maximum likelihood, a posterior and Bayesian estimation, central concepts like conjugate distributions and Bayesian networks are reviewed.

## III. PROPOSED WORK

The proposed Sentiment Analysis on twitter data is based on these important parts viz Data Extraction, pre-processing of extracted data and labelling and tracking.

To uncover the tweets we first extract the tweets then apply the preprocessing on the tweets which are extracted. That perform the three steps (1)slang word translation,(2) non-english word removal,(3)URL removal. After that assign the label and tracking the publc sentiments. Using the FB-LDA we have to distill out the foreground topics from the background topics.

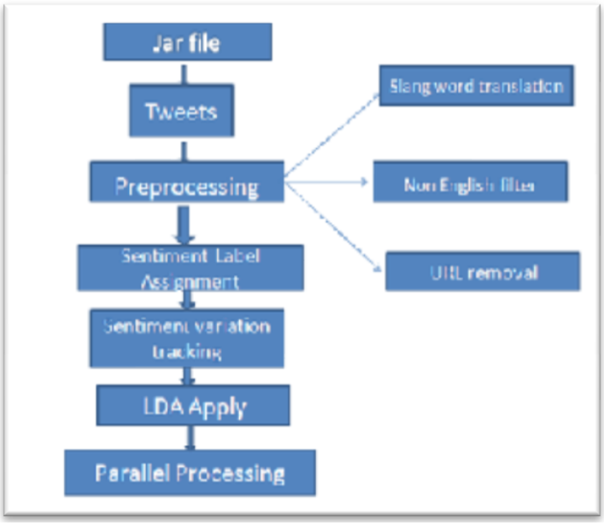


Fig.2 system Architecture

## IV. METHODOLOGY

Sentiments are the words or sentences that represent view or opinion that is held or expressed that can be positive, negative or neutral.To analyze public sentiment variations and find possible reasons behind these variations, for this propose Latent Dirichlet Allocation (LDA) based models: Foreground and Background LDA (FB-LDA).To extract tweets associated with the target, it will undergo the complete dataset and extract all the tweets that contain the keywords of the target. Compared with regular text documents, tweets square measure usually less formal and infrequently written in manner. Sentiment analysis tools applied on raw tweets typically come through terribly poor performance in most cases. Therefore, pre-processing techniques on tweets square measure necessary for getting satisfactory results on sentiment analysis.

**Preprocessing:**

(1) **Slang words translation**: Tweets typically contain plenty of slang words (e.g., lol, omg). These words square measure typically necessary for sentiment analysis however might not be enclosed in sentiment lexicons. Since the sentiment analysis tool i'm about to use is predicated on sentiment lexicon, it convert these slang words into their commonplace forms exploitation the web Slang Word Dictionary1 so add them to the tweets.
(2) **Non-English tweets filtering**: Since the sentiment analysis tools to be used solely work for English texts, it'll remove all non-English tweets before. A tweet is taken into account as non-English if over twenty % of its words (after slang words translation) don't seem within the antelope Aspell English dictionary.
(3) **Computer address removal**: plenty of users used URLs in their tweets. These URLs complicate the sentiment analysis method for this will plan to take away them from tweets.

**Sentiment label assignment**:

To assign sentiment labels for every tweet a lot of with confidence, we resort to 2 progressive sentiment analysis tools. One is that the Senti Strength tool. This tool relies on the LIWC sentiment lexicon. It works within the following way: initial assign a sentiment score to every word in the text in line with the sentiment lexicon; then select the most positive score and the utmost negative score among those of all individual words within the text; compute the total of the utmost positive score and therefore the maximum negative score, denoted as Final Score; finally, use the sign of Final Score to point whether or not a tweet is positive, neutral or negative.

**Sentiment variation tracking:**

once getting the sentiment labels of all extracted tweets about a target, we are able to track the sentiment variation exploitation
some descriptive statistics. Previous work on burst detection usually chooses the variation of the full variety of tweets over time as associate indicator. However, in this work, we tend to have an interest in analyzing the fundamental measure during that the positive (negative) sentiment climbs upward whereas the negative (positive) sentiment slides downward. During this case, the full variety of tweets isn't informative any longer since the quantity of positive tweets and negative tweets could amendment consistently.

## V. RESULT AND ANALYSIS

After entering the keyword and extracting we get the tweets from twitter and after applying the preprocessing we gate some labels from that we get positive, negative and neutral tweets. With the help of that we find the variations between that tweets.

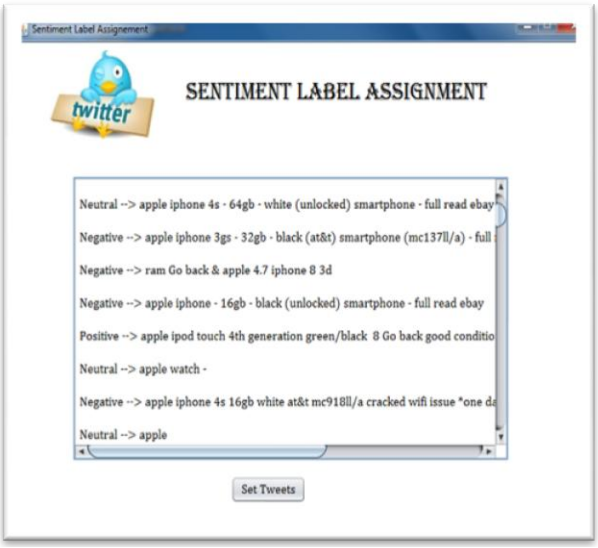


Fig.4 Snapshot for Positive, negative and neutral tweets

## VI. CONCLUSION AND FUTURE SCOPE

This model proposes a Latent Dirichlet Allocation (LDA) based mostly model, Foreground and Background LDA (FB-LDA), to distil foreground topics and filter out long background topics. These foreground topics will offer potential interpretations of the sentiment variations. The projected models also can be applied to different tasks like finding topic variations in the sets of documents. At this stage that much work is done. In this project will try to perform the general public sentiments using multi-core programming and trying to improve the performance and efficiency of the system. Using parallel processing it take minimum time to show the output as compare to serial processing. This work is incomplete without parallel processing we will show the output after research work is done.

## REFERENCES

[1] S. Tan, Y. Li, H. Sun, Z. Guan, X. Yan, J. B. Chun Chen, and X. He, "Interpreting the Public Sentiment Variations on Twitter", IEEE Transactions On Knowledge And Data Engineering, Vol. 26, No. 5, May 2014.

[2] X. Chen, M. Vorvoreanu, and K. Madhavan, "Mining Social Media Data for Understanding Students' Learning Experiences", IEEE Transactions on Learning Technologies,

[3] B. Gokulakrishnan, P. Priyanthan, T. Ragavan, N. Prasath, A. Perera, "Opinion Mining and Sentiment Analysis on a Twitter Data Stream" , The International Conference on Advances in ICT for Emerging Regions - ICTer 2012 pp. 182-188.

[4] M. Anjaria, R. M. Reddy Guddeti, "Influence Factor based Opinion Mining of Twitter Data Using Supervised Learning", IEEE 2014.

[5] H. Becker, M. Naaman, and L. Gravano, "Learning similarity metrics for event identification in social media", in Proc. 3rd ACM WSDM, Macau, China, 2010.

[6] D.M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation", J. Mach. Learn. Res., vol. 3, pp. 993–1022, Jan. 2003.

[7] J. Bollen, H. Mao, and A. Pipe, "Modelling public mood and emotion: Twitter sentiment and socio-economic phenomena", in Proc. 5th Int. AAAI Conf. Weblogs Social Media, Barcelona, Spain, 2011.

[8] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market", J. Comput. Sci., vol. 2, no. 1, pp. 1–8, Mar. 2011.

[9] D. Chakrabarti and K. Punera, "Event summarization using tweets", Proc. 5th Int. AAAI Conf. Weblogs Social Media, Barcelona, Spain, 2011.

[10] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision", CS224N Project Rep., Stanford: 1–12, 2009.

[11] T. L. Griffiths and M. Steyvers, "Finding scientific topics", Proc. Nat. Acad. Sci. USA, vol. 101, (Suppl. 1), pp. 5228–5235, Apr. 2004.

[12] D. Hall, D. Jurafsky, and C. D. Manning, "Studying the history of ideas using topic models", Proc. Conf. EMNLP, Stroudsburg, PA, USA, 2008, pp. 363–371.

[13] G. Heinrich, "Parameter estimation for text analysis", Fraunhofer IGD, Darmstadt, Germany, Univ. Leipzig, Leipzig, Germany, Tech. Rep., 2009.

[14] O'Connor, R. Balasubramanyan, B. R. Routledge, 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.

 