Twitter Data Mining to Extract Student's Learning Problem

Zalak M. patel
Sanchalchand patel college of engineering
Student of master in computer engineering
Gujarat Technological University
Ahemdabad- Gujarat- India
Email: zalakpatel157@gmail.com

Vishal P Patel
Sanchalchand patel college of engineering
Mentoring master in computer engineering
Gujarat Technological University
Ahemdabad- Gujarat- India
Email: vppatel.ce@spcevng.ac.in

Abstract:

Now-a-days million of people have mostly focus on social media platforms to share their own thoughts and opinions to their day to day life, business, celebrity, education etc. people are shared their positive and negative opinions on social media platform. In this discuss about to extract the sentiment from a micro blogging service like Twitter. So new generation students are easily share their learning experiences on the social media platform. In this there is a problem to find out meaningful tweets that include positive and negative emotions of students. Earlier researches are done on only to find the problems regarding the student's learning experience. In this sentiment analysis used for finding the opinion of student's learning experience on a social media(Twitter) like positive, negative or neutral. Text mining is the area of data mining which deal with extracting useful information from the large database or document. To find student's learning experience used multilable classifier and ontology model to decide positive and negative sentiment.

Keywords – Opinion Mining, Twitter, Sentiment Analysis on opinion, Social Media, classification, ontology model

1. Introduction

Micro blogging websites are social media site(Twitter, Face book) to which user makes short and frequent posts. Twitter is one of the famous micro blogging services where user can read and post messages which are 140 characters in length. Twitter messages are also called as Tweets. Students are mostly share their emotions on the social media sites. Students are freely share their learning experience on social media platform. This Tweets ate difficult to understand because it contain informal grammar, sarscam, incomplete sentences etc. These tweets are also

useful for educators, any decision maker to know about student's learning experiences. Sometime students feel shy or afraid in the classroom to share their problems regarding study. At that time social media provide a platform to students to share their own opinion. And to know about contents of tweets are also useful for understand student's experiences. Student's experiences are two types good or bad. Traditionally educational research have been using surveys, focus group, interviews, classroom activity to collect the data of student's learning experiences[1]. These method are time consuming. To extract the experiences from social media platforms are effective.

For that purpose many mining techniques are used to extract meaningful information from the text documents.

The research goal of this study include section II related work. Section III define the proposed workflow. Section IV system analysis and design and section V give the conclusion of research work

2. Related work

Xin chen, mihaela, krishna madhavan[2] consider the complexity of student's experience reflected from social media content required human interpretation. Sometime human can not able for predict the tweets of students on a social media platform because of it's contain misspelling, slang words etc. So authors suggest the solution is Used qualitative analysis(Latent Dirichllet Allocation) and naive-bayes multi label classification algorithm to classifier tweets reflecting student problem.

Geetika Gautam, Divakar yadav [3] find out the problem that to identify customer's reviews using tweets are difficult to understand by any person. To solve this problem author select the feature vector list. Used various machine learning classifier and semantic approach to find polarity. Naïve bayse gives accurate result than all other classifier.

Luiz F.S Colrtta,nadia F. F. da[4] consider the problem of Stand alone Support Vector Machine(SVM)not give accurate result for finding solution using tweets for a student's experience. So authors suggest that Used combining classifier and cluster ensembles (C3E) to find out students problem so accuracy is improved to find out experience.

Sara Keretna, Ahmad Hossny, Doug[5] find out the problem that recognizing the identity of the users in social network is difficulty. So authors are suggest the method of authenticate the genuine account versus fake account using writeprint, which is the writing style biometric.

Neha R. Kasture, Poonam B. Bhilare[6]consider the problem is the expression of the verbal throught differs individually, To identifying the right sentiment from the bulk of data becomes the real challenge. Authors suggest that use

logical approach to analyze the sentiment of the text available on social media.

3. proposed work

For this work collect the data using Twitter hash tag like #Engineering problems from the Twitter API. We take some symptoms to classify tweets into positive and negative categories.

CATEGORY	MOST PROBABLE SYSMTOMS THAT OCCURS IN TWEETS
Heavy Study Load	hour, homework, exam, day, class, work, negtoken, problem, study, week, too much, all, lab, still, out, time, page, library, spend, today, long, school, due, engineer, already, disgusting
Lack of Social Engagement	negtoken, Friday, homework, out, study, work, weekend, life, class, engineer, exam ,drink, break, Saturday, people, social, lab, spend, tonight, watch, game, miss, party, sunny, beautiful
Negative Emotion	hate, f***, shit, exam, negtoken, week, class, hell, engineer, suck, study, hour, homework, time, equate, FML, lab, sad, bad, day, feel, tire, damn, death, hard
Sleep Problems	sleep, hour, night, negtoken, bed, all-night, exam, homework, nap, coffee, time, study, more, work, class, dream, lady engineer, late, week, day, long, morning, wake, awake, nosleep
Diversity Issues	girl, class, only, negtoken, guy, engineer, Asia, professor, speak, English, female, hot, kid, more, too much, walk, people, teach, understand, chick, China, foreign, out, white, black
Knowledge about instruments	3D printer, scanner, laptop, computer
Job opportunity	Awesome, job, opportunity, nice, congratulations, internship
Feeling happy	Like, feel, parent, study

Latest technology	Window 8, Linux, window 10

Table 1. List of category with their symptoms

There are many steps for getting student's learning experience:

(1) Data collection:

To collect the student's data from Twitter social media using Twitter API. Twitter API provide the data to the users.

(2) Pre-processing data:

This data contains many un useful and un understanding words. So pre- processing become more important phase to getting clean tweets about students.

Remove casing, Hash tags, Username, URL:

Casing means capital latter in the sentence. Capital word not give information about students in tweets. So remove the casing. In the tweets there is used the hash tags to express views about users, like #engineering problem, #business etc. This hash tags is not useful in our work so remove it. Username is also not used for finding the solution so remove it. URL is also not important for us to in our work so remove it.

Removes stop words and emoticons: Stop words means am, is, are, have, has etc, in the sentence. These all are not useful so remove it.

Now a day people are used different symbols like smiles to express their view on the social media. There is not useful to us so remove it.

Replace slang words: Students are used slang word to express their thought on social media. Like Happyyyyyyy, Goooood etc. So remove this extra letter and replace in proper word. Like Goooood is replace by Good.

Convert stream word in its root forms: Students are used streaming word like nationality, coming etc. So using the porters streaming algorithm convert the stream words to its root forms. Like nationality convert in national.

(3) Emotion ontology model:

This is used for the differentiate positive and negative words using semantic similarity approach. There is used semantic similarity approach to find out synonyms of the word to categories the students experiences. Collect the more word of positive and negative category and it is used to decide label of tweets like positive or negative.

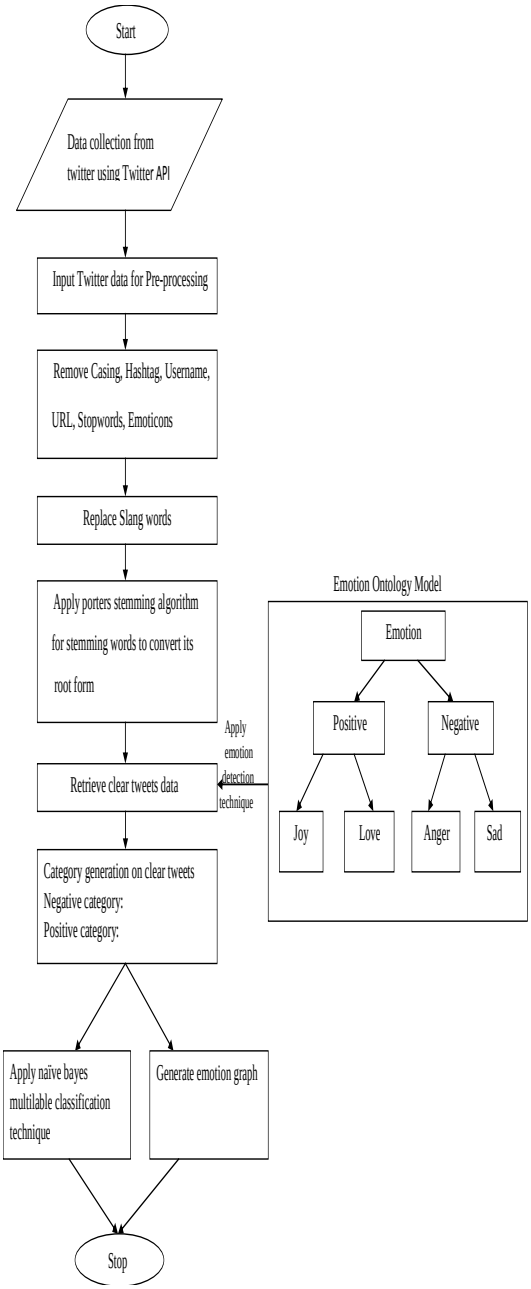

Fig. 1 Proposed workflow

(4) Retrieve clear tweets:

Using this stage retrieve the clear tweets using all above stapes.

(5) Category generation on clear tweet data:

There are two types of category are generated: Positive and negative views of students. Negative views are categories as Heavy study load, Lack of social engagement, Negative emotion, Sleep problem, Diversity issues. Positive views are categories as Knowledge about instruments, Job opportunity, Feeling happy, Latest technology.

(6) Generate emotion graph:

This stage is used for generating the emotion graph based on positive and negative tweets.

(7)Apply naïve bayes(NB) multi label classification:

There is used NB classifier to classify tweets category wise. Here used NB multi label classifier to classify tweets in a multiple category.

4. System analysis and design:

The proposed work is implemented in java environment and take dataset from Twitter API. Below fig. 2 display the GUI of our proposed work.

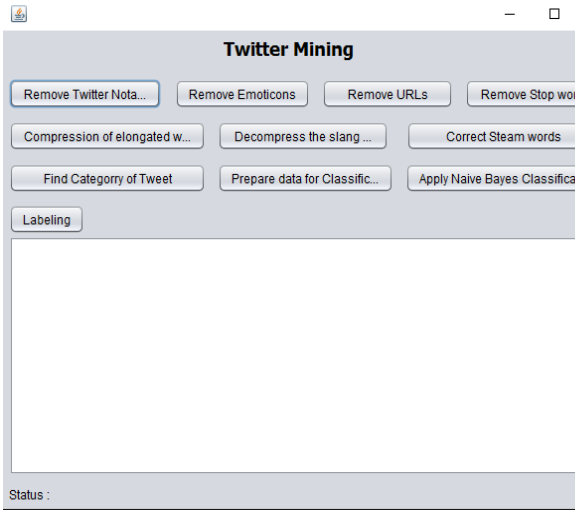

Fig. 2 GUI of proposed work

First we take a dataset from Twitter API using #Engineering problem. Them apply Pre-processing approaches to clean the tweets. Then categories tweets from positive and negative categories. Then apply the naïve bayes classifier and measure the performance.

4.1 Performance measurement:

Performance is measured by precision, recall, f-measure and accuracy.

These parameters are measured by following matrix.

	Tweet select by system	Tweet not select by system
Expected Tweet	True Positive(TP)	False Negative(FN)
Not Expected Tweet	False Positive(FP)	True Negative(TN)

Table. 2 Sentiment matrix

Precision, Recall, F-measure and Accuracy measured by following equations:

Precision: It gives the result of exactness value.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall: It gives the result of completeness value.

$$\text{Recall} = \frac{TP}{TP + FN}$$

F-measure: It is the harmonic value of recall and precision. The value of F-measure is between 0 and 1.

$$\text{F-measure} = 2 * \frac{\text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}}$$

Accuracy: It is measure the correctly classified value.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

To calculating the value of TP, TN, FP, FN below matrix is made.

	Tweet select by system	Tweet not select by system
Expected Tweet	0.842	0.158
Not Expected Tweet	0.045	0.234

Below graph present the performance measurement of the work.

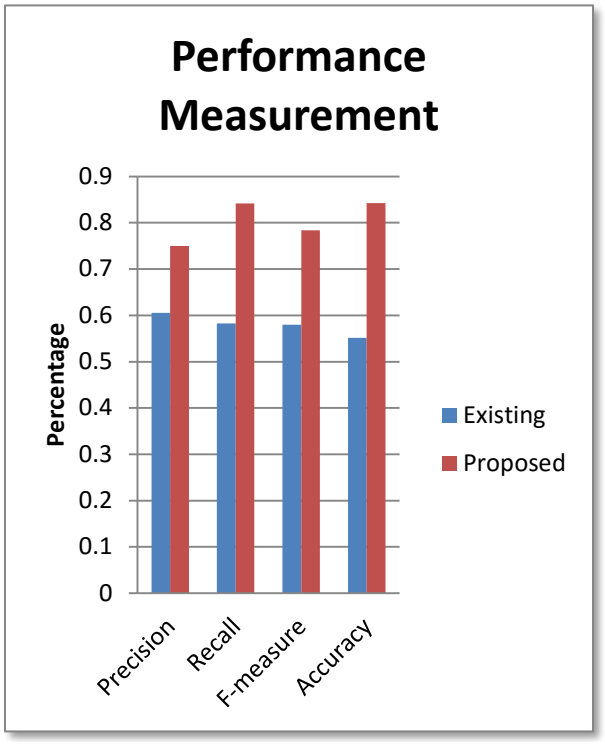

Fig. 3 Graph for performance measurement

5. Conclusion

This research presents sentiment analysis about student's learning experience using Twitter data. Using student's opinion on Twitter we find out the student's emotion (positive, negative). For that select the various positive and negative categories and find the student's emotions. For this used naïve bayse multilable classifier to find correct polarity (positive, negative). To applying ontology model we find the particular tweet is positive or negative. This research is useful for educators, policy makers and students to taking some decision about college. And conclude that student just not post their bad experience but also post good experience about their college life. For the future work take dataset using #lady engineer like related hash tags.

5. References

[1] M. clark, s.sheppard, c.aaatman, " Academic pathways study processes and realities", in proceeding of the American society for engineering education annual conference and exposition, 2008

[2] Xin chen,mihaela,krishna madhavan, "Mining social media data for understanding student learning experience", IEEE transactions on learning technologies, volume 7, issue 3 IEEE 2014

[3] Geetika Gautam, Divakar yadav "Sentiment analysis of Twitter data using machine learning approaches and semantic analysis ", pp. 978-1-4799-5173-4 IEEE 2014

[4] Luiz F.S Colrtta,nadia F. F. da silva,"Combining classification and cluster for tweet sentiment analysis.", Brazilian Conference on Intelligent Systems IEEE 2014

[5] Sara Keretna, Ahmad Hossny, Doug Creighton, "Recognising User Identity in Twitter Social Networks via Text Mining", IEEE International Conference on Systems, Man, and Cybernetics, IEEE 2013

[6] Neha R. Kasture, Poonam B. Bhilare, "An Approach for sentiment analysis on social networking sites", 2015 International Conference on Computing Communication Control and Automation, 2015 IEEE

[7] Nur Azizah Vidyaa, Mohamad Ivan Fananyb, Indra Budib, "Twitter Sentiment to Analyze Net Brand Reputation of Mobile Phone Providers", The Third Information Systems International Conference, SCIENCE DIRECT 2015

[8] steven c, harris, lanqin zheng, kinshik," Multi dimensional sentiment classification in online learning environment", 2014 IEEE sixth international conference on technology for education

[9] Geetanjali S. Potdar1, Prof R. N. Phursule2, A Survey Paper on Twitter Opinion Mining, International Journal of Science and Research (IJSR) Volume 4 Issue 1, January 2015

[10] G. Vinodhini, R. M. Chandrasekaran "Sentiment Analysis and Opinion Mining: A Survey" Assistant Professor, Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar-608002, Volume 2, Issue 6, June 2012, IEEE paper.