Research Article

Detecting Polarity Score Of The Course Feedback Text Reviews Using Customized Sentiment Lexicon

[1]Melba Rosalind. J , [2]Dr. S. Suguna

[1]Research Scholar, Madurai Kamaraj University, Madurai -2
melbarosalind@ldc.edu.in
[2]Assistant Professor in Computer Science
Sri Meenaskhi Govt. Arts College for Women(A), Madurai – 2
kt.suguna@gmail.com

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 16 April 2021

Abstract: Commenting on social media about any product, person or event has become a common practice in current trend. This opinion yields and also predicts the reputation about the particular element. So, it is considered most important to detect the exact wavelength of the reviews passed in the social media. This paper aims to facilitate the domain-based sentiment detection by enhancing the existing sentiment lexicon. After enhancing it has attained increased accuracy of 85.1% than the existing lexicon in classifying the positive and negative reviews with fine-tuned pre-processing techniques. And also, the BoW detection has been improved than the existing lexicon. The online course reviews are considered for sentiment polarity detection.

Key terms: opinion mining, sentiment polarity, pre-processing, lexicon, text reviews, BoW

I. Introduction

The rapid increase of web usage has provided wide forum for people to express and exchange ideas and opinions. It also holds the huge productive data to be analysed. So recently more research is carried out in discovering the trend of sentiment about a people, place, product, events, service, organization and many more. In most cases, [4] such large number of information seems unstructured for average internet user. However, it attracted many sentiment analysis researchers towards developing such systems that could assist in analyzing user's reviews efficiently. User generated reviews poses different challenges due to the specialized nature of the online text. Basically, sentiment analysis is carried out at two levels. They are document level and sentence level. The main motivation of our work is to classify the course reviews by sentiment scoring using lexicon-based approach and to increase the efficiency of the general lexicon with some modifications. [10] A basic kind of sentiment analysis is sentiment categorisation – categorising pieces of text into positive and negative sentiment polarity (or valence or orientation). Researchers have investigated sentiment categorisation at the document level (including product reviews), as well as sentence level and even text passage level (including phrases and clauses). There are two main approaches for sentiment analysis [3]: 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 assign sentiment scores to the opinion words describing how Positive, Negative and Objective the words contained in the dictionary are. Bag of Model (BoW) creates a vocabulary of words and each word is associated with the occurrence count. With this, dimensional feature vectors are created for each document. The word feature can be unigram, bigram, trigram and up to n-grams. In this paper, we have considered the unigram word features for analysis with customized sentiment lexicon. One drawback of lexicon-based approaches is that the contextual and domain-specific semantic orientation of a word is generally ignored [6]. Sentiment Analysis (SA) spins around subjectivity classification. In [8] the education domain, students convey their opinions about a teacher's teaching abilities. Due to the unavailability of automated tools to process such feedback, these comments are not properly utilized. [2] Instructors are constantly looking for ways to understand and address the challenges that their students face during the learning process.

The [3] application of sentiment analysis in students' comment was used in various objectives; teaching evaluation, course-online evaluation and teacher evaluation. [9] Sentiment analysis or opinion mining is especially helpful to discover the opinion on product or services. The standard approach is to consider the sentiment (or polarity), negative or positive, as the target of a classification problem.

The paper is presented as follows. Section 2 summarizes the literature study. Section 3 presents the proposed System architecture and proposed algorithm. Section 4 explains the methodology. Furthermore, Section 5 shows the performance measures and comparison of results of our study and section 6 shows evidence of the visualization. Eventually, Section 7 briefly summarizes the overall conclusion and the plan of future work.

II. Literature Study

The application of SA starts from the area of commerce to service domain crossways in the world. It is widely used research topic.

[1]In this paper, a database of sentiment words has been used for analysis of opinion. Every sentiment word in the database has been given a value. When a sentiment word is detected in a sentence the value saved in the database is used for evaluating the cumulative opinion value. The collaborated opinion is evaluated by analysing teacher's remarks word by word and then implementing the algorithm proposed. The evaluated opinion value for a student can be utilized while giving marks to the student. The [2] proposes a system that accepts students' learning diaries as input and then fragments the diaries by the date of each diary entry. It then extracts the emotions present in the diary entry, their negative and positive attributes, and the topics present. An emotional score (eScore) is calculated for each one of Plutchik's eight categories represented in each diary entry to identify how these emotions evolve over the period of time when the diaries are written.

[3] Presents a Lexicon based sentiment analysis that evaluates the level of teaching performance from students' textual feedback comment. A database of English sentiment words is constructed to identify the polarity of words as a lexical source. Average Polarity Score of all Comments is calculated in the range from -3 to +3.

[4] Proposes an integrated rule-based framework for sentiment analysis with emphasis on emoticon classification, proper management of modifiers and negations, performs SWN (SentiWordNet)-based sentiment classification, and improves the classification accuracy and enhances the performance of sentiment classification for domain specific words using domain specific classifier. The Manual Scoring Annotation Scheme is also implemented for assigning polarity scores to those words which are not available in SWN. The proposed method is quite generalized and can classify the sentiments in cross domain.

Authors [5] handle feature-based opinion mining with respect to various features, like Module, Teacher, Exam; Resources are taken for TF-IDF based opinion mining using RapdMiner. Performance of SVM, KNN, NB and NN is presented. The result infers that the feature exam has most negative comments. KNN shows the best precision result of 100%, NB gives the best recall and accuracy result of 97.07% and 99.11% respectively. [6] The hybrid model for sentiment analysis is trained using unigrams, bigrams, TF-IDF and lexicon-based features. Proposed Hybrid Approach (TF-IDF with Domain Specific Lexicon) outperforms other Sentiment Analysis APIs. TFIDF + lexicon scores give highest Accuracy and F-measure values 0.934, 0.926.

[7] Author has developed a domain-specific lexicon, based on a combination of probabilistic and information theoretic weights and compared with SentiWordNet and Generic lexicons. Amazon product reviews of 15 different categories are considered. The domain-specific lexicons achieve an accuracy of 90.09% on average, which is an improvement of 3.25% over the generic lexicon that achieves 86.84%, and it is an improvement of 10.08% over the SentiWordNet lexicon that achieves 80.23% accuracy.

[8] Sentiment Analysis (SA) of students' textual feedback is carried out using Lexicon-based approach. The sentiment score is calculated by summing the Word Attitude and Word Frequency of word in a feedback. A Knime workflow is developed for sentiment analysis. The performance of the sentiment score metric against the Likert based scores is compared and sentiment score results are good in terms of accuracy, recall and f-measure.

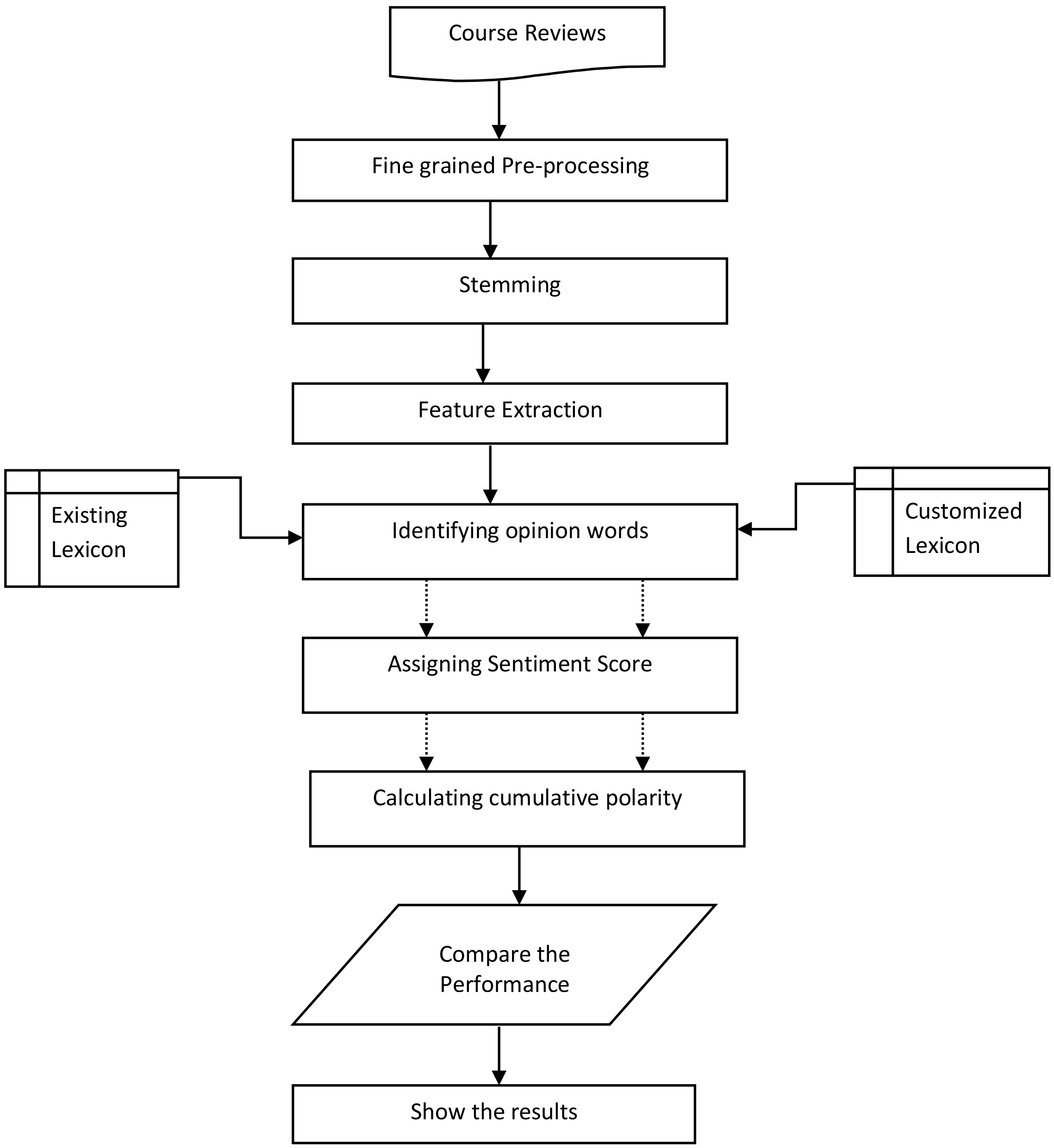
[9] Bing Sentiment lexicon is used to classify the MOOC reviews with category labels through NB prediction model. Unigram features are used. Setting Laplace smoothing reduced the number of not helpful reviews erroneously classified as helpful reviews. Performance metrics such as accuracy, error rate, precision, recall, and F-measure are used to compare the application of different techniques to the Naive Bayes model. Polarity [13] of the words in the collected BBC News articles is computed next using the wordNet lexical dictionary. The sentiment score of whole news article has been calculated using the "extract sentiment" operator. Also by using Score sentiment function based on WordNet and SentiWordNet dictionary, total sentiment score of news article is calculated. A generic, [14] extendable, domain-independent, lexicon-based framework for SA by using existing tools and techniques except for the proposed neutral bias module; The proposed framework over four datasets (52,039 reviews) with five lexicons and results are shown and it outperforms each of individual lexicons. They proposed the impact of each lexicon on the overall accuracy, for any specific dataset, using any specific pre-processing method, given a set of other lexicons. They have also manually annotated the Opinosis dataset for polarity scores which will be helpful for researchers who carry out SA with opinion summarization.

III. Proposed System Architecture

The overall process of proposed model is shown in the system architecture. It consists of data collection, pre-processing, stemming and unigram feature extraction and sentiment polarity calculation using existing and customized sentiment lexicon. Finally the results are compared.

Figure 1. System Architecture

3a. Proposed Algorithm The sequence of process involved in the proposed lexicon-based classification method for implementing the customized sentiment analysis are shown as follows:



We applied following steps to improve the existing Bing lexicon. We implemented the following algorithm, Sentiment Scoring using Customized Sentiment Lexicon (CSL) in R environment using notable text and sentiment analysis packages.

Algorithm Sentiment Scoring using Customized Sentiment Lexicon (CSL)
Input: Course reviews
Output: Sentiment polarity Score

Begin

1. Compose CSL
   a. Add new words with sentiment labels to existing Bing
   b. Modify sentiment labels for specific words
2. Read all entries in the corpus
3. i=0
4. While (i<=(length_ of_corpus))
5. Pass the corpus, C for cleaning
6. Clean the corpus(C) using steps 7 to 11
7. Replace all the special characters with space delimiter
8. Lowercase the text
9. Remove punctuation and special characters
10. Remove numbers and unwanted whitespaces
11. Remove stop words
12. Stem the document using porter's stemming method
13. End While
14. Convert the clean corpus (C) into DTM with length>=3
15. Change DTM into tidy format
16. Do
17. Generate unigram tokens
18. If opinion word found
19. Pass to CSL
20. Assign sentiment score +1 for positive, -1 for negative
21. Else
22. Skip next word
23. End if
24. While(till the end of DTM)
25. if (document D does not contain any opinion word)
26. Discard D
27. Else
28. Calculate cumulative sentiment_polarity for every document d
29. End if
30. Write the polarity score
31. Classify the document based on polarity score
32. End

IV. Methodology

We tried to improve the model performance by implementing the customized lexicon related to academic domain. There are [10] two basic approaches for automatic sentiment categorization – machine-learning approach and lexicon based approach. Machine-learning methods often employ a 'bag-of-words' approach of using words (usually lemmatized or stemmed) in the corpus as independent features in a feature vector to represent the documents. Multiword terms are also sometimes used as features. The value for each word or term feature is usually taken as the term frequency tf (i.e. the number of occurrences of the word in the document) or tf *idf (tf multiplied by the inverse document frequency – the number of documents containing the term). Here, we follow the bag-of-words approach for comparing the performance of existing and customized lexicon. Following steps elaborate the process of the system.

4a. Data Collection

Here, the course reviews given by students of Coursera online portal are collected through Kaggle open repository. Table 1 shows the summary of dataset. 'R Programming' course is selected and total text reviews given for that course is taken for sentiment analysis. In the dataset, the reviews are labelled with predefined sentiment labels as 1 to 5 where 4, 5 denotes positive, very positive and 1,2 denotes negative, very negative text and 3 denotes the neutral text. Using the stems of words the sentiment words are identified and sentiment score is assigned accordingly.

Table 1 Summary of Dataset

Total Number of reviews for the course	1292
Total Number of sentences	2715
Average number of sentences per review	2.4
Total Number of tokens	14902
Maximal Term Length	628

The scope of this work is to boost the performance of sentiment analysis by the following steps:
Step 1: Fine-grained pre-processing
Step 2: Enhancing the existing sentiment lexicon
Step 3: Sentiment polarity calculation

4b. Fine-grained pre-processing

After importing the dataset into R environment, the fine grained text cleaning process was carried out. This includes the following process steps:

Lowercasing – All text reviews are converted into lowercase letter to prepare the data in a consistent manner.

Removing numbers – numbers are not interesting feature in mining the text reviews hence it is removed.

Replacing special characters and punctuation – numbers and punctuation has less contribution for opinion mining, so these are replaced with empty space.

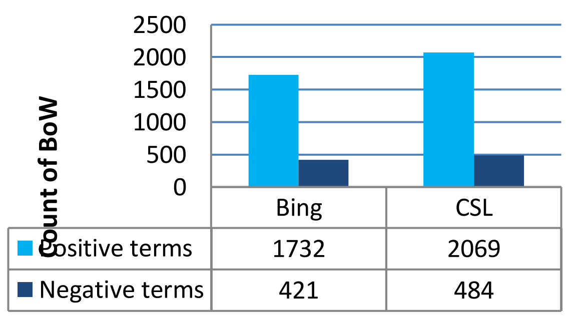
Removing stop words – existing English stop words like *the, but,* etc., are removed from the corpus. Some more words like "year", "mon", "id", "yday", "wday", "mday", "datetimestamp", "within", "coursera" are also removed to enhance the accuracy.

After the above mentioned process, cleaned corpus is transformed to document-term matrix (DTM). Pre-processing steps significantly improves the generation of bag of words. Table 2 presents the improvement of bag of words generated after pre-processing the text reviews for both existing and customized lexicon. The numbers of words generated by the customized lexicon for both positive and negative sentiment are 2069, 484. It is higher than the existing lexicon generation. It identifies only 1732 positive, 421 negative words. Figure 2 shows the total count of BoW generated. [4] Noise reductions steps assist in resolving the data sparseness issue efficiently. Misspelled words present in the review also influence the impact of sentiment polarity score.

Table 2 Bag of Words

Lexicon	**Bag of Words(BoW)**		
	+ve	-ve	Total
Bing	1732	421	2153
Customized Sentiment Lexicon (CSL)	2069	484	2553

Figure 2 Bag of Words (BoW)



4c. Enhancing the Existing Sentiment Lexicon

R [12] tidytext package provides three general purpose lexicons. They are AFINN, Bing and NRC. They are mostly preferred for sentiment analysis. In which,

- AFINN lexicon allots scores with the range between -5 and 5. Positive numbers denote positive sentiment whereas negative numbers denote negative sentiment. It contains 2477 terms in the lexicon.

- Bing lexicon assigns scores into positive (+1) and negative (-1) categories. It contains 6786 words in the lexicon.
- NRC lexicon assigns scores into ten categories including Plutchik emotion categories apart from positive and negative terms. It contains 14,182 unigram words and eight.

Some other dictionaries are also developed by the researchers with specific domain. [11] They are psychological Harvard-IV dictionary, Henry's finance-specific dictionary, Loughran-McDonald finance-specific dictionary and QDAP dictionary. It [10] is well known that a sentiment lexicon needs to be customized for a domain to obtain optimal results; many authors have explored methods to optimize a general-purpose sentiment lexicon for a particular domain or to adapt or extend a sentiment lexicon constructed for one domain to another domain. So, in this work, we added notable sentiment words related to academic domain that are missing in the existing Bing lexicon by examining the students' feedback reviews. The words found in the course reviews with the highest frequency are added to existing lexicon. The words added are related to educational domain. Few new words added in the customized lexicon are given in Table 3. For few words the sentiment orientation is also modified. For example, in the comment, *'quite tough for someone without R-programming background, should add in more explanation in the course to help beginner passing the quiz'* the word *'tough'* projects a negative orientation about the course difficulty whereas it is given as positive in the lexicon. So, we changed it as negative. Similarly in the other review, *'Amazing course. Lots of new tricks learned'. The assignments were challenging and a lot of fun.' T*he word *'tricks' e*xpresses a positive feedback so it is modified with positive sentiment in the lexicon. Similarly for some other words like scratch, challenging also polarity is modified.

Table 3 Sample of new words added in the customized lexicon

Word	Sentiment
rote	Negative
cram	Negative
cover	Positive
rudimentary	Negative
concise	Positive

Conversely, [7] the sentiment of some words change considerably from one domain to another. There are also certain conflicting words present which influence the sentiment orientation based on the domain it is used. For instance according to Bing lexicon, the word '*crash'* is coded as negative word whereas in academic domain, the same word crash used like *crash course* means quick or short term course and perceived as a positive word. Certain reviews that do not contain opinion expressing words are discarded automatically. For instance, the review, '*Everything initial to learning R Programming*.' is not considered for sentiment analysis. The [7] generic score of the variable words often falls in between the minimum and maximum domain-specific score, which reinforces that generic lexicons cannot capture the variation in sentiment for some words. They also present that few words are domain-specific, meaning that they can only be found in a certain domain.

4 d. Sentiment Polarity Calculation

We used Bing sentiment lexicons which are based on unigrams. These lexicons contain totally 6786 English words out of which 2005 are positive words, 4781 are negative words. The ratio of negative words is higher than the positive words. Here, the sentiment polarity is calculated using both Bing and customized lexicons.

For each of the sentence in a given course review, the bags of words are generated. The detected BoWs are assigned with scores for positives or negative opinion. We have calculated the sentiment polarity score by summing all of the scores of opinion words detected in a sentence. And all the consecutive sentences in the review are summed up to find the cumulative score. The Cumulative value expresses the orientation of the review statement. If the cumulative score is lesser than zero it falls under negative category and if t is above zero it is positive. The modifier and intensifiers are not considered in this study and we analyze only the unigrams features of academic domain. Table 4 presents the sentiment score calculation for some reviews. By adding certain opinion words to the existing lexicon we enhanced the sentiment classification. Table 4 shows the difference in sentiment polarity scoring. Here, for sample, for the word tough, the sentiment value is changed as 'negative' whereas it is found as 'positive' in the lexicon. Similarly for the word '*crash*' the sentiment value is changed as 'positive'. The customized Bing lexicon shows the significant difference in sentiment analysis.

Table 4 Sentiment Score Calculation

Sample Reviews	Polarity detected using Bing Lexicon	Polarity detected using Customized Bing Lexicon	Sentiment Label assigned in the dataset
Bit tough and the professor's were rushing through the course in videos.	1	-2	Neutral
This is a crash course into R-Programming. If you have never did any programming this course will be a challenge! But it is a good way to get going with R-programming.	0	2	Positive

V. Performance Measures

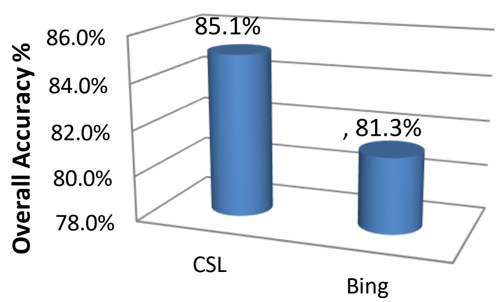
Training the model is not required for this system. So, for evaluating the sentiment polarity detection, we have taken recall and overall accuracy measures.

Accuracy: It denotes the Number of items correctly identified as either truly positive or truly negative out of the total number of items.

Recall or True Positive Rate: It denotes Number of items correctly identified as positive out of the total actual positives

The comparative results show that the accuracy measure of customized lexicon is significantly enhanced as 85.1 after strengthening the domain based sentiment scores whereas for the existing lexicon it shows 81.3%. The overall accuracy % for both is displayed in Figure 3.

Figure 3 Overall Accuracy



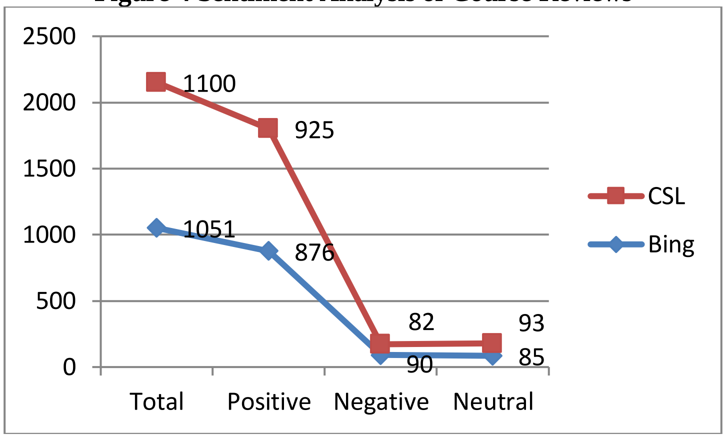
The recall measure of sentiment classification is improved significantly after applying customized sentiment lexicon in the same dataset. We present the classification recall measure % of existing and customized lexicon in Table 5 based on lexicon method.

Table 5 Recall Measure %

Lexicon	Recall %		
	Positive %	Negative %	Neutral %
Bing	83.9%	67.1%	74.5%
CSL	88.6%	61.2%	81.5%

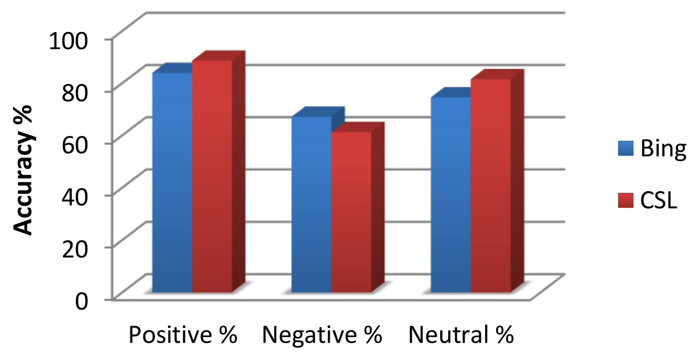
The polarity score detection of positive and neutral reviews is higher than the existing Bing lexicon whereas negative is less than the existing model. The outcome of our sentiment analysis is graphically shown in Figure 4. It is observed that a majority of course reviews fall into the positive categories and very minor percentage of reviews are having negative and neutral sentiments.

Figure 4 Sentiment Analysis of Course Reviews



The recall measure also shows that the customized sentiment lexicon outperforms in detecting positive and neutral sentiments which is 88.6% and 81.5%. The negative polarity scoring is 61.2% less than existing since bing lexicon value that is 67.1%. Bing has larger negative lexicon coverage so it performs well in detecting the negative tokens. The result shows that the customized model shows significant improvement than the existing lexicon in recalling the sentiment features. Figure 5 displays the positive, negative and neutral sentiment classification recall % of both Bing and Customized bing lexicons. Negative sentiment detection has to be enhanced.

Figure 5 Recall Performance



VI. Visualization

The visualization graphs are generated for the evidence of positive and negative words detection. Figure 6 shows the unigram features with negative sentiment; Figure 7 shows the unigram features with positive sentiment. Figure 8 shows the separation of positive and negative words in different colors. Figure 9 shows the top 10 frequent unigram words of the corpus.

Figure 6 Unigram Features with Negative Sentiment

Figure 8 Positive and Negative Word cloud

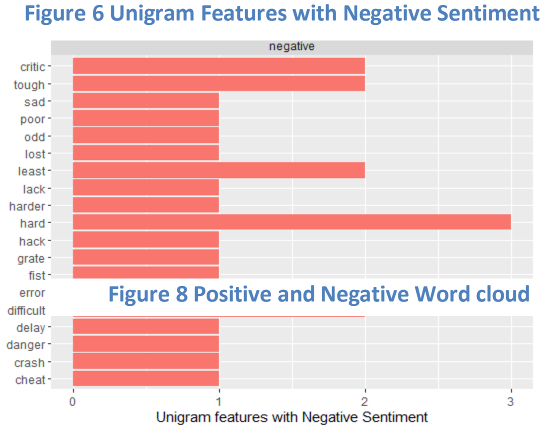
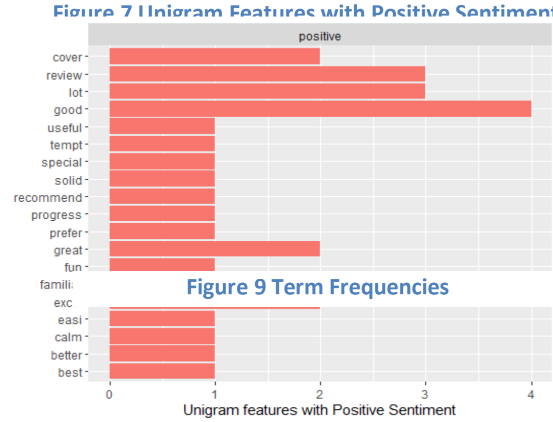
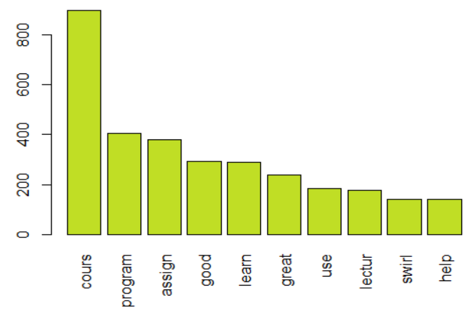


Figure 7 Unigram Features with Positive Sentiment

Figure 9 Term Frequencies





VII. Conclusion and Future work

Detecting exact sentiment of course reviews can be fine tuned by efficient pre-processing procedures and adding domain specific words to the existing lexicon. The experiments show that the customized lexicon, CSL performs better than the existing lexicon at the document level. With the customized sentiment lexicon, detection of opinion words and polarity scoring is also enhanced. Our study will be helpful to enhance the sentiment classification using lexicon based approach for unigram features. Our model outperforms the existing lexicon and gives significant accuracy of 85.1% after customizing the lexicon with academic domain opinion words.

By introducing specific domain specific opinion terms in the lexicon can further improve the model. This model can also be further extended with bigrams. This study also helped us to identify the other issues like contribution of spelling mistakes in opinion reviews classification. These issues can be resolved with proper technique. Our future work will focus on predicting and evaluating the sentiments features classification using machine learning approaches.

References

1. Deepali Virmani, Vikrant Malhotra, Ridhi Tyagi, *Sentiment Analysis Using Collaborated Opinion Mining*, 2016
2. Myriam Munezero et al., *Exploiting Sentiment Analysis to Track Emotions in Students' Learning Diaries*, ACM 2013, http://dx.doi.org/10.1145/2526984
3. Khin Zezawar Aung, Nyein Nyein Myo, *Sentiment Analysis of Students' Comment Using Lexicon Based Approach*, 2017 IEEE
4. Asghar MZ, Khan A, Ahmad S, Qasim M,Khan IA, *Lexicon-enhanced sentiment analysis framework using rule-based classification scheme*, 2017, doi:10.1371/journal.pone.0171649
5. Dhanalakshmi V., Dhivya Bino, Saravanan A. M., *Opinion mining from student feedback data using supervised learning algorithms*, 2016 IEEE
6. Zarmeen Nasim et al., *Sentiment Analysis of Student Feedback Using Machine Learning and Lexicon Based Approaches*, 2017 IEEE
7. Kevin Labille, Susan Gauch, and Sultan Alfarhood, *Creating Domain-Specific Sentiment Lexicons via Text Mining*, 2017 IEEE
8. Quratulain Rajput et al., *Lexicon-Based Sentiment Analysis of Teachers' Evaluation*, Applied Computational Intelligence and Soft Computing 2016, http://dx.doi.org/10.1155/2016/2385429
9. Fetty Fitriyanti Lubis et al., *Improving Course Review Helpfulness Prediction through Sentiment Analysis*, IEEE 2017
10. Christopher SG Khoo, Sathik Basha Johnkhan, *Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons*, Journal of Information Science, 2017
11. Maurizio Naldi, *A review of sentiment computation methods with R packages*, arXiv: 1901.08319v1 [cs.CL] 24 Jan 2019
12. R Core Team (2020). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. URL https://www.R-project.org/.
13. Soonh Taj, Baby Bakhtawer Shaikh, Areej Fatemah Meghji, *Sentiment Analysis of News Articles: A Lexicon based Approach*, 2019 IEEE

14. Mohammed Elsaid Moussa, Ensaf Hussein Mohamed & Mohamed Hassan Haggag, *A generic lexicon-based framework for sentiment analysis*, INTERNATIONAL JOURNAL OF COMPUTERS AND APPLICATIONS, DOI: 10.1080/1206212X.2018.1483813