Student Feedback Mining System Using Sentiment Analysis

R.Menaha

IT Department

Dr.Mahalingam College of Engineering and Technology

Pollachi, Tamilnadu, India

R.Dhanaranjani, T.Rajalakshmi, R.Yogarubini

IT Department

Dr. Mahalingam College of Engineering and Technology

Abstract- Academic industries used to collect feedback from the students on the main aspects of course such as preparations, contents, delivery methods, punctual, skills, appreciation, and learning experience. The feedback is collected in terms of both qualitative and quantitative scores. Recent approaches for feedback mining use manual methods and it focus mostly on the quantitative comments. So the evaluation cannot be made through deeper analysis. In this paper, we develop a student feedback mining system (SFMS) which applies text analytics and sentiment analysis approach to provide instructors a quantified and deeper analysis of the qualitative feedback from students that will improve the students learning experience. We have collected feedback from the students and then text processing is done to clean the data. Features or topics are extracted from the pre-processed document. Feedback comments about each topic are collected and made as a cluster. Classify the comments using sentiment classifier and apply the visualization techniques to represent the views of students. This proposed system is an efficient approach for providing qualitative feedback for the instructor that enriches the students learning.

Keywords- Students Feedback, Text processing, Clustering, Topic extraction, Sentiment analysis.

1. INTRODUCTION

Students provide feedback in quantitative ratings and qualitative comments related to preparation, contents, delivery methods, punctual, skills, appreciation, and learning experience. The delivery methods and preparation component refers to instructor's interaction, delivery style, ability to motivate students, out of class support, etc. The content refers to course details such as concepts, lecture notes, labs, exams, projects, etc. The preparation refers to student's learning experience such as understanding concepts, developing skills, applying acquired skills, etc. The paper correction refers to correction of mistakes and providing solutions to overcome it. The punctual refers to the class timing and assignment or record submission. The appreciation refers to the comments given when something is done perfectly. Analyzing and evaluating this qualitative data helps us to make better sense of student feedback on instruction and curriculum.

Recent methods for analyzing student course evaluations are manual and it mainly focuses on the quantitative feedback. It does not support for deeper analysis. This paper focus on providing qualitative and quantitative feedback to analyze and provide better teaching to improve the student's performance.

The paper will be structured as follows: Section 2 will review the techniques used in text processing and sentiment analysis approach in the background. Section 3 will describe the related works of the current research about the student feedback mining system. Section 4 will provide the proposed system of this paper. Section 5 will have experiments and future works to be implemented and finally we concluded in section 6.

2. BACKGROUND

Text mining approach is useful in the sentiment analysis process. In this section, we provide a brief description about the methods that we adopted to extract the keywords from the students feedback document.

2.1 Tokenization

Tokenization is the act of breaking up a sequence of strings into pieces such as words, keywords, phrases, symbols and other elements called tokens. Tokens can be individual words, phrases or even whole sentences. In the process of tokenization, some characters like punctuation marks are discarded [4].

2.2 Stop word removal

Stop words are words which are filtered out before or after processing of natural language data. These words are removed to extract only the meaningful information [4]. The list of stop words may be 'the, is, at, which, on, who, where, how, hi, before, after' etc.

2.3 Clustering

Clustering is the process of making a group of abstract objects into classes of similar objects [4]. A cluster of data objects can be treated as one group. While doing cluster analysis, we first partition the set of data into groups based on data similarity and then assign the labels to the groups. The main advantage of clustering over classification is that, it is adaptable to changes and helps single out useful features that distinguish different groups.

2.4 Classification

Data classification is the process of organizing data into categories for its most effective and efficient use [4]. A well-planned data classification system makes essential data easy to find and retrieve. This can be of particular importance for risk management, legal discovery, and compliance.

2.5 Sentiment Analysis

Sentiment analysis refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. Sentiment analysis is widely applied in review and social media for a variety of applications, ranging from marketing to customer service. Sentiment analysis aims in determining the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. The attitude may be his or her judgment or evaluation, affective state, or the intended emotional communication.

3. RELATED WORKS

Francis F. Balahadia; Ma. Corazon G. Fernando; Irish C. Juanatas [7], developed the teacher's performance evaluation tool using opinion mining with sentiment analysis. They collected the feedback from the students and identified the strength and weakness of the particular teacher. They evaluated the qualitative and quantitative data and provided sentiment score of the teacher in a school.

Nabeela Altrabsheh; Mihaela Cocea; Sanaz Fallahkhair [13], reduced the stress and time consuming of analyzing the feedbacks while teaching. To overcome it, they processed automatically using sentiment analysis. They used Support Vector Machine(SVM) to provide a higher level of pre-processing.

Alok Kumar; Renu Jain Feedback [1], proposed an automatic evaluation system based on sentiment analysis. Feedback is collected in the form of running text and sentiment analysis is performed to identify important aspects using supervised and semi supervised machine learning techniques.

4. PROPOSED SYSTEM

We proposed a system to mine the feedback given by the students and obtain knowledge from that and present that information in qualitative way. Feedback was collected for a course; those feedbacks were pre-processed using text processing techniques. In preprocessing, the feedback files are generated as a flat file. The flat file is tokenized into sentences and the keywords are listed after removing the stop words. We have identified the frequency of each word and extract the topic which has the highest frequency count. Similar comments in each topic are clustered and then the clustered words are classified into positive or negative comments. The classified comments are generated as a chart for easy visualization.
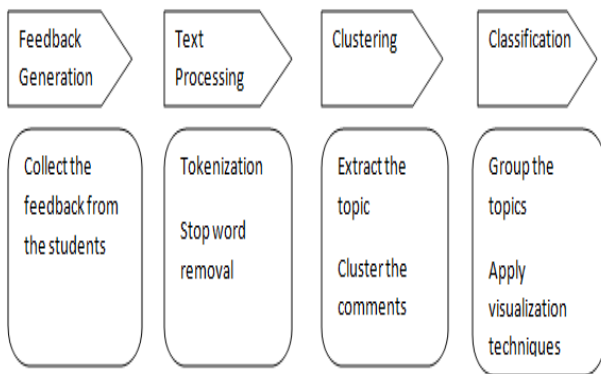
Fig.:1 Block diagram of proposed system

4.1 Keyword Extraction

To clean the data, the collected feedback is subjected to tokenization and stop word removal. The following sequence of steps shows that how do we performed preprocessing.

Step 1: Read each student feedback document and append it into a document D.

Step 2: Tokenize D based on. or, to identify the sentence.

Step 3: On each sentence, remove the stop words.

Step 4: Update the document D.

Step 5: Now the document D is removed from stop words.

Topic Extraction

From the preprocessed document, the parts of a sentence like adjectives, verbs, adverbs, pronouns, nouns, proper noun etc., is removed to identify the topics available in the student feedback. The topic might be teaching, project, communication, interaction, punctuality etc., The following algorithm we have used for the extraction of topics from the feedback document.
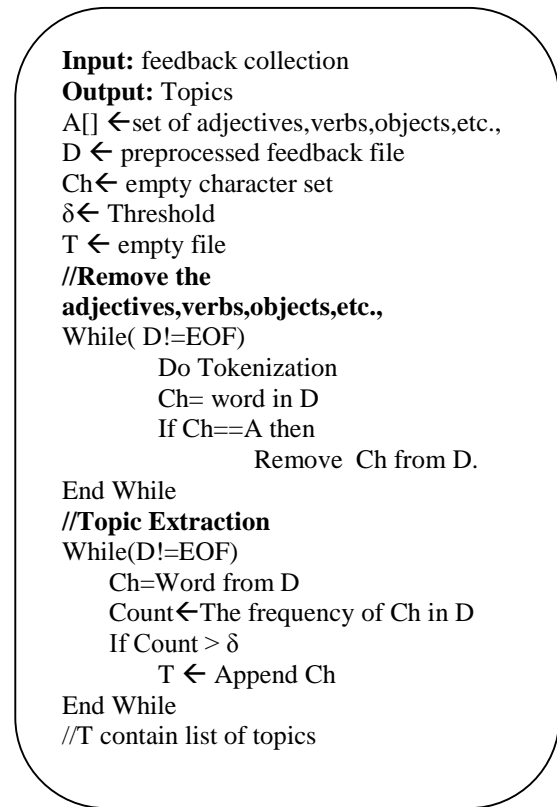
Input: feedback collection
Output: Topics
A[] ← set of adjectives,verbs,objects,etc.,
D ← preprocessed feedback file
Ch← empty character set
δ← Threshold
T ← empty file
//Remove the adjectives,verbs,objects,etc.,
While(D!=EOF)
    Do Tokenization
    Ch= word in D
    If Ch==A then
        Remove Ch from D.
End While
//Topic Extraction
While(D!=EOF)
    Ch=Word from D
    Count←The frequency of Ch in D
    If Count > δ
        T ← Append Ch
End While
//T contain list of topics

Fig 2.:Topic Extraction Algorithm

We have used a threshold δ to limit the number of topics. The frequency of each word in D is counted by using the equation 1. If a word exceeds the δ then it is identified as topic.

$$\mu(w) = \sum \frac{N(w)}{T}$$

Where μ(w) is the frequency of the word w in document D. N(w) is the number of times the word w appeared in D. And T is the total number of words in D.

4.2 Clustering

Each Topic from T is identified as cluster. Read each student feedback document for a course and identify the comments given by the students related to the topic and make it as cluster. We have used pattern matching for comparing the topics with the student feedback document. Semantic similarity can also be used to compute the relatedness between the topic and the feedback comments. The identified topics which are made as cluster by using our approach are Faculty interaction, Punctual, Project, Lab, etc.

4.3 Classification

Read each cluster comments and classify those comments as either positive or negative. An array list is maintained for positive and negative comments. The student feedback comments are compared with the list to conclude that comment is either positive or negative on that topic. The following algorithm shows the sequence of steps employed to perform classification.

Input: Clusters
Output: features
P[] ← set of positive words(good,best,etc.,)
NA[] ← set of negative words(bad,poor,etc.,)
N← No of clusters.
C1 to CN← cluster sets , Ch ← empty character set
//Classification
Classify(C1 to CN)
While(i<N)
    Read Each Cluster from C1 to CN
    Ch← Read each comment from a cluster
    If Ch contains P then
        Positive_count++
        Else if Ch contains NA then
            Negative_count++
End While

Fig.:3 Classification Algorithm

5. EXPERIMENTS

In this paper, we build a Student Feedback Mining System to analyze topics and their sentiments from student generated feedback. The feedbacks are collected from the students for a single course for easy evaluation and to improve student's learning file.

We tokenized the feedbacks into sentences and removed the stop words. Topics were extracted from the feedback document. The following table shows the extracted topics and comments related to it, which were made as cluster is shown in the table 1. From each cluster, the comments are classified as positive or negative which is shown in the following table 2.

S.No	Topic	Clustered comments
1	Faculty interaction	Topics, Teaching, Entertainment, Speaking, Friendly, Help.
2	Punctual	On time, In time
3	Lab	Exercises, Assignments, Experiments, Test, Mini projects,
4	Delivery style	Fast, Slow, Medium, Teaching.
5	Paper correction	Feedbacks, Partial, Impartial, Good

Table 1 Clustering

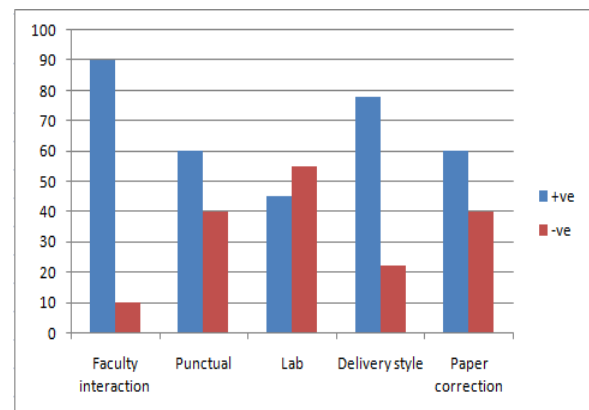| S.No	Topic	+ve	-ve
1	Faculty interaction	90	10
2	Punctual	60	40
3	Lab	45	55
4	Delivery style	78	22
5	Paper correction	60	40

Table 2 Classification

Fig.:4 Chart representation

5. DISCUSSIONS AND FUTURE WORKS

As future work, the proposed system can be extended to include semantic similarity for

clustering the student feedback .And also different visualization techniques would be used to project the results.

6. CONCLUSION

In this paper, a Student Feedback Mining System is build to analyze topics and their sentiments from student generated feedback. This system uses preprocessing, topic extraction, clustering, classification to represent the student views in a graphical way. This system will be useful to improve the students learning and instructor's methods of delivery.

7. REFERENCES

1. Alok Kumar; Renu Jain in *Sentiment analysis and Feedback Evaluation*

2. AnshulMittal,ArpitGoel in*Stock prediction using twitter Sentiment analysis.*

3. Brennan, J. & Williams, R. (2004) *Collecting and Using Student Feedback*. A Guide to Good Practice (LTSN, York).

4. Dhanalakshmi V in *Opinion mining from student feedback data using supervised learning algorithms.*

5. Donovan, J., Mader, C. E., & Shinsky, J. (2012). *Constructive Student Feedback: Online vs. Traditional Course Evaluations.*

6. Elaine Keane & Iain Mac Labhrainn , *Obtaining Student Feedback on Teaching & Course Quality* , CELT, April 2005

7. Francis F. Balahadia; Ma. Corazon G. Fernando; Irish C. Juanatas in *Teacher's performance evaluation tool using opinion mining with sentiment analysis*

8. GokarnIlaNitin ,Asst.Prof.GottipatiSwapna , Prof.VenkyShankararaman in *Analyzing Educational Comments for Topics and Sentiments: A Text Analytics Approach.*

9. HarshaliP.Patil,MohammadAtiquein *Sentiment analysis for Social media.*

10. K. P. Mohanan, *the place of student feedback in teaching evaluation* http://www.cdtl.nus.edu.sg/publications/stud feedback/StudFeedback_Teach Quality.pdf

11. Mark McGuire,ConstanceKampf Aarhus University in *Using Social Media Sentiment Analysis to Understand Audiences: A New Skill for Technical Communicators?*

12. M.S.Neethu,R.Rajasree in *Sentiment analysis in twitter using machine learning techniques.*

13. Nabeela Altrabsheh; Mihaela Cocea; Sanaz Fallahkhair in *Sentiment Analysis: Towards a Tool for Analysing Real-Time Students Feedback*

14. Tan Li Im, PhangWai San, Chin Kim On Center of Excellence in Semantic Agents Universities Malaysia in *Rule-based Sentiment Analysis for Financial News*

15. Yao, Y., & Grady, M. L. (2005). *How do faculty make formative use of student evaluation feedback?*: A multiple case study. *Journal of Personnel Evaluation in Education*, *18*(2), 107-126.

16. Zhao, Y., Karypis, G., & Du, D. Z. (2005). *Criterion functions for document clustering* (Doctoral dissertation, University of Minnesota.).

17. Object Refinery Limited. "JFreeChart". JFreeChart. Jan 2009. Web. 17 Apr. 2015 http://www.jfree.org/jfreechart/

18. The Apache Software Foundation. "Apache Poi Project". Apache POI. Jan 2005. Web. 17 Apr. 2015. https://poi.apache.org/

WEB REFERENCES

1. *URL:http://stackoverflow.com*
2. *URL:http://javaworld.com*
3. *URL:http://programcreek.com*
4. *www.wikipedia.com*