A Lexicon Based Sentiment Analyzer Framework for Student-Teacher Textual Comments

Khin Zezawar Aung

Application Department, University of Computer Studies, Mandalay, Myanmar

Abstract- Opinion mining is a process for tracking the mood of the people about any particular topic by review. Sentiment analysis tries to determine the sentiment of a writer about some aspect and also the overall contextual polarity of a document. This paper presents the sentiment analysis in collaboration with opinion extraction, summarization, and tracking the records of teachers. This paper modifies the existing algorithm in order to obtain the collaborated opinion result. The aim of this paper is to analyze the students' text comments using lexicon based sentiment analysis to predict teacher performance. A database of sentiment words is created as a lexical source to get the polarity of words. In this study, students give their comments on their teacher. Finally, the result of opinion about the teachers are represented as very high, high, moderate, low and very low by evaluating the feelings expressed by students.

Index Terms- Collaborated Opinion, Opinion Mining, Sentiment Analysis, Students' Comment,

I. INTRODUCTION

Opinion mining also called sentiment analysis is a process of finding users opinion about particular topic or a product or problem. A topic can be a news, event, product, movie, reviews, location hotel etc [8].

In general, opinion may be the result of a people's personal feelings, beliefs, sentiments and desires etc. For this reason, opinion mining becomes popular research topic. Analyzing students' comments using sentiment analysis techniques can identify the students' positive or negative feelings. Students' feedback is important because it can help the lecturers to understand the students learning behavior. Sometimes students do not understand what the lecturer is trying to explain, thus by providing feedbacks, students can indicate this to the lecturer.

The sentiment analysis task may be done at different levels; document-level, sentence-level and aspect or feature level. In Document level the whole document is classified either into positive or negative class. Sentence level sentiment classification classifies sentence into positive, negative or neutral class. Aspect or Feature level sentiment classification concerns with identifying and extracting product feature from the source data [3]. This system uses sentiment analysis in sentence by sentence to determine teaching performance by applying the students' comments in university of computer studies, Mandalay, Myanmar. Finally the collaborated opinion about the teacher is represented as very high, high, moderate, low and very low.

II. RELATED WORKS

In recent years, there has been a huge burst of research activity in the areas of sentiment analysis and opinion mining. Hu's and Dave's [9] research focused on extracting opinion from remarks. Hu's research was a product feature based research. It aimed at extracting product features and gave product based summary.

Various tools are also available now for opinion extraction, sentiment analysis and opinion summarization. There have been researches regarding development for better algorithms for such tools. Many algorithms have been proposed in order to understand and implement opinion mining and sentiment analysis. Researchers have developed models for identifying the polarity of words, sentences and whole document [8].

Ku, Liang and Chen [8] proposed algorithm for opinion extraction, opinion summarization and tracking the opinion which may be used for multiple languages. The opinion extraction algorithm takes value of opinion holder into consideration whereas in this paper the value of opinion holder is taken to be one.

R. McDonald [11] proposed document level analysis which has benefit of finer level of classification. He used a structural model with different parameters for document level analysis. He used Viterbi's algorithm for solving the models with different parameters or arguments.

Deepali Virmani, Vikrant Malhotra, Ridhi Tyagi [4] proposed sentiment analysis algorithm for calculating collaborated opinion value. They used bottom_up approach to identify opinions present. They used sentiment word database for analysis of opinion. For such analysis, most of the paper used thesaurus and word net.

LI Caiqiang and Ma Junming [7] proposed online education teacher evaluation model based on opinion mining. This model gets an overall evaluation of each teacher. They didn't indicate the words polarity strength in their work.

Sagum et al. [1] proposed an application of language modeling for faculty comment evaluation. They applied sentiment analysis to determine the rating of positivity or negativity of the comments given by the computer science students in a Faculty Evaluation Form. To get the polarity of words, SentiWordNet was utilized as a lexical source. To identify the polarity, the authors constructed the language model in which the probability of positivity, negativity or neutral occurrence is stored based on data that was fed beforehand. The objective of this paper is to evaluate the level of teaching performance from student comments by using lexicon based approach.

III.   METHODS IN SENTIMENT ANALYSIS

There are two main approaches for sentiment analysis: machine learning based and lexicon based. Machine learning based approach uses classification technique to classify text. Lexicon based method uses sentiment dictionary with opinion words and match them with data to determine polarity. They assigns sentiment scores to the opinion words describing how Positive, Negative and Objective the words contained in the dictionary are.

A.   Lexicon Approach

Opinion words are divided in many categories. Positive opinion words are used to express some necessary things, and negative opinion words are used to describe unnecessary things. Opinion phrases and idioms are also there which together are called *opinion lexicon.*

B.   Dictionary Based Approach

A small set of opinion words is manually collected. Then, this set is grown by finding their synonyms and antonyms in the WordNet [6] and thesaurus. After the system has found new words these words are added to the seed list and the next process starts. This process stops when no new words are found. To remove or correct the errors manual inspection process will be done.

C.   Polarity word Dictionary Construction

In our dictionary the synonym is considered as a positive word and antonym is considered as a negative word. There are 165 positive words and 104 negative words in our lexicon. In our work, the score of opinion words are defined by an expert. The sentiment score ranges from zero to six.

Same word in different domains can have different meanings. In our lexicon 'fast' is negative opinion and its opinion score is 3. For example 'I feel IPAD2 is fast running speed and higher screen resolution'. In this sentence 'fast' is the positive opinion in phone review domain but negative opinion in student review domain. For example 'I understand her teaching but speed of presentation is very fast for other students'. In this sentence 'fast' is negative opinion and 'very' depicts a stronger emotion. The opinion score of 'very fast' is 2. The Dictionary contains the strength of the polarity of every word.

Negation words (e.g. no, not, neither, nor) are very important in identifying the sentiments, as their presence can reverse the polarity of the sentence. For example 'teaching performance is not good'. Although the word 'good' depicts a positive sentiment the negation 'not' reverses its polarity. In the proposed approach, whenever a negation word is encountered in a comment, the opinion score is increased or decreased by a certain amount.

Blind negation words (e.g. needed, required) are also very important in identifying the sentiments. For example 'Her teaching method needed to be better', 'better' depicts a positive sentiment but the presence of the blind negation word 'needed' suggests that this sentence is actually depicting negative sentiment. In the proposed approach whenever a blind negation word occurs in a sentence its polarity is immediately labeled as negative and assigned the opinion score to 3.

D.   Proposed Algorithm

This paper modifies the existing algorithm for identifying the polarity of comments. In the existing work polarity of comments word by word in a sentence was not considered. This algorithm analyses the comments word by word. The algorithm is applied on every comment to identify the polarity of each comment. The algorithm generates a numeric value for the opinion. If the opinion value are high the opinion are considered positive. Lower opinion value represents negative remarks. Sentiment words are identified and a combined value is given to each sentence. A database is maintained to identify the sentiment words. The database along with the sentiment word saves an associated value for the opinion word. The value assigned to each sentiment word is based on how much strong or weak sentiment is being used. If a sentiment word emotes strongly positive opinion higher is its value in the database. A sentiment word that represents strong negative opinion lower is its value in the database. When a sentence is analyzed, for each sentiment word found in the sentence, its opinion value is fetched from the database. Then the collaborated opinion value of that sentence is estimated. If there is negation in a sentence the value of opinion score is decreased/ increased by a certain amount. If there is blind negation word in a sentence, the value of opinion score is immediately determined as negative opinion. The following is the algorithm that is being used.

Input: students' comments
Output: collaborated opinion scores
Begin
1. For each comment
2. Split sentence
3. For each word in sentence
4. If word is blind negation Then
5. Give negative sentiment value from the Database (i.e. =3)
6.   Else If word is positive sentiment Then
7.   Give positive sentiment value from the Database (i.e. >=4)
8.   Else If word is negative sentiment Then
9.   Give negative sentiment value from the Database (i.e. <=3)
10.   Else If word is negation Then
11.   Give reverse sentiment value (i.e. increase or decrease by a certain amount)
12. End If
13. End For
14. Calculate average opinion scores for each comment
15. End For
16. Calculate   average   collaborated   opinion   scores   for   all comments
17. If average collaborated opinion scores is less than 2 Then
18. Give opinion result is "Very Low"
19. Else If average collaborated opinion scores is greater than or equal 2 And less than 3Then
20. Give opinion result is "Low"

21. Else If average collaborated opinion scores is greater than or equal 3 And less than 4 Then
22. Give opinion result is "Moderate"
23. Else If average collaborated opinion scores is greater than or equal 4 And less than 5 Then
24. Give opinion result is "High"
25. Else If average collaborated opinion scores is greater than or equal 5Then
26. Give opinion result is "Very High"
27. End If
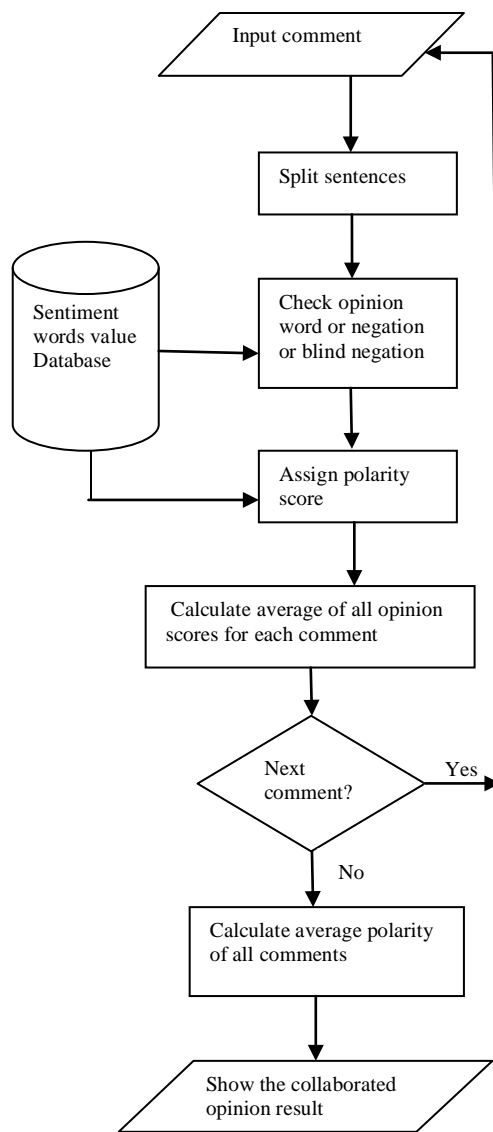28. Show the opinion result
End

IV. SYSTEM ARCHITECTURE

Figure 1. System Architecture

Firstly, we need to input the student's comment to our system. Then preprocessing step is done. This step splits the sentences from one comment. Then this system compares each word in this sentence may be blind negation word or opinion word or negation word by using sentiment word value database. The presence of the blind negation word indicates negative sentiment. If a negation word is found in sentence, the opinion score is increased or decreased by a certain amount. And then assign the polarity scores to each word. After that this system calculates the average opinion scores for each comment for one teacher. If there is next comment to input, the above steps are done repeatedly. If there is no new comment to enter, this system calculates the average polarity of all comments. Finally, this system displays the collaborated opinion result about same teacher.
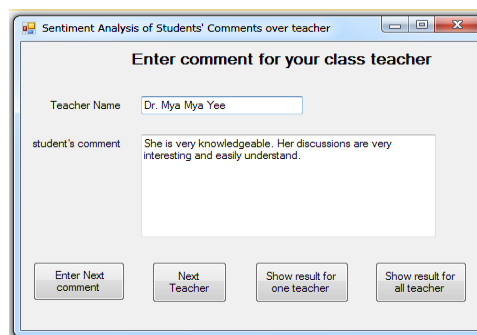
V. EXPECTED RESULT

Figure 2. Input Comment Form

Initially, we gathered comments from the Faculty Evaluation Survey for university of computer studies, Mandalay in Myanmar. There are 1120 students' feedbacks of 30 teachers from this university.

The polarity scores of words in each sentence were summed up, and divided by the total number of subjective words. Then, it was classified as very low, low, moderate, high or very high. Then, the polarities of all comments are averaged. After averaging, the collaborated opinion about the teacher is displayed.

In this section, three comments are tested for sample calculation based on same teacher as shown in Figure 2.

Comment 1: She is very knowledgeable. Her discussions are very interesting and easily understand.

Comment 2: She is very helpful during writing process of papers. But her teaching speed is fast.

Comment 3: Her explanations are clear. Her class notes and outlines are very helpful. She gives enjoyable discussion.

The opinion words of comment 1 are very knowledgeable, very interesting and easily understand. The average of all opinion scores in this comment is 5. So opinion of first student about the teacher is very high.

The opinion words of comment 2 are very helpful, fast. The average of all opinion scores in this comment is 4. So opinion of second student about the teacher is high.

The opinion words of comment 3 are clear, very helpful, and enjoyable. The average of all opinion scores in this comment is 4.333333. So opinion of third student about the teacher is high.

The average polarity score for all comments is 4.444444. So, the result of collaborated opinion about the teacher is high. The collaborated opinion scores of five teachers are displayed in figure 3.
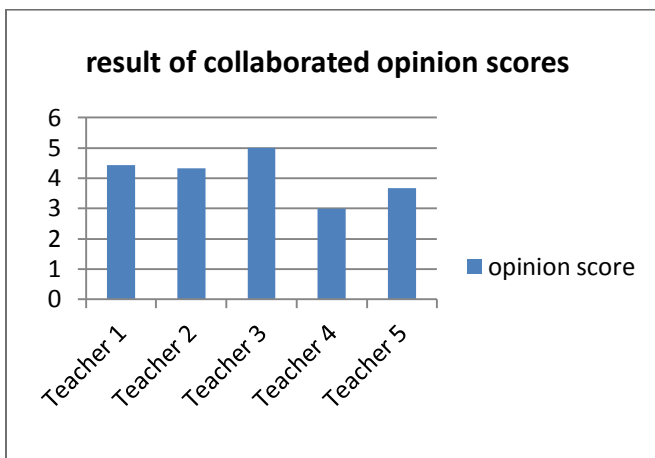
Figure 3. Collaborated opinion results

VI. CONCLUSION AND FUTURE WORKS

Sentiment analysis is now a very useful task across a wide variety of domains. This system used sentiment analysis algorithm to evaluate the performance of teacher from students' comment. A database of English opinion words is constructed for student-teacher domain to identify the polarity of words and their scores. Every opinion word in the database has been given a value. This system calculates the average polarity of all comments which represents their sentiment degrees. Finally this study shows the opinion result of teacher that is represented as very high, high, moderate, low and very low by using bar chart.

Future research may need to consider the analysis of a large scale dataset using aspect based sentiment analysis and a database of Myanmar opinion words will be constructed for such domain of sentiment analysis.

REFERENCES

[1] A. Sagum, Jessiree Grace M. de Vera, Prince John S. Lansang, Danica Sharon R. Narciso, Jahren K. Respeto, "Application of Language Modelling in Sentiment Analysis for Faculty Comment Evaluation," *Proceedings of the International Multi Conference of Engineers and Computer Scientists*, vol.1, March 2015.

[2] C. Kaushik and A. Mishra, "A Scalable, Lexicon Based Technique for Sentiment Analysis," *International Journal in Foundations of Computer Science & Technology (IJFCST),* September 2014.

[3] D. Ankitkumar M, Prof. R. R. Badre, Prof. M. Kinikar, "A Survey on Sentiment Analysis and Opinion Mining," *International Journal of Innovative Research in Computer and Communication Engineering*, November 2014.

[4] D. Virmani, V.Malhotra, R. Tyagi, " Sentiment Analysis Using Collaborated Opinion Mining," *Bhagwan Parshuram Institute of Technology*, New Delhi, India, January, 2014.

[5] G. Angulakshmi, Dr. R. ManickaChezian, "An Analysis on Opinion Mining: Techniques and Tools," *International Journal of Advanced Research in Computer and Communication Engineering*, vol3, July 2014.

[6] G. Miller, R. Beckwith, C. Fellbaum, D. Gross, K.Miller, "WordNet: an on-line lexical database," *Oxford Univ. Press*, 1993.

[7] L. Caiqiang and M. Junming, "Research on Online Education Teacher Evaluation Model based on Opinion Mining," *National Conference on Information Technology and Computer Science (CITCS),* 2012.

[8] Lun-Wei Ku, Yu-Ting Liang and Hsin-Hsi Chen, "Opinion Extraction, Summarization and Tracking in News and Blog Corpora," *American Association for Artificial Intelligence*, 2006.

[9] M. Hu and B. Lu, "Mining and summarizing customer remarks," *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2004, pp. 168-177.

[10] P. Bharathisindhu, S.S.Brunda, "Identifying E-Learner's Opinion Using Automated Sentiment Analysis in E-Learning," *International Journal of Research in Engineering Technology*, vol 3, January 2014.

[11] R. McDonald , K. Hannan, T. Neylon, M. Wells, and J. Teynar, "Structured models for fine-to-coarse sentiment analysis," *Proceedings of the Association for Computational Linguistics (ACL),* June 2007, 432-439.

[12] V. Soni, M.R.Patel, "Unsupervised Opinion Mining From Text Reviews Using SentiWordNet," *International Journal of Computer Trends and Technology (IJCTT)*, vol 11, May 2014.

AUTHOR

First Author – Khin Zezawar Aung, Master of Information Science (M.I.Sc), University of Computer Studies, Mandalay, Myanmar, zezawar12@gmail.com