An Analysis of Students Opinions on Social Networking Site for Understanding Student's Learning Practices

Prof. Mangesh R. Balpande, Prof. Deven Ketkar, Prof. Mahesh Patil

[1]Assistant Professor, IT Department, FAMT, Ratnagiri, [2] Assistant Professor, IT Department, FAMT, Ratnagiri, [3] Assistant Professor, IT Department, FAMT, Ratnagiri

[1]Information Technology Department.

[1]Finolex Academy of Management of Technology, Ratnagiri, Maharashtra, India

***Abstract*— Recent year's rapid growth in internet use, social networking sites become important medium of communication. By using social networking sites such as Facebook, twitter, LinkedIn etc., millions of messages are exchanged. Different users share their personal opinions or views about various issues. They also discuss several current hot trending topics on Twitter and Facebook. Many people including youngsters, businessmen, sports players and film industry people use different social networking sites for sharing their views on different trends. By using this information that we generate from different posts, making it an important base for tracking and analyzing sentimentation of students. This information is useful in decision making and opinion mining. In this work, we have moved one step further to interpret sentiment variations using twitter tweets. We observed that emerging topics (named foreground topics) within the sentiment variation periods are highly related to the genuine reasons behind the variations. We select the most representative tweets for foreground topics and develop another generative model called Reason Candidate and Background LDA (RCB-LDA) for ranking them with respect to their "popularity" within the variation period.**

Opinion mining also known as sentiment analysis plays an important role in determining the sentiments involved in various content. For example, if anyone wants to buy a new car, a buyer will always check reviews. Based on that reviews he/she will take a decision. These kind of decisions are based on others experiences. Student sentiment analysis is currently a very significant trend in the area education system. Many universities can make use of these data for taking opinion on their services that they provide to student. Student sentiment analysis is important task in the area of natural language processing. Natural language processing involves giving artificial intelligence to computers and is concerned with promoting an understanding of human languages for machines' use. Student sentiment analysis extracts opinions, sentiments and emotions from text and analyze them this information is very useful for governments, businesses and individuals. Hence, need arises to develop an intelligent system which mine such huge content and classify them into Negative, Positive, Neutral type. Student sentiment analysis is the automated mining of opinions, attitudes, emotions from text and other posts, and database sources through Natural Language Processing (NLP).

Keywords — **Student sentiment Analysis, Public sentimental Social Sites, Twitter, Emerging topic mining.**

I. INTRODUCTION

Extensive growth of user generated messages on internet, Social site like Twitter where millions of users used to share their opinion regarding some topic. Figure 1 indicates that web has huge amount of data and social networks has part of that huge data. We can perform sentiment analysis on Social Sites where data has provided a platform where timely student sentiment can be exposing in an economical and effective way, which is tedious for decision making in various domains. For example, a company could analyze the public sentiment in tweets 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 succession of movies sentiment.

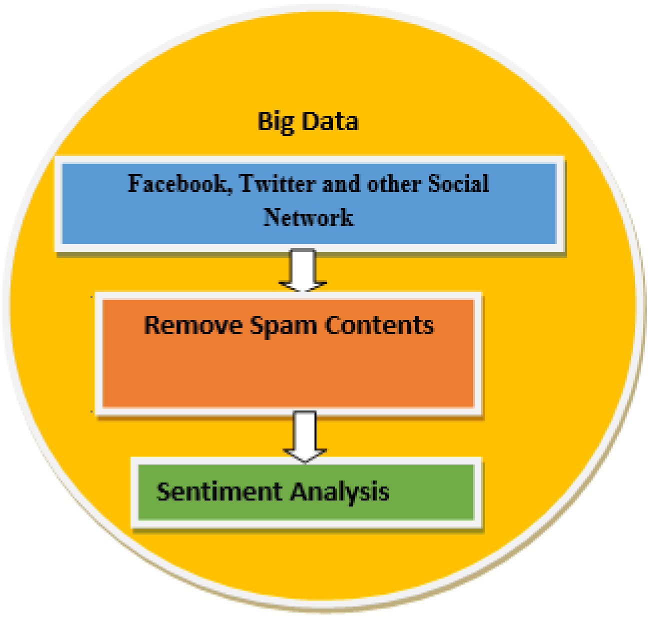

Fig. 1 Visualization of Social Network Analysis

Due to the tremendous use of social media services, there is a great opportunity to understand and analyze the sentiment of the students by analyzing its large-scale data as well as opinion-rich data. Sentiment analysis on tweets can be done by many approaches. Various methods such as machine-learning and lexicon-based approaches have been widely used for sentiment analysis on twitter like sites. Machine-learning approaches to student sentiment analysis needs to train the data.

Searching for student's opinions via surveys and polls has been an expensive and time-consuming task. The proliferation of web 2.0 has changed the way students express their opinions and feelings. This so called user-generated content posted in blogs, forums, product review sites and social network is most publicly available and easy to obtain. Thus, there is a growing need for automated analysis of this kind of data. This is a challenging task with foundations in Natural Language Processing (NLP) and text mining referred to as sentiment analysis. Many research studies in sentiment analysis are concerned with product reviews from websites like twitter is a most popular worldwide social website, which provides a micro blogging services and social networking, enable its users to update their status in tweets, follow people they are interested in (e.g. Sachin Tendulkar) and retweet other‘s posts and even communicate with them directly. The student sentimental analysis on Twitter data has provided an economical and effective way to expose timely student sentiments, which is critical for decision making in various domains areas. For instance, a universities can study the sentiments of student in Tweets to obtain users ‘feedback towards its education system. There are several streams of research investigating the role of Twitter. Twitter has attracted attention in both academia and industry for research area. Previous research mainly focused on tracking public sentiment. There have been a large number of research studies and industrial applications in the area of public sentiment tracking and modelling. Previous research like O‘Connor focused on tracking public sentiment on twitter and studying its correlation with consumer confidence and presidential job approval polls. On twitter, any user can publish a message referred to as 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 stock markets. They reported that events in real life indeed have a significant and immediate effect on the public sentiment on twitter. One valuable analysis is to find possible reasons behind sentiment variation, which can provide important information for decision-making. E.g. 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.

II.LITERATURE SURVEY

Many Researchers carried out research work in Social Network Analysis and sentiment analysis. Sentiment Analysis (SA) is a text processing technique to derive an opinion or mood intention based on the terms used in a real language sentence. Number of researchers have concentrated on generating statistical inference from social network data using sentiment analysis models. Bo Pang and Lilliam Lee [2] provided an insight full discussion on sentiment analysis. In this, they have considered the ratio in positive words and total words to estimate the opinion.

Social media technologies take on many different forms including social network, micro blogging, weblogs, magazines, internet forums, social blogs, photographs, video, rating and social bookmarking. Public and private opinions about variety of subjects are expressed and spread continually via numerous social media. Sentiment analysis is used to determine the attitude of a writer with respect to some topic. The attitude may be his or her judgment, the intended emotional communication or the emotional state of the author when writing. A basic task in sentiment analysis is classifying the polarity of a given text at the word, sentence, and document whether expressed opinion in a word, sentence or in document has sentence feature positive, negative, or neutral. Classification of sentiments looks, for instance, at emotional states such as 'happy', 'angry', 'sad' and 'neutral'. Sentiment analysis has become popular in judging opinion of consumers towards various brands. The way in which consumers express their opinion on social networking websites helps to judge this opinion .When it comes to sentiment or opinion or emotion, we are not concerned with the topic of the text but the positive or negative opinion it express. People can freely express their opinion in social media as blogs, micro blogs, reviews, forum discussion and social network sites towards any person, events, product, service, news or organization. All these platforms are source of huge amount of valuable information that we are interested to analyze.

Several prior studies have estimated and made use of aggregated text sentiment. The informal study by Lindsay (2008) focuses on lexical induction in building a sentiment classifier for a proprietary dataset of Facebook wall posts (a web conversation/micro blog medium broadly similar to Twitter), and demonstrates correlations to several polls conducted during part of the 2008 presidential election. We are unaware of other research validating text analysis against traditional opinion polls, though a number of companies offer text sentiment analysis basically for this purpose (e.g., Nielsen Buzzmetrics). There are at least several other studies that use time series of either aggregate text sentiment or news, including analyzing stock behavior based on text from blogs, good and bad news (Gilbert and Karahalios 2010), news articles (Koppel and Shtrimberg 2004;Lavrenko et al. 2000) and investor message boards (Antweiler and Frank 2004; Das and Chen 2007). Dodds and Danforth (2009) use an emotion word counting technique for purely exploratory analysis of several corporations.

Twitter is making it a valuable platform for tracing and analyzing public sentiment. It provides information for decision making in various domains with respect to current issues in society. In this work, we interpret sentiment variations over various topics from society. Recent topics within the sentiment variation periods are related to genuine reasons behind the variations. Based on this observation, Latent Dirichlet Allocation (LDA) based model, Foreground and Background LDA (FB-LDA), to distill foreground topics. It filters out longstanding *background topics*. Foreground topics can give interpretations about the levels of sentiment variation. This proposed system selects most representative tweets data for foreground topics and develop model called Reason Candidate and another generative model called Background LDA (RCB-LDA) to rank them with respect to their popularity'within the variation period. Latent Dirichlet Allocation (LDA) based models to analyze tweets in significant variation periods, and identify possible root cause for the variations. This model can be termed as Foreground and Background LDA (FB-LDA) Mode l, and it can filter both background topics and extract foreground topics from tweets in the specified variation period, by the use of an auxiliary set of background tweets generated just before the variation. Reason Candidate and Background LDA (RCB-LDA). RCB-LDA first extracts representative tweets for the foreground topics (obtained from FB-LDA) as reason candidates. After that it will associate each remaining tweet in the variation period with one reason candidate and rank the reason candidates by the number of tweets associated with them. (21-2 night)(80%).

There are many papers, which describe different classification techniques for sentiment analysis. Sentiment classification can be formulated as a supervised problem with two class labels (positive and negative). In (Pang, Lee and Vaithyanathan 2002), the authors apply supervised learning methods such as naïve Bayesian and support vector machines (SVM) to classify movie reviews into two classes. Most unsupervised sentiment classification approaches try to generate a general or domain dependent opinion lexicon for words or opinion phrases. In (Riloff and Wiebe 2003), the authors collected subjectivity clues as a part of their work. The clues were then used in (Wiebe, Wilson and Cardie 2005) to detect semantic orientation. In this paper, a bootstrapping process was proposed, where high precision classifiers use known subjective vocabulary to separate subjective and objective sentences from a no annotated text collection. The aspect extraction method refers to the concept of determining opinion targets and their attributes which are mentioned in a document or a sentence. Many information extraction techniques have been applied so far.(91%).

III. PROPOSED METHODOLOGY

A. General Architecture

Today's almost all Social Networking sites have been widely used for expressing opinions or emotion in the public domain with help of internet. Twitter has been the point of attraction to several researchers in important areas. The main two-fold contributions of this paper are: to the best of our study, our research and our knowledge is the first work that tries to analyze and interpret the public sentiment variations in micro blogging services like twitter. Two novel generative models are developed to solve the reason mining problem. The two proposed models are general: they can be applied to other tasks such as finding topic differences between two sets of documents.

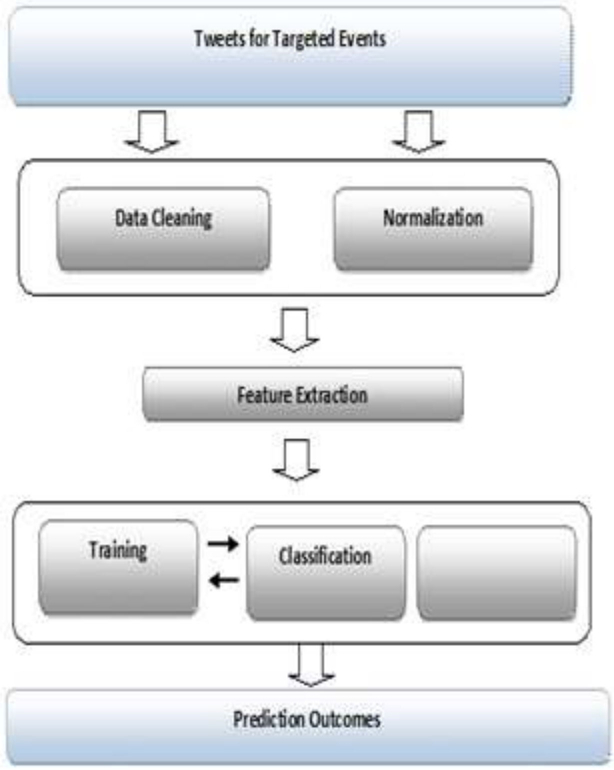

Fig. 2 High Level System Flow

Fig. 2 shows an example of High level system flow. To analyze variations in public sentiments There are two Latent Dirichlet Allocation (LDA) based models: (1) Foregroundand Background LDA (FB-LDA) and (2) Reason Candidate and Background LDA (RCB-LDA). NaïveBayes, SVM, MaxEnt, ANN classifiers with features extracted from Twitter data using feature extraction methods such as Unigram, Bigram and Hybrid (Unigram + Bigrams) for sentiment analysis. In order to remove stop words and to extract features from text, we perform data cleaning and normalization on given set. We extract target based extended features model by modifying it and twitter user data from the normalized data [7]. Vectors are used in part of chunks to train the classifier as a part of incremental training. The sentiment analysis results are incorporated with influence factor of supervised learning to predict the results using prediction model.

B. Proposed Architecture

In our work, we have proposed following three steps for sentiment tracking-:

1) We extract tweets related to our interested targets (e.g. Arvind Kejrival, Delhi Election 2015 *etc*), and preprocess the raw extracted for more cleaned for sentiment analysis.

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

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

IV. MODULES

A. Tweets Extraction and Preprocessing: Our First phase starts with extracting tweet lines related to the targeted issue, we go through the whole collected raw dataset and extract all the core lines tweets which contain keywords of the targeted issues. Compared with regular text documents, tweets are generally somewhat informal and often written in an adhoc manner like it may contains short forms, some abbreviation. Sentiment analysis tools applied on raw tweets but often achieve very poor performance in most cases. Hence, there is a need of preprocessing techniques on tweets that 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 sentiment analysis tools to be used only work for English texts, we remove all non-English tweets in advance as these non English words 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) do not appear in the GNU Aspell English Dictionary.

3) URL removal: A lots of users may include various URLs in their tweets. These URLs may complicate our sentiment analysis process. Hence, we have decided 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 SentiStrength tool [8]. This tool is based on the LIWC sentiment lexicon[10]. It works in the following way: first assign a sentiment score to each word in the text according to the sentiment lexicon; then choose maximum positive score and maximum negative score among those of all individual words in the text; compute the sum of 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 unsupervised technique, classification is done by comparing 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, analyze 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 to 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 training set.

(a) If token is present in dictionary,

i. If token is positive, then s _ s + w.

ii. If token is negative, then s _ s − w.

4. Look at aggregate text sentiment score s,

(a) If s > threshold, then classify the text as positive

(b) If s < threshold, then classify the text as negative.

V. CONCLUSIONS

Overall, we conclude that social network based behavioral analysis parameters can increase the prediction accuracy. However, 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 student sentiment variations and finding possible reasons causing these educational problem and overcome it. Our study can inform educational administrators and other relevant decision maker to gain further understanding of engineering student problems. Our purpose is to achieve deep understanding of student learning experiences.

REFERENCES

1] G. Siemens and P. Long, "Penetrating the Fog: Analytics in Learning and Education, Educause Rev., vol. 46, no. 5, pp. 30-32, 2011.Bo Pang. Lilliam Lee, *"Seeing Stars: Exploiting class relationships for sentiment categorization with respect to rating scales"*, 2002.

2] M. Rost, L. Barkhuus, H. Cramer, and B. Brown, "Representation and Communication: Challenges in Interpreting Large Social Media Datasets," Proc. Conf. Computer Supported Cooperative Work, pp. 357-362, 2013.

3] M. Clark, S. Sheppard, C. Atman, L. Fleming, R. Miller, R. Stevens, R. Streveler, and K. Smith, "Academic Pathways Study: Processes and Realities," Proc. Am. Soc. Eng. Education Ann. Conf. Exposition, 2008.

4] C.J. Atman, S.D. Sheppard, J. Turns, R.S. Adams, L. Fleming, R. Stevens, R.A. Streveler, K. Smith, R. Miller, L. Leifer, K. Yasuhara, and D. Lund, Enabling Engineering Student Success: The Final Report for the Center for the Advancement of Engineering Education. Morgan & Claypool Publishers, Center for the Advancement of Engineering Education, 2010.

5] Bernard J. Jansen, Mimi Zhang, Kate Sobel and AbdurChowdury,‖*Micro-blogging as online word of mouth branding"*, 27th International Conference Extended Abstracts on Human Factors in Computing Systems, New York, 2009,pages 3859-3862.

6] J.C. Na, H. Sui, C. Khoo, S. Chan, and Y. Zhou.‖*Effectiveness of simple linguistic processing in automatic sentiment classification of product reviews"*,Advances In Knowledge and organization, 2004, pages 49-54.

7] T. Minka and J. Lafferty, ―*Expectation-propagation for the generative aspect model*‖, in *Proc. 18th Conf. UAI*, San Francisco, CA, USA, 2002.

8] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas, ―*Sentiment strength detection in short informal text*‖ ,*J. Amer. Soc.Inform. Sci. Technol.*, vol. 61, no. 12, pp. 2544–2558, 2010.

9] Nat'l Academy of Eng., The Engineer of 2020: Visions of Engineering in the New Century. National Academies Press, 2004.

10] E. Goffman, The Presentation of Self in Everyday Life. Lightning Source Inc., 1959.