INSPECTING STUDENTS' ACADEMIC REALITIES USING SOCIAL MEDIA PLATFORM

Priya Lande[1], Vipul Dalal[2]

[1] *Student, Department of Computer Engineering, Vidyalankar Institute od Technology, Maharashtra, India*
[2] *Associate Professor, Department of Computer Engineering, Vidyalankar Institute of Technology, Maharashtra, India*

Abstract

The informal conversation of the students used on social media has been studied vastly now-a-days to understand their experiences. Twitter has been examined in which the hashtag #EnggProblems is mined to extract tweets and to classify the problems of the students into categories: Heavy study load, sleep problems, lack of social engagement, diversity issues, negative emotion and there is one more category that is 'others' .many classification algorithms were applied on the tweets extracted from the said hashtag, out of which naïve bayes multi label classifier gives the best results. It has been observed that most tweets get classified into others category as a result of which students' problem cannot be identified clearly. This paper proposes an approach where SentiWordNet has been used to assign scores to each tweet and classify them as positive, negative or neutral. This helps to understand the sentiment of the student tweeting and identify at risk students

Key Words: *Student, classification, SentiWordNet, education.*

----------***----------

1. INTRODUCTION

Students consider online social media as their resort to entertainment. They voice out their views, concerns on the social media. Many social networking sites are popular like Facebook, Twitter. Students form a community or a group regarding any particular topic and everyone is free to express their views regarding that specified topic. In Twitter, this concept is represented as a hashtag. Twitter is said to be one of the popular social networking websites. The data students input to twitter in their posts is said to be a tweet. A student can tweet about himself personally, about his friends, and also he can tweet about his learning experience. These tweets when collected, they hold a large amount of information which gives scope to analyze this data, as a result of which, we can understand about what a student is going through. A hashtag begins with '#' it represents a certain context under which related tweets are tweeted. For example, a hashtag #friendsforever will have tweets which are related to friendship. Offline procedures to collect such data are available but they are time consuming and do not hold for repetition and faster frequency. [1] [2] hashtag related to students' life can give us an idea about their educational experience. A student can be an engineering student. Engineers are chosen for tweet analysis [3]. #EnggProblems consists of all the tweets related to a student's learning experience. We can analyze it by categorizing the tweets into specific problems which majorly apparent. A number of classifiers can be applied to the data. Naïve Bayes, Support Vector Machine and maximum margin multiplier, etc. Classification can be single label or multi-label. Single label classification classifies a data point in one class only whereas in multi-label classification can classify a data point into more than one class.

2. LITERATURE SURVEY

In [4], the classification of varied sentiments was processed on smaller chunks of text retrieved from twitter. The paper puts forth a supervised classification scheme which makes use of the hashtags and smileys taken from twitter to train the labels. The data on which this process is carried out identified the type of sentiment. The classification result of this is smileys where categorization obtained can be mixed.

In [5], the emoticons are extracted and labels are trained. Diverse multi-label classifiers, feature extractors are exploited to execute this approach. Naïve Bayes, SVM, maximum entropy classifier are used, with unigrams and bigrams to be feature extractors, where both the classifier and feature extractor serve as different units. This serves for trying different combinations of classifier

In [6], they have suggested the method of tweet classification, which is done by usage of information to classify the tweets. The users may also personalize the tweet views pertaining to their interest. Short texts do not offer statistical recurrence of words. In such state, Bag of words scheme doesn't give adequate results. Domain [articular features are extracted from the author's profile. This sufficiently classifies the text into a collection of classes like news, deals etc.

From the hashtag #EnggProblems tweets were streamed and all the duplicate tweets were removed and 19,799 tweets were gathered in total. Inductive content analysis was conducted over the large tweets collected; this led to formation of categories- heavy study load, sleep problems, negative emotion, diversity issues, lack of social engagement and 'others'. 'Others' is a special category which works for categorization of random tweets which do not hold up to any other above specified category. Naïve Bayes multi-label classifier has proved to be a better

classifier than other state-of-art classifiers [7] it is applied on the large dataset. This helped in classification of students' problems into the specified categories. This method not only classifies the students' problem but also helps in deeply understanding their experiences. On the contrary it is observed that most of the tweets get classified into the 'Others' category. This doesn't serve the purpose of understanding the students' problems extensively. In the fig. 1 beginning from the leftmost bar categories are: 1.Others, 2. Sleep problems, 3. Negative Emotion 4.Diversity Issues 5.Lack of Social engagement. 6. Heavy study load.

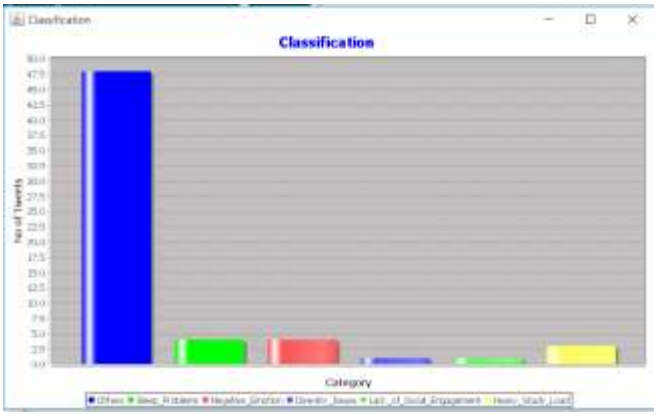

Fig -1: Result of Naïve Bayes Multi-Label Classifier.

3. PROPOSED METHOD

English nouns adjectives, verbs and adverbs form the synsets of the WordNet. Synset is the subjective collection of distinct synonyms. These synsets are linked together with some lexical connection. WordNet has large applications in the area of word sense disambiguation. It facilitates the provision of organized word groups with no loss of reliability and coverage. WordNet can corroborate the word sense and evaluate their pertinence with rest, depict their meaning. But to get by the sentiments too, WordNet has to be broadened with more detail.

3.1 SentiWordNet

It is the expansion of WordNet; all the values are related to negative, positive or neutral value [8]. The version of SentiWordNet advances from version 1.0 to 3.0. This interface is available online free of cost. It marks each synset with a value for each category between 0.0 and 1.0. Total of each category is 1.0, so that a non-zero value can be possessed by each synset for each and every sentiment. This is as a result of some synsets having a positive , negative or a neutral value again this depends the area in which they are been used. SentiWordNet advances the WordNet by a different angle. It can be applied in the area of sentiment analysis also. WordNet analysis can be given an extra scope by using SentiWordNet.

Opinion mining generally targets to inspect opinions of an author. People exploit the online media express their views and opinions. There are communities, blogs which discuss about varied topics which can be contemporary too. Twitter is one of the most popular microblogging platforms, where users are mostly active. Many industries utilise opinion mining and get an insight in the customers feedback given online, which help them to improve on their way of working and thereby gain profit.

SentiWordNet cannot handle multiple word sets, so it is advised that pre-processing is carried out. In our work we use SentiWordNet to score each tweet and that tweet is classified into positive, negative or neutral. In this way we can understand each incoming tweet from the 'others' category as positive or negative. This will put more light over the sentiments of the student as to positive or negative. This will also further mine the 'others' category which was an overhead for the classification. Application of opinion mining simplifies the issue and guides towards the sentiments of the students. The following picture shows the result after application of SentiWordNet dictionary on the 'others' category.

.

Fig-2 Tweets scored by Sentiwordnet

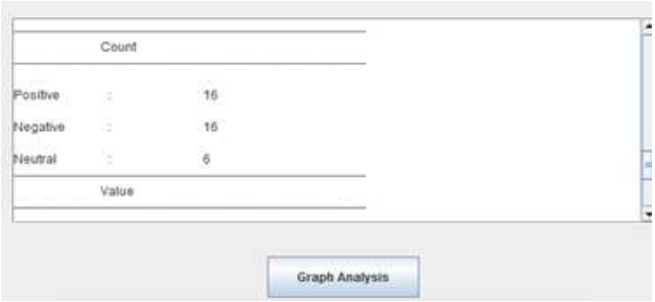

Fig-3 Summary generation as positive negative and neutral reviews

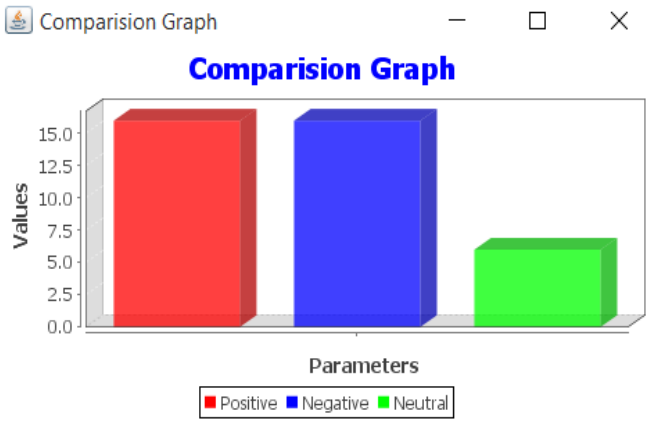

Fig-4 Graph analysis of above generated summary

3.2 Observations

Fig 2. shows the score given to each tweet after the application of Sentiwordnet. Fig. 3 is the generation of summary as positive and negative and neutral reviews from the tweets. Fig.4 shows the graph analysis of these observed reviews. It is observed that for Sentiwordnet, input is given as the whole data from the 'others' category; each word's score is referred and tagged with the help of the Sentiwordnet. Each word has got synonyms that depict a finite score which can be positive, negative or neutral. More the positive synonyms are more the word is positive and ultimately is the tweet. Similar is the case for negative tweets. Here it is seen that a clear and informative result is obtained where it can be distinguished whether a tweet's sentiment is retrieved. It gives insight regarding the scenario of the educational experience of the students. If there are more of negative tweets encountered then accordingly improvements can be carried out in the academic system.

4. CONCLUSION AND FUTURE DISCUSSION

Naïve Bayes multi-label classification gave results which did not provide enough information regarding students' problems. All the five categories did not cover the input as expected. The maximum of tweets were classified into the 'others' category which was a long tail. The data from others category was mined further using Sentiwordnet 3.0 dictionary which classified the tweets from 'others' category into positive, negative and neutral. The results obtained, provided a clear understanding of the sentiment of the students while posting the tweets. This atleast gave an insight into the scenario of the educational experiences of the students.

REFERENCES

[1] M.Clark, S.Sheppard, C.Atman, L. Fleming, R. Miller, R. Stevens, R. Streveler, K. Smith, "Academic pathway study: Processes and Realities", in *Proceedings of American Society for Engineering Education and Annual Conference and Exposition*, 2008.

[2] 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 Students' Success: the final report for the Center for the Advancement of Engineering Education," Morgan and Claypool Publishers, Center for the Advancement of Engineering Education, 2008.

[3] C Moller- Wong and A. Eide, "An engineering student retention study," Journal of Engineering Education, vol.86, no. 1, pp. 7-15, 1997.

[4] D. Davidov, O. Tsur and A. Rappoport, "Enhanced Sentiment learning using Twitter Hashtags and smileys," in *Proceedings of 23rd International Conference on Computational Linguistics: Posters*, 2010, pp. 241-249.

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

[6] 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 international ACM SIGIR conference on research and development in information retrieval*, New York, NY, USA, 2010, pp. 241-842.

[7] Xin chen, Krishna Madhvan, Mihaela Vorvoreanu: Mining Social Media Data for Understanding Students' Learning Experiences in *IEEE Transactions on Learning Technologies*, 2014.

[8] Opinion Mining using Sentiwordnet, Julia Kreutzer and Neele Witte Sentiment Analysis HT 2013/14 Uppsala University.

BIOGRAPHIES

Priya Lande
Studying Masters in Computer Engineering from Vidyalankar Institute of Technology, Mumbai.

Vipul Dalal
Working as an Associate Professor at Vidyalankar Institute of Technology, Mumbai.