Analyzing Sentiment in Student-Teacher Textual Comments Using Lexicon Based Approach

Khin Zezawar Aung
Faculty of Information Science, University of Computer Studies, Mandalay, Myanmar zezawar12@gmail.com,
Nyein Nyein Myo
Faculty of Information Science, University of Computer Studies, Mandalay, Myanmar
vicky.mdy@gmail.com

Abstract

Opinion mining, which is also known as Sentiment Analysis, is an interesting field to analyze people's opinions, sentiments, attitudes and appraisals. There are several approaches to analyze the textual data using sentiment analysis. The unstructured textual feedback comments are very complex and the evaluation of these comments is a difficult task. Manual analysis of opinion result takes too long to determine. So, an automated textual analysis is performed using lexicon based approach to predict the teaching performance. Most of the existing sentiment lexicon can't identify some words concerning with the educational domain. There is no teaching sentiment lexicon publicly available for educational domain. So, teaching sentiment lexicon is created for educational domain to get the polarity of words. The experimental results show that the proposed lexicon is more effective than other lexicons for educational domain.

1. Introduction

Sentiment analysis (SA) is a kind of text mining that classify the text into positive, negative and neutral category. SA is now a very helpful task across a wide variety of domains. It can be done at Document level, Sentence level, and Aspect or Feature level [1]. Document level sentiment classification classifies the whole document into positive or negative category. Sentence level sentiment classification classifies sentence into positive, negative or neutral category. Feature level sentiment classification relates to identifying and extracting feature from the source data. This system performs lexicon based sentiment analysis in sentence by sentence to determine the teaching performance by applying the student-teacher textual feedback.

Evaluation of teaching by students is a norm in education institutions. The teaching evaluation generally involves a Likert scale based question or close-ended question in which the students will indicate their opinions by checking how strongly they agree or disagree with fixed questions using multiple choices. Most of the answers of Likert-scale based question contains five levels: strongly agree, agree, neutral, disagree and strongly disagree. Students can choose the desire level. Another evaluation consists of textual feedback from students using open-ended questions [10]. The textual feedback provides an opportunity to students [9] to highlight certain aspects which are not directly covered by Likert-scale based questions. The feedback provides useful insights to both the teacher and the academic administration as the conventional Likert-scale responses are incapable of capturing such aspects [4].

This paper is organized as follows. The related works on sentiment analysis is presented in section 2. The sentiment analysis approaches are discussed in section 3. Section 4 presents the proposed lexicon for teaching evaluation. Section 5 presents the architecture of the proposed system. The experimental result of the teaching evaluation is presented in section 6. Conclusion of this paper is described in section 7.

2. Related Works

This section discusses about the literature reviews of sentiment analysis in many domains. The typical lexicon based sentiment analysis system generally requires three steps. Preprocessing step is done firstly to get the correct result. Secondly, how to identify the opinion word and assign the polarity score are the important part of the lexicon based sentiment analysis approach. Finally, the overall polarity scores are calculated to get the opinion result. So this section presents about the related works of lexicon based approach, machine learning approach and hybrid approach.

In 2015, P. Kaewyong et al. [10] implemented the automatic interpretation system of students' comments to evaluate the teacher performance. They made the qualitative analysis and quantitative analysis of students' freestyle text comment from RateMyProfessor.com using lexicon based sentiment analysis. They consider three items for teacher evaluation such as helpfulness, clarity and easiness to represent the teacher characteristic in each course. The visual correlation results showed that the moving curve between three pairs of those quantitative and qualitative results of students' feedbacks are relatively moving in the same direction. They proof their concept using Pearson's correlation and Spearman's rank to confirm the relationship between each pair of two variables.

In the study of Q. Rajput et al. [12], students' feedback is analyzed automatically using lexicon based approach to improve the quality of teaching in 2016. The Likert scale based quantitative data and textual feedback data from open-ended questions were applied in their research work. The authors used the modified sentiment dictionary from the MPQA corpus to define the polarity of words for the academic domain. They multiplied the word frequency and word attitude to get the overall attitude of a word. Their paper provides new insight into a teacher's performance which is not available from Likert scale based quantitative data using word clouds visualization techniques. The precision, recall and F-measure of their proposed sentiment analysis approach were 0.94, 0.97 and 0.95 respectively for positive sentiment class. The highest precision for negative sentiment class was 0.93 and the highest recall was 0.97 for neutral sentiment (mix of positive and negative).

In 2017, P. Pugsee et al. [11] proposed satisfactory conclusion for cosmetic product review comments to analyze the positive or negative opinion by using sentiment analysis. They used cosmetic sentiment lexicon, modified from SentiWordNet, which contains the special cosmetic product words such as cakey, crepey, orangey and hot girl, etc. These words are not contained in SentiWordNet, so they added these words to SentiWordNet to select the subjectivity of the product review comments. They used Naïve Bayes classifier to generate the classification model. Their experimental results show that the similar proportion of positive and negative data is necessary to train to get the high accuracy and precision of both positive and negative comments.

M. El-Masri et al. [6] proposed a web-based tool for Arabic sentiment analysis in 2017. The authors present a new web based tool that uses sentiment analysis for Arabic text. Their tool is a genetic tool and can be used over multiple domains. This tool allows user to use sentiment analysis to a given topic. User can select parameters including the time of the tweets, preprocessing, features and machine learning techniques. The authors labelled the tweets into positive, negative, neutral and both (both positive and negative equally) using the popular lexicons such as Arabic lexicons, General Inquirer, Bing Liu, MPQA, Slang lexicon, etc. But they didn't test the accuracy of the label data outputs from using these lexicons. There are several limitation in their research. They extracted only 25 tweets and then labelled manually to compare the results of lexicon based and machine learning techniques. They suggest that the lexicon could be improved by adding more terms.

The objective of this paper is to analyze the sentiment in student-teacher textual feedback by using lexicon based approach to evaluate the teaching performance and construct the teaching sentiment lexicon for educational domain.

3. Sentiment Analysis Approaches

This section presents about the sentiment analysis approaches. Three main approaches can be done for sentiment analysis: machine learning based, lexicon based and hybrid approach. Machine learning based of sentiment analysis approach uses training dataset and evaluate the performance of the learned model on the test dataset. This approach uses classification technique to classify text. Lexicon based approach of sentiment analysis uses a lexicon or a sentiment dictionary that contains a list of words associated with sentiment polarity.

Lexicon based approach involves calculating the orientation for a document from the semantic orientation of words or phrases [8]. In generally, the lexicon can be constructed either manually or automatically. M. Hu and B. Liu [7] used an online lexical resource WordNet to predict the semantic orientation of an opinion word. Taboada et al. proposed the lexicon-based approach that determines the polarity of a word by using their constructed dictionaries [8].

Hybrid approach that combine the use of sentiment lexicon and machine learning methods. Most of the hybrid approach uses sentiment lexicon for labelling the training data [2] [11] and uses the classifier to predict the sentiment polarity on the evaluation datasets.

3.1. Opinion Lexicon

The lexicon-based approach depends on opinion (or sentiment) words, which are words that express positive or negative sentiments [5]. Choosing the sentiment lexicon to rely on is very important. The following section describes some popular sentiment lexicon.

3.1.1. Liu Lexicon

Liu lexicon consist of a set of around 6800 English words classified into positive and negative opinion groups. Liu et al. utilized the adjective synonym and antonym sets in WordNet to predict semantic orientation of adjectives. Firstly, a small list of seed adjectives tagged with either positive or negative labels is manually created. This seed adjective list is actually domain independent. For example, great, fantastic, good are positive adjectives; and bad, dull are negative adjectives. The list will be then expanded using WordNet, resulting in a list of 4783 negative terms and 2006 positive terms including misspellings, morphological variants, slang, and social-media markup which are useful for social network data analysis. However Liu lexicon cannot cover all of real world problems in terms of sentiment analysis for educational domain.

3.1.2. Afinn Lexicon

Afinn lexicon was initially set up in 2009 for tweets downloaded for online sentiment analysis in relation to the United Nation Climate Conference (COP15). The old version termed AFINN-96 distributed on the Internet has 1468 different words, including a few phrases. The newest version, AFINN-111 contains 2477 unique words and 15 phrases. AFINN uses a scoring range from −5 (very negative) to +5 (very positive). For ease of labeling the author only scored for valence, leaving out, e.g., subjectivity/objectivity, arousal and dominance. The words were scored manually by the author. The word list in Afinn lexicon initiated from a set of obscene words [3]. Most of the positive words were labeled with +2 and most of the negative words with –2, strong obscene words with either –4 or –5.

4. Proposed Lexicon

Opinion words are defined in teaching sentiment lexicon. Small set of opinion words in students' comments are collected manually and assigned the polarity scores by two experts who have experience in teaching evaluation. The opinion words are increased by searching in the online dictionary, thesaurus for their synonyms and antonyms. For example, synonyms for care are foster, protect, watch, guard, nurse etc. The antonyms for care are neglect, disregard, hate, etc. In this teaching sentiment lexicon synonym is considered as a positive word and antonym is considered as a negative word [10]. But scores of all synonyms are defined differently. The score of great is stronger than the score of good. The word good has opinion score +2 but the word great has opinion score +3. The reliability of the score of opinion words are given by a language expert from the Department of Languages.

This proposed lexicon is based on the idea of Taboada et al. [8]. There are 909 words in teaching sentiment lexicon. Total intensifier words are 30, total positive words are 579 words, total negative words are 296 words and total blind negation word is 4 words. The sentiment score ranges from -3 to +3. The score ranging from +1 to +3 are considered as positive; whereas any negative score ranging from -1 to -3. Blind negation words are need, needed, require, required. Some example words are shown in Table 1.

Table 1. Example words in teaching sentiment lexicon

Opinion word	score	Intensifier word	(%)
guard	+2	actually	+50%
interactive	+2	somewhat	-30%
unexcited	-1	really	+25%
care	+2	very	+50%
knowledgeable	+2	extraordinarily	+50%
low	-1	easily	+25%
understand	+2	much	+50%
realize	+2	fairly	+25%
sleepy	-2	extremely	+100%
complex	-3	absolutely	+50%
comprehend	+2	fully	+50%
aware	+2	completely	+50%
fast	-1	entirely	+50%
complicated	-3	greatly	+100%
available	+2	incredibly	+100%
lazy	-2	totally	+50%
cryptic	-2	quite	+25%
spend	-1	strongly	+50%
long	-2	little	-50%
respond	+2	less	-25%
concern	+2	truly	70%
explanatory	+2	genuinely	+50%

A heuristic technique is used to calculate the semantic orientation score of combining words for automated analysis of student-teacher textual feedback comments. In the following equations, Ws is the semantic orientation score of combining words. Sinf is the intensifier value of word based on 100%. Os is the score of opinion word from teaching sentiment lexicon.

$$Ws=Os \quad (1)$$

$$Ws=(100\%+Sinf) * Os \quad (2)$$

$$Ws=(100\%+Sinf) * (100\%+Sinf) * Os \quad (3)$$

$$Ws=Ws * (-1) \quad (4)$$

If only opinion word is found in sentence, the corresponding positive scores or negative scores are assigned using equation (1). If one intensifier word and one opinion word are found together in the same context, moreover the index of intensifier word must be the reduction of 1 index of the opinion word, then equation (2) is used to get the semantic orientation score of combining words. If two intensifier words and one opinion word are found together in the sentence, moreover the index of first intensifier word must be the reduction of the 2 index of opinion word, equation (3) is used. If a negation word in front of the opinion word is found in the sentence, reversed polarity scores are given by (4).

The semantic orientation score of combining words in all sentences are summed up to get the total polarity scores by (5). In (5), PTs is the total polarity score of all words in all sentences from one comment. Wsi is the semantic orientation score of combining words for i^{th} term in one comment. i is the order of combining opinion words from 1 to n. n is the total number of combining opinion words in all sentences from one comment. T is set of teaching sentiment terms from teaching sentiment lexicon. PTsi is the total polarity score of i^{th} term for all comments. N is the total number of opinion words in all comments. P is the average polarity scores of all sentences in all comments.

$$PTs = \sum_{i=1}^{n} \mathbf{Wsi} \quad , (Wsi \in T) \quad (5)$$

$$P= \sum_{i=1}^{N} \mathbf{PTsi} \;/N \quad (6)$$

The average polarity score of all comments is calculated by (6).

If P>0 the opinion result is defined as positive. If P<0 the opinion result is defined as negative. If P=0, the opinion result is defined as neutral.

5. System Architecture

Figure 1 shows the model of sentiment analysis for student-teacher textual comment. Firstly the teaching evaluation summary of teachers from Stanford University and Baruch College located in New York are collected. Input were the textual comments for of students' feedback comments, which were composed of one or more sentences connected to attitude of instructor, strengths and weaknesses of the textbook, assignments and exams, and the course over-all. Next step is preprocessing. The aim of preprocessing is to remove the unwanted data. This step includes the following tasks:

Case conversion: This step changes all text into lower case.

Split sentence: The comment is divided into sentences.

Replace "n't" with "not". The presence of negation words such as (no, not, neither, nor, nothing, never, none, without) can reverse the polarity scores by using equation (4).

Punctuation removal: Punctuations in a text do not provide the correct result. So, punctuation characters from the word can remove to get the better accuracy.

Tokenization: This process breaks a stream of text into a list of words.

Next step is opinion word identification. This system compares each tokenized word in the comment with blind negation word or positive opinion word or negative opinion word or intensifier word by using teaching sentiment lexicon. And then the polarity scores are assigned to each word by using teaching sentiment lexicon.

The presence of the blind negation word indicates negative sentiment value. After that the total sentiment score of all sentences are calculated by using equation (5) and the average polarity score of all sentences can be calculated by using equation (6). Finally, the opinion result for teacher is displayed for user. This processes are done using the proposed lexicon and Afinn lexicon.

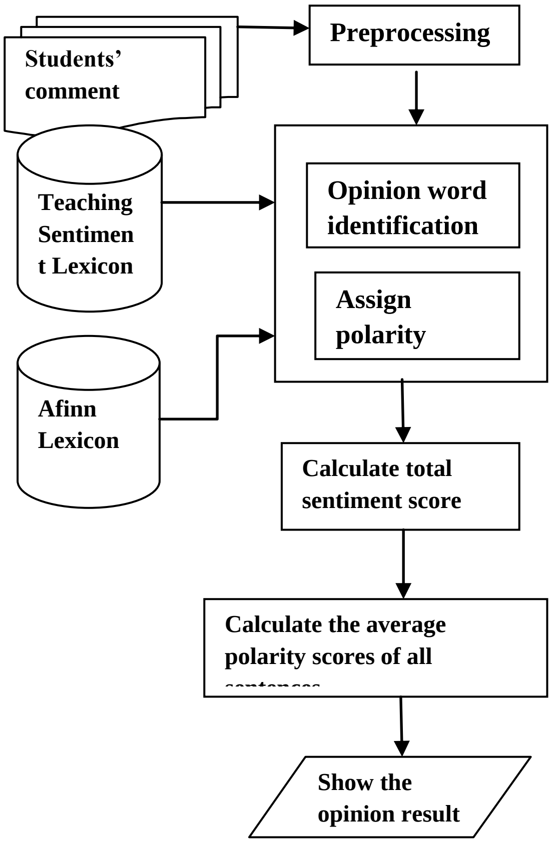

Figure 1. The model of sentiment analysis

6. Experimental Results

Our case study involves 1441 students' feedback comments. Some students' comments and opinion result using the proposed lexicon and Afinn lexicon are depicted in Table 2. Afinn lexicon can't include the words such as explain, interactive, ask, understand, practiced, knowledgeable, available, slow, fast, talented, complex, complicated, sleepy, overcome, concern, etc.

AFINN lexicon considers the sentiment valence scores. The proposed lexicon also considers the valence scores of opinion word. In Liu lexicon, all positive words and negative words are defined separately using text file.

There is no valence score for positive words and negative words. So, the comparison of opinion result using AFINN lexicon is more appropriate than other lexicons such as Liu lexicon, LIWC (Linguistic Inquiry and Word Count), GI (General Inquirer), SentiWordNet lexicon. The SentiWordNet lexicon is very noisy. A large majority of synsets have neither positive nor negative polarity.

Table 2. Sample students' comments and opinion result using two lexicons

Students' feedback comments	Opinion result using the proposed lexicon	Opinion result using Afinn lexicon
I really appreciated the care that Professor Abramitzky showed in making sure his students learned.	positive	positive
The professor was very professional, every day in class he had a detailed plan of what we would be learning. Took time to answer questions.	positive	neutral
Jarrod made the math actually seem understandable. Put it in terms we could understand.	positive	neutral
Very concerned about student's well-being in the course.	positive	neutral
Course materials are confusing and assignments took forever to finish.	negativ e	negative
He was incredibly considerate and caring.	positive	neutral
Explains both the mathematics and intuitive portions of a concept.	positive	neutral
Regular problem sessions by TA would be helpful. Only class lectures were not enough to solve homework problems. More examples should be given at other regular sessions.	negativ e	negative

6.1. Performance Evaluation

The teaching evaluation summary of Jarrod Pickens, who is a mathematics professor, from the Department of Mathematics, Baruch College located in New York, professor Ram Abramitzky from Department of Economics and Kevin J. Ross from Department of Statistics, from Stanford University are collected to calculate the performance. All feedback comments of three professors are manually identified by readers. The performance of this system is tested using four evaluation values: accuracy, precision, recall and F-Measure.

Accuracy (A) = (TP+TN)/ (TP+FP+FN+TN)

Precision (P) =TP/ (TP+FP)

Recall (R) = TP/ (TP+FN)

F-Measure (F1) = (2*P*R)/ (P+R) where, TP is True positive, FP is False positive, FN is False negative. TN is True negative.

The performance of students' comment analysis is shown in Table 3, Table 4 and Table 5.

Table 3. Precision, Recall, F-Measure and Accuracy of Proposed Lexicon

Test File	Precision	Recall	F1	Accuracy (%)
Ram	0.99	0.99	0.99	98%
Kevin	0.98	0.98	0.98	96%
Jarrod	0.98	0.98	0.98	97%

Table 4. Precision, Recall, F-Measure and Accuracy Using Afinn Lexicon

Test File	Precision	Recall	F1	Accuracy (%)
Ram	0.91	0.98	0.95	90%
Kevin	0.98	0.81	0.87	77%
Jarrod	0.95	0.97	0.96	92%

Table 5. Average Precision, Recall, F-Measure and Accuracy Using Proposed Lexicon and Afinn Lexicon

Lexicon	Precision	Recall	F1	Accuracy (%)
Proposed Lexicon	0.98	0.98	0.98	97%
Afinn Lexicon	0.95	0.92	0.93	86%

The experimental results show that the accuracy of the proposed lexicon, teaching sentiment lexicon, is better than Afinn Lexicon.

7. Conclusion

This paper presented a sentiment analysis of student-teacher textual feedback comment using lexicon based approach. The presented approach uses teaching sentiment lexicon to identify the polarity of words for academic domain. Every opinion word in teaching sentiment lexicon has been given a value. The sentiment value is ranged from -3 to +3. This system can analyzes automatically the students' feedback comments into positive, negative and neutral using the proposed lexicon and Afinn lexicon. The accuracy of the proposed lexicon is better than other lexicons for educational domain. This paper suggested that most of the existing lexicon needs to add the teaching sentiment words for academic domain to achieve the better result.

References

[1] B. Liu, "Sentiment Analysis and Opinion Mining", Morgan & Claypool Publishers, May 2012.

[2] C. Pong-inwong, & W. Songpan (Rungworawut), "TeachingSenti-Lexicon for Automated Sentiment Polarity Definition in Teaching Evaluation". 10th International Conference on Semantics, Knowledge and Grids (SKG), Beijing: IEEE, 2014, pp. 84 – 91.

[3] Finn Arup Nielsen, "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs", ESWC2011 Workshop on Making Sense of Micro posts: March 2011, pp. 93-98. http://arxiv.org/abs/1103.2903

[4] G. I. Nitin, G. Swapna and V. Shankararaman, "Analyzing Educational Comments for Topics and Sentiments: A Text Analytics Approach", 2015 IEEE Frontiers in Education Conference (FIE), 2015.

[5] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu, "Combining lexicon-based and learning-based methods for twitter sentiment analysis," Technical Report HPL-2011-89, 2011.

[6] M. El-Masri, N. Altrabsheh, H. Mansour, A. Ramsay, "A web-based tool for Arabic sentiment analysis", 3rd International Conference on Arabic Computational Linguistics, ACLing 2017, November, 2017, pp. 38-45.

[7] M. Hu and B. Liu, "Mining and summarizing customer reviews," ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2004, pp. 168-177.

[8] M. Taboada, J. Brooke, M. Tofiloski, K. Voll and M. Stede, "Lexicon-based methods for sentiment analysis", Association for Computational Linguistics, 37(2), 2011, pp. 267-307.

[9] N. Altrabsheh, M. Cocea, S. Fallahkhair, "Sentiment analysis: towards a tool for analyzing real-time students feedback", 2014 IEEE 26th Internationaal Conference on Tools with Artificial Intelligence, 2014, pp. 419-423.

[10] P. Kaewyong, A. Sukprasert, N. Salim and F.A. Phang, "The possibility of students' comments

automatic interpret using Lexicon based sentiment analysis to teacher evaluation", 3rd International Conference on Artificial Intelligence and Computer Science (AICS2015), 2015, pp. 179-189.

[11] P. Pugsee, P. Sombatsri and R. Juntiwakul, "Satisfactory Analysis for Cosmetic Product Review Comments," Association for Computing Machinery (ACM), Phuket, Thailand, May, 2017.

[12] Q. Rajput, S. Haider, and S. Ghani, "Lexicon-based sentiment analysis of teachers' evaluation," Applied Computational Intelligence and Soft Computing, vol. 2016, September 2016, pp. 1-12.