International Journal of Latest Trends in Engineering and Technology
Vol.(7)Issue(3), pp. 392-395
DOI: http://dx.doi.org/10.21172/1.73.551
e-ISSN:2278-621X

Analyze engineering student's Twitter posts to understand issues and problems in their educational experiences

Mr.S.D.Rane[1] , Prof.U.A.Nuli[2] and Prof.N.S.Mahajan[3]

Abstract-Students' informal conversations on Twitter useful for know the student educational experiences concerns about the learning process. Data from Twitter and other social networking environments can provide valuable knowledge to inform student learning. However, the growing large scale of data required automatic data analysis and mining techniques. Proposed new system which will be help to developing a workflow to integrate both qualitative analysis and large-scale data mining techniques. This will be focusing on engineering students' Twitter posts to understand issues and problems in their educational life. First conduct a qualitative analysis on samples taken from tweets related to engineering students' college life. In proposed system will used a multi-label classification algorithm to classify tweets reflecting students' problems such as heavy study load, lack of motivation, stress, lack of social engagement, sleep deprivation and others.

Keywords – Data mining, Text classification, Tweets, Social networking.

I. INTRODUCTION

Social media sites Twitter have become important venues for the young generation to communicate and exchange information about their learning process. On various social media sites, such as twitter, whatsapp students discuss and share their everyday problems in an informal manner. Students' digital footprints provide implicit knowledge and a whole new perspective for educational researchers to understand students' experiences outside the controlled classroom environment. This understanding can inform institutional decision-making on interventions for at-risk students, improvement of education quality in college and thus enhance student recruitment and retention ratio.

Traditionally, educational researchers uses methods such as surveys, interviews, classroom activities to collect data related to students' learning experiences. These methods are usually required more time. The scale of such studies are usually limited.

Here we chose to give more focus on engineering students' posts on Twitter about problems in their educational experiences because Engineering colleges and departments have been struggling with student recruitment and retention issues, Based on understanding of issues and problems in students'

[1] *IT Department Dkte's TEI Ichalkaranji (Maharashtra), India*
[2] *CSE Department Dkte's TEI Ichalkaranji (Maharashtra), India*
[3] *IT Department Dkte's TEI Ichalkaranji (Maharashtra), India*

life, policymakers can make decisions on services that can help students overcome such problems and issues.

Here we propose system for proposes a workflow for a qualitative research methodology and large-scale data mining techniques . We use qualitative data from human interpretation for data mining algorithm , so that we can gain deeper understanding of data and get quality result.

The rest of the paper is organized as follows. Proposed embedding and extraction algorithms are explained in section II. Experimental results are presented in section III. Concluding remarks are given in section IV.

II. LITERATURE SURVEY

Goffman's theory for social performance used to drawn the value of information data on the web. Most fundamental aspects of this theory is the idea of front-stage and back-stage of people's social performances. Compared with the front stage, the relaxing atmosphere of back-stage usually applauds more spontaneous actions. For students, compared to classroom settings, social media are relative relaxing back-stage. When students post content on social media sites, they usually post what they think and feel at that moment. In this sense, the data collected from online conversation may be more authentic and unfiltered than responses to formal research prompts. Many studies show that social media users may purposefully manage their online identity to "look better "than in real life [1] [2].Discus about importance of analyzing social media as a communicative rather than representational system [3]. In automatic twitter sentiment classification the Twitter messages classified as either positive or negative with respect to a twitter query term. This is useful for customer to re-search the sentiment of products before purchase, or companies to monitor the public sentiment of their brands [4]. Predicting the popularity of messages as measured by the number of future retweets and make clear what kinds of factors influence information propagation in Twitter. Treat the problem as a classification task. First, required to train a binary label classifier with positive and negative examples of messages this will be retweeted in the future. Second, train a multi-label classifier to predicts the volume range of future retweets for a new message. To build classifiers, required to investigate classification of features to determine which ones can be used as predictors of popularity, including the content and relevant information of messages and graph structural properties of users [5]. Developed a workflow to integrate both qualitative analysis and large-scale data using mining techniques. More Focused on engineering students' Twitter posts to understand issues and problems in their educational life [6]. Introduce an approach to classify tweets into important categories by using the author information and features within the collect tweets messages. With such a system, users can subscribe to or view only certain types of tweets based on their interest [7].

III. PROPOSED SYSTEM

A. Proposed system

In the proposed system we will be focusing on engineering students' Twitter posts to understand issues and problems in their educational experiences. First we will conduct a qualitative analysis on samples taken from tweets related to engineering students' college life. Use a multi-label classification algorithm to classify tweets reflecting students' problems such as heavy study load, lack of social engagement, and sleep deprivation.

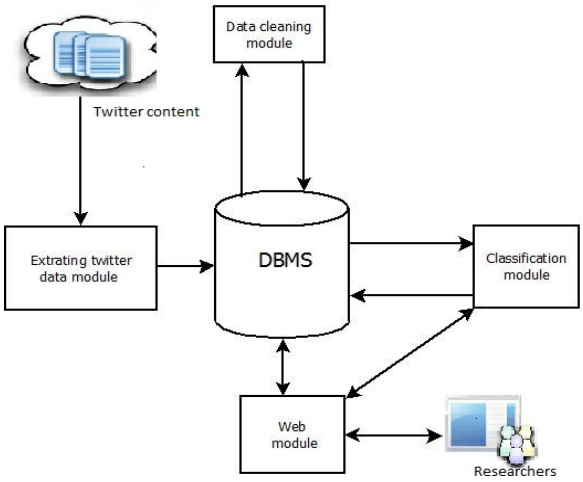

Figure1: System Architecture

The architecture of proposed system is shown in figure1. It will be describe in following modular form.

Modules in the proposed system are as follows

Extracting twitter data
Data cleaning
Tweet classification
Web module

Extracting twitter data

In this module use the twitter search API to collect the twitter data. For collection of tweet provide input terms as engineeringProblems to this module.

Data cleaning

In this module pre-processed the texts and find useful text before training the classifier. Preprocessing includes removed all the #engineeringProblems hashtags. For other co-occurring hashtags, only removed the # sign, and kept the hashtag texts, removed all words that contain non-letter symbols and punctuation, remove the common stopwords. After removing unnecessary symbols from collected tweet all the data are inserted in to database. Each tweet is auto-assign a tweet ID. The tweet content, Twitter user ID, time stamp are stored in separate columns with a primary key as the tweet ID in the tweet table.

Classification module

In this module we will do the classification modeling. The classification model can be build base on the researcher classification results and then applied to the other dataset. Researcher can choose which data set he would like to overview. Researcher can explore the data and defines categories. He can also specify the sample size and randomly select the specific number of tweets from database for analysis. Researcher will going to use Naïve Bayes classification because it is most efficient for our example data. Base on the researcher annotation of each tweet, the Naïve Bayes classifier build a probabilistic model for each word in each category. After model Build researcher will choose dataset which means to apply the model trained using sample of dataset to the entire dataset.

Procedures of the multi-label Naïve Bayes classifier. Suppose there are a total number of N words in the training document collection (in our case, each tweet is a document) $W = \{w_1, w_2,..., w_N$

}and a total number of P categories C = {c_1, c_2,..., c_p }.If a word w_n appears in a category c for s times, and appear in categories other than c for s' times, then based on the Maximum Likelihood Estimation, the probability of this word in a specific category c and probability of this word in categories other than c. For a document T in the testing set, there are K words W_T is a subset of W. To classify this document into category c or not c. assume independence among each word in this document, and any word wik conditioned on c or c' follows multinomial distribution.we use the Bayes Theorem, to find the probability that T belongs to category c or c'. If p(c | T) is larger than the probability threshold Th, then T belongs to category c, otherwise, T does belong to category c'. Select categories for tweets are heavy study load, lack of social engagement, negative emotion, sleep problems, diversity issues and other.

IV.CONCLUSION

It is provides a workflow for analyzing social media data for education purposes that overcomes the major limitation of manual qualitative analysis and large scale computational analysis. It is useful to understanding of engineering student's college experience

REFERENCES

[1] Hsin-Ying Wu, Kuan-Liang Liu and Charles Trappey, "Understanding Customers Using Facebook Pages: Data Mining Users Feedback Using Text Analysis", IEEE 18th International Conference on Computer Supported Cooperative Work in Design, 2014, pp. 346-350..

[2]. I-Hsien Ting, Shyue-Liang Wang, Hsing-Miao Chi and Jyun-Sing Wu, "Content Matters: A study of hate groups detection based on social networks analysis and web mining",IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2013,pp. 1196-1201

[3] M. Rost, L. Barkhuus, H. Cramer, and B. Brown, "Representation and communication: challenges in interpreting large social media datasets," in Proceedings of the 2013 conference on Computer supported cooperative work, 2013, pp. 357–362.

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

[5] L. Hong, O. Dan, and B. D. Davison, "Predicting popular messages in twitter," in Proceedings of the 20th international conference companion on World wide web, 2011, pp. 57–58.

[6] Xin Chen, Mihaela Vorvoeanu, and Krishna Madhavan, "Mining Social Media Data for Understanding students learning Experience, IEEE Transactions 2014.

[7] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, "Short text classification in twitter to improve information filtering," in Proceedings of the 33rd nternational ACM SIGIR conference on Research and development in information retrieval, New York, NY, USA, 2010, pp. 841–842.