Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

Literature Survey of Lexicon based Method and Machine Learning Methods of Sentiment Analysis of Student's Feedback

Ms. Shital A. Patil, Dr. Krishnakant P. Adhiya
[1]Research Scholar,
Department of Computer Engg.,SSBT, COET Bambhori ,Jalgaon,
[2]Professor,
Department of Computer Engg., SSBT, COET Bambhori , Jalgaon

Article Info
Page Number: 2257-2261
Publication Issue:
Vol 72 No. 1 (2023)

Article History
Article Received: 15 November 2022
Revised: 24 December 2022
Accepted: 18 January 2023

Abstract
Internet-based applications are rapidly used in educational sector for understanding opinion of students using comment and feedback given by students**.** Process of gathering and analysing student's opinions, thoughts, and impressions regarding various aspects is called sentiment analysis. Student's opinions can be beneficial to Organisation and making decisions based on opinion. However, multiple challenges have been faced evaluation procedure.
Using natural language processing and text mining, Sentiment analysis identifies and extracts subjective information from the text**.** This paper discusses a literature survey of the various method of Sentiment Analysis. This paper also discusses various methods of SA of Student's feedback.

Keywords: -: Feature extraction, feature selection, NLP, text processing, sentiment analysis, dependency features, classification, Aspect Based Sentiment analysis (ABSA), Sentiment Analysis (SA)

1. Introduction

Tracking the mood of the people for about any particular topic by review is called sentiment Analysis. Business and e-commerce applications, such as product reviews, hotel review and movie ratings are popular application of opinion mining. Proposed research work focus on students' feedback comments. Feedback analysis is more important to measure the performance of teacher. Analyzing students' comments using sentiment analysis approaches can classify the students' positive or negative feelings. Students' feedback can highlight various issues students may have with a lecture [17].

2. Literature Survey

For the SA subtask, many studies employed readily available lexicons like WordNet, SentiWordNet and SenticNet. Some authors contributed by proposing lexicons for certain topics. The disadvantage of lexicon-based approaches is that they can become application-specific or fail to recognize new characteristics.

Muslimah Wook, et al. [1] proposed a system opinion mining-based student feedback analysis using machine learning and a lexicon-based approach. The findings of the pedagogical evaluations are revealed using a lexicon-based method and an information extraction technique included in this system. One popular approach to quantifying textual data, linguistic analysis, may be used to determine the overall tone of a student's comments. Word polarity was determined using the Vader Lexicon, an English attitude word collection. The instructor's teaching evaluation findings were obtained by evaluating these sentiment polarities, which contained intensifier words derived from students' input. Opinion findings were improved because of the system's ability to comprehend novel elements like capitalization and emoji. In the end, this brand-new method will supply valuable data to universities to enhance the quality of classroom instruction and syllabi.

Ganpat Singh Chauhan et al. [2] aspect-based sentiment classification for improvement of teaching-learning for students. Every day, an enormous quantity of explanations and background is provided on social networking sites. To extract emotion from free-form text, sentiment analysis is often used. The sentimental analysis of social media

sites has been linked to a reduction in using more conventional methods of soliciting comments and suggestions. Most efforts have been put towards processing user comments, mostly consisting of classifying the positive or negative sentiments using lexicon-predicated or machine-learning algorithms at the document level. Findings indicate that sentiment analysis is useful for gathering student feedback on various topics, although it is seldom used in the classroom.

Laishram Kirtibas Singh et al. [3] describe a literature study of student feedback analysis. The students' responses may largely gauge the state of the workplace and the teacher's effectiveness; by listening to and learning from their comments, teachers can refine their practices and better serve their students. Learners sometimes need a better method for providing feedback, whether it's given verbally or via counselling. Using a student processing model, whether online or offline, makes grading and analysis of student feedback simpler and has the potential to enhance classroom instruction. Surveys and data processing provide significantly more timely and precise feedback. Emotion analysis is a method for determining if a student's emotional state is good, negative, or neutral.

Aadesh N.D. [4] proposed a system of student feedback-based sentiment classification using machine learning techniques SVM, NB and ANN Classifier, and Random Forest are only some of the machine learning algorithms suggested for doing content analysis on student comments. These methods of machine learning are also compared to one another. Based on the data, the Multinomial Naive Bayes Classifier achieves the highest accuracy in practice. They may use the results of this study to enhance the quality of education and the campus overall. Student feedback on a conference, workshop, etc., may also be gathered using this method.

Zemun Kastrati et al. [5] proposed a weekly supervised learning approach for aspect-class student feedback datasets. To successfully identify the component categories highlighted in the unlabeled students' reviews, the proposed method leverages weakly supervised annotating of MOOC-related characteristics and proliferates the weak supervision signal. As a result, it drastically lessens the primary bottleneck of all deep learning techniques: the necessity for manually labelled data. Experiments are conducted using two different quality education datasets: one with around 105k student ratings acquired from Coursera and another with 5989 students' input from more conventional classroom settings. Experimental findings show that our suggested approach achieves excellent outcomes in aspect category recognition and aspect emotion classification.

Saida Ulfa et al. [6] proposed a system of student feedback learning for sentiment classification. This research examined how sentiment analysis methods may be used for online student feedback. Twelve papers discussed the use of sentiment analysis to analyze student responses. ZemunKastrati et al. [7] proposed a system of student feedback for sentiment analysis using NLP and deep learning techniques. Several machine learning and deep learning methods are compared on a dataset consisting of student comments. Several natural language processing methods were analyzed and rated during the sorting process. These included the elimination of stop words, the creation of lemmas, the extraction of stems, and the analysis of dependency parsing. It covers the key obstacles currently in the fields of emotion classification and recommendation.

Rosario Catelli et al. [8] proposed a system lexicon-based and Bert system that were analyzed with machine learning techniques. Using an ad hoc dataset, this research suggests comparing these products in the Italian market, one of the biggest in the world. This study's small dataset reflects the reality that many languages besides English and Chinese must deal with. While evidence usually shows the importance of machine translation like BERT built on deep cognitive networks, it opens several questions about the efficacy and advancement of these alternatives, especially in comparison to those based on dictionaries.

Qik Lin et al. [9] propose two lexical-based methods, knowledge-based and machine learning-based, to automatically extract opinions from short reviews. To begin, evaluating classroom teaching needs methods for sentiment analysis at the sentence and paragraph levels. Second, appropriate teacher modelling may be obtained by amalgamating many forms of student assessment data such as multiple-choice answers, evaluative texts, academic history, intellectual, social network, and research abilities.

Irfan Ali Kandhro et al. [10] proposed a system of student feedback analysis using numerous natural language processing and machine learning techniques. Various classifiers, including ANN, SVM, KNN and multiple supervised machine learning classification algorithms, were utilized to create the model. The outcomes were examined using assessment criteria such as the Confusion Matrix, Precise, Recall, and F-score.

Michelangelo Misuraca et al. [11] proposed a system of opinion-mining techniques for mining educational data on student feedback. Here it provides a method for determining the emotional tone of student evaluations. After

justifying the proposal's mathematical establishment and determining the polarity for positive/negative sentiment categorization. Numerous machine learning algorithms have been used for classification, such as ANN, SVM, NB and random forest. The SVM outperforms the highest accuracy at 96.50% on the heterogeneous dataset.

Metadata Mohammed Almosawi and Dr Salma A. Mahmood [12] proposed a lexical-based sentiment analysis approach to determine the polarity of students' feedback. Using a free-form online survey, they could compile a comprehensive dataset. This data collects a lexicon of 2,217 terms related to higher learning in both the Iraqi dialect spoken in southern Iraq and contemporary Arabic. The suggested method of sentiment analysis was 98% accurate. Various ML methods were used, including Naive Bayes, Support Vector Machines, and the Kernelized Naive Neighbor Finder. Concerning the lecturer's performance, the gathered data shows that 60% of the students had a favorable impression, whereas 40% had a negative one.

HtarHtar Lwin et al. [13] proposed a system of education data mining using feedback analysis with machine learning techniques. In this article, we'll look at how to put into practice a method for analyzing feedback that uses both quantitative ratings and qualitative remarks. It utilized numerical ratings and free-form text comments as our datasets of choice. When grouping ratings, the K-means clustering technique is used. After obtaining a tagged dataset via the clustering process, classification models may be constructed utilizing various classification techniques. It employed the Naive Bayes classifier for training a model, then used 10-fold cross-validation to sort comments in the test dataset by positive and negative sentiment. Additionally, it tested the model by trying to categorize some free-form languages that had not been classified.

Rahil Nawaz et al. [14] AI and machine learning-based feedback analysis for educational data mining and student feedback data. Additionally, it proposes a unique automated analytic methodology for mining useful insights from students' open-ended replies to survey questions.

Giuseppe Varvara et al. [15] proposed a system challenge of dental education on student feedback data using NLP and machine learning techniques. The purpose of this research was to ascertain how undergraduates see the incorporation of these strategies into their coursework. All in all, 353 first- through sixth-year students at the students. The students were surveyed using an online survey created in Italian using "Google Forms" and sent via email. There were three sections to the questionnaire: the first asked for demographic information like age, sex, and course year; the second contained multiple-choice questions about how they felt the e-learning was going; and the third contained two free-form questions about the benefits and drawbacks of the new methods of instruction

3. **Methods**

The feature selection methods that have been investigated are as follows:

3.1 Term frequency (TF)

The term frequency count is used to pick features in this method. The term frequency of each feature is computed for each category. For feature selection, a threshold is specified. In each category, features with a word frequency higher than '2' are chosen. As a consequence of this, a term frequency matrix for each category is produced. In addition, a compound matrix comprising words and their occurrence counts in all categories is created. A binary train matrix is produced from this matrix, with '1' representing non-zero term frequency.

3.2 Weighted Term Frequency

These techniques utilized the weight of each term is calculated using equation in this method (1). The conditional probability of a term is X (t,k), where X t is the overall occurrence count of a term "t" in all categories and X (t,k) is the occurrence count of term "t" in category "k." The weight of "t" rises if the percentage of occurrence of a word "t" in category "k" is higher than the other categories. For each category, a weight threshold is set. To create a binary train matrix, terms (features) with a weight higher than threshold are chosen. Kim Schouten et al. [14] perform a weighted term computation as well. This paper takes a similar approach to [14], proposing a hybrid method for feature selection that uses correlation to prevent feature duplication. Weights are utilized to identify the category of the test phrase in [8], and weights are employed for feature selection and creation of a binary train matrix in this method.

$$\text{weight}(t) = \frac{X_{t,k}}{X_t} \qquad (1)$$

3.3 Term Frequency with Correlation Coefficient

To improve the classification accuracy of the classifier, features must be useful but not redundant. The term frequency matrix produced in I is utilized in this approach. The information gained from this matrix is useful, however it is also redundant. To prevent duplication, each feature's correlation is computed in relation to other characteristics in the same category. To calculate correlation, the Pearson correlation coefficient is employed.

$$C0weight\ [t_i] = \frac{n\ (\sum X[]Y[]) - (\sum X[]) * (\sum Y[])}{\sqrt{[n\sum X^2 - (\sum X)^2]}\sqrt{[n\sum Y^2 - (\sum Y)^2]}} \quad (2)$$

The correlation of each term with regard to other terms is calculated using Equation (2), where "x[] and y[]" are vectors of term t_i and t_{i+1}, respectively, containing term frequency with respect to each category. The average of the correlation of a word "t" with other terms in that term category. To create a binary train matrix, terms with a correlation value of less than or equal to 0.85 are chosen.

3.4 Weighted Term Frequency with Correlation Coefficient

The weighted matrix produced is utilized to create a new matrix containing the weight of a word in relation to each term category in this method. The correlation of each word with regard to other terms is calculated using Equation (2), where "x[] and y[]" are vectors of term t_i and t_{i+1}, respectively, including the weight of term "t" in each category. Finally, as previously stated, a binary train matrix is produced (iii). This article makes a contribution by proposing a supervised method for term extraction that chooses important features while avoiding duplication by assessing feature correlation. The obtained findings indicate that the weighted word frequency with correlation method has a higher F-score than the other approaches. It is discovered in this study that characteristics chosen using weighted word frequency are more relevant, but also redundant. By measuring the correlation between characteristics in a term category, redundancy is eliminated.

4. Conclusion

Using natural language processing and text mining, Sentiment analysis identifies and extracts subjective information from the text. In this paper, we discusses a literature survey of the various method of Sentiment Analysis. This paper also discusses various methods of SA of Student's feedback .

References

[1] Wook, Muslimah, et al. "Opinion mining technique for developing student feedback analysis system using lexicon-based approach (OMFeedback)." Education and Information Technologies 25.4 (2020): 2549-2560.
[2] Chauhan, Ganpat Singh, Preksha Agrawal, and Yogesh Kumar Meena. "Aspect-based sentiment analysis of students' feedback to improve teaching–learning process." Information and communication technology for intelligent systems. Springer, Singapore, 2019. 259-266.
[3] Singh, Laishram Kirtibas, and R. Renuga Devi. "Student feedback sentiment analysis: A review." Materials Today: Proceedings (2021).
[4] Dsouza, DaneenaDeeksha, et al. "Sentimental analysis of student feedback using machine learning techniques." International Journal of Recent Technology and Engineering 8.14 (2019): 986-991.
[5] Kastrati, Zenun, Ali Shariq Imran, and ArianitKurti. "Weakly supervised framework for aspect-based sentiment analysis on students' reviews of MOOCs." IEEE Access 8 (2020): 106799-106810.
[6] Ulfa, Saida, et al. "Student feedback on online learning by using sentiment analysis: a literature review." 2020 6th International Conference on Education and Technology (ICET). IEEE, 2020.
[7] Kastrati, Zenun, et al. "Sentiment analysis of students' feedback with NLP and deep learning: A systematic mapping study." Applied Sciences 11.9 (2021): 3986.
[8] Catelli, Rosario, Serena Pelosi, and Massimo Esposito. "Lexicon-based vs. Bert-based sentiment analysis: A comparative study in Italian." Electronics 11.3 (2022): 374.
[9] Lin, Qika, et al. "Lexical based automated teaching evaluation via students' short reviews." Computer Applications in Engineering Education 27.1 (2019): 194-205.
[10]Kandhro, Irfan Ali, et al. "Student feedback sentiment analysis model using various machine learning schemes: a review." Indian Journal of Science and Technology 12.14 (2019): 1-9.
[11] Misuraca, Michelangelo, GermanaScepi, and Maria Spano. "Using Opinion Mining as an educational

analytic: An integrated strategy for the analysis of students' feedback." Studies in Educational Evaluation 68 (2021): 100979.
[12] Almosawi, Mrtdaa Mohammed, and Salma A. Mahmood. "Lexicon-Based Approach For Sentiment Analysis To Student Feedback." Webology (ISSN: 1735-188X) 19.1 (2022).
[13] Lwin, HtarHtar, et al. "Feedback analysis in outcome base education using machine learning." 2020 17th International conference on electrical engineering/electronics, computer, telecommunications and information technology (ECTI-CON). IEEE, 2020.
[14] Nawaz, Reheel, et al. "Leveraging AI and Machine Learning for National Student Survey: Actionable Insights from Textual Feedback to Enhance Quality of Teaching and Learning in UK's Higher Education." Applied Sciences 12.1 (2022): 514.
[15] Varvara, Giuseppe, et al. "Dental education challenges during the COVID-19 pandemic period in Italy: Undergraduate student feedback, future perspectives, and the needs of teaching strategies for professional development." Healthcare. Vol. 9. No. 4. MDPI, 2021
[16] Mayur Wankhade, Annavarapu Chandra Sekhara Rao,Chaitanya Kulkarni1 "A survey on sentiment analysis methods, applications, and challenges" Artifcial Intelligence Review (2022) 55:5731–5780
https://doi.org/10.1007/s10462-022-10144-1
[17] Khin Zezawar Aung, Nyein Nyein Myo "Sentiment Analysis of Students' Comment Using Lexicon Based Approach" 978-1-5090-5507-4/17/$31.00 ©2017 IEEE ICIS 2017, May 24-26, 2017, Wuhan, China,IEEE Computer Society

Ms. Shital Abhimanyu Patil, completed B.E from Godavari college Of Engineering, Jalgaon. and ME Computer Science & Engineering from SSBT COET Bambhori. She is working as Assistant Professor in SSBT's College of Engineering and Technology since 2006.She is pursuing PhD in Computer Science & Engineering from Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon. Her areas of interest are Machine Learning, Data Analytics and Natural Language Processing.

Dr. Krishnakant Prabhudas Adhiya has completed PhD degree in Computer Science and Engineering from North Maharashtra University Jalgaon. Currently he is working as Professor, Department of Computer Engineering, SSBT's College of Engineering & Technology, Jalgaon and recognized PhD Guide in Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon. He has 44 research papers in reputed peer reviewed journals in addition to 14 papers in International Conferences to his credit. He is Professional Member in ACM, Life member of ISTE. His research interests are Wireless Sensor Networks and Security, Machine Learning, Block Chain and Natural Language Processing