Online-Journals.org

iJET International Journal of Emerging Technologies in Learning

iJET | eISSN: 1863-0383 | Vol. 18 No. 18 (2023) | OPEN ACCESS

https://doi.org/10.3991/ijet.v18i18.42521

PAPER

An Evaluation Framework for Online Courses Based on Sentiment Analysis Using Machine Learning

Jinjing Zeng[1,2], Kailing Luo[3,4], Yu Lu[5](✉), Mingfen Wang[5]

[1]Information Resources Research Institute, Fuzhou University, Fuzhou, China

[2]Library, Fujian Agriculture and Forestry University, Fuzhou, China

[3]College of Education, Fujian Normal University, Fuzhou, China

[4]Fuzhou Gulou Experimental Primary School, Fuzhou, China

[5]Concord University College, Fujian Normal University, Fuzhou, China

media@fjnu.edu.cn

ABSTRACT

Online course evaluation is critical to both course selection and teaching effectiveness for students and teachers. However, the current online course evaluation methods have been criticized for neglecting learners' needs and their inefficiency. Therefore, a course evaluation framework based on sentiment analysis using machine learning is proposed in this study to analyze a large number of online course review comments from learners. Initially, massive open online course review comments were collected through web crawling. Then, sentence- and aspect-based sentiment analyses were performed. Finally, a list of aspect terms that reflected the learners' requirements was compiled based on the model-generated outcomes. The model was utilized to evaluate an online intellectual property law online course. Results demonstrate that the training models built in this study achieve over 90% accuracy and that 90%–95% of learners are satisfied with the intellectual property law online course. The learners are particularly satisfied with the teacher's teaching style and course schedule. However, the models also highlight the insufficient interactivity in the class and the scarcity of novel course cases. The proposed framework provides a learner-centric approach to evaluating online courses, thereby enhancing the credibility of online course evaluation. This framework also serves as a practical reference for online course recommendation and construction.

KEYWORDS

online course evaluation, sentiment analysis, machine learning, course review comment, intellectual property law

1 INTRODUCTION

Online courses are introduced to address certain challenges in learning, such as offline teacher shortages and the uneven distribution of high-quality educational resources, by leveraging their openness and ability to share knowledge [1, 2]. Following the COVID-19 pandemic, online learning has gained tremendous popularity among students seeking educational materials online [3]. Course selection plays a crucial role in achieving desired learning outcomes, and informed decision making

Zeng, J., Luo, K., Lu, Y., Wang, M. (2023). An Evaluation Framework for Online Courses Based on Sentiment Analysis Using Machine Learning. *International Journal of Emerging Technologies in Learning (iJET)*, 18(18), pp. 4–22. https://doi.org/10.3991/ijet.v18i18.42521

Article submitted 2023-05-22. Revision uploaded 2023-07-01. Final acceptance 2023-07-02.

relies on reliable course evaluations. However, the traditional online course evaluation methods, such as direct manual grading, are susceptible to significant subjective biases and arbitrary distinctions, resulting in discrepancies between the assigned grades and actual course quality [4].

To mitigate the impact of manual grading on course evaluation, there has been growing interest in studies that assess course quality by interpreting and analyzing course review comments. Previous studies [5] show that these review comments serve as valuable sources of feedback that encompass a wide range of course aspects, including the overall situation, learning difficulties, teaching quality, and curriculum resources. These review comments represent the authentic experiences and recommendations of learners who have completed a course. However, a common tendency among learners is to skim through only a limited number of abundant review comments, potentially resulting in the oversight of vital information that is crucial to their decision making when selecting courses.

To enhance the objectivity and comprehensiveness of evaluation results, sentiment analysis is conducted on course review comments to assess the teaching quality, platform construction level, and user satisfaction [6]. By leveraging sentiment analysis, courses can be evaluated in an objective manner. However, the extraction of aspect terms from review comments poses a burdensome task that warrants further attention.

Machine-learning-based methods have significantly reduced the manual effort required for sentiment analysis, leading to their widespread application in analyzing review comments across various domains, such as social media, product evaluation, and business investment. These methods have also been adopted to analyze the sentiment of course review comments. For example, Wang et al. [7] proposed the lite bidirectional encoder representation from transformers (BERT) bi-directional long short-term memory (ALBERT-BiLSTM) sentiment analysis model for multiple open online course (MOOC) review comments. Alaa et al. [8] used the salp swam algorithm, a long short-term memory classifier, to predict the emotions of students based on their course feedback. Qi et al. [9] used latent Dirichlet allocation (LDA) to develop a course evaluation system that describes various aspects in online course review comments. While these machine-learning-based methods have significantly improved the efficiency of sentiment analysis for course evaluation, they often suffer from limitations in aspect classification logic. Therefore, the data obtained through machine learning should be interpreted using existing evaluation systems to enhance the persuasiveness of course evaluation results.

The current course evaluation systems primarily target teachers or online platforms, but they often lack valuable feedback from learners. In addition, only a few course evaluation methods consider learners' requirements to effectively capture the key aspects and aspect-based sentiment of course review comments. An evaluation system that can seamlessly combines both of these approaches must be built to effectively address the needs of stakeholders involved in the course evaluation process.

To address these gaps, this study proposes a comprehensive online course evaluation framework that incorporates learners' requirements. This framework utilizes machine learning techniques, including sentence- and aspect level sentiment analyses, to analyze the sentiment polarity of a large volume of online review comments. The evaluation results aim to answer the following critical questions for any online course:

A) How can learner satisfaction with a course be gauged?
B) How can the course's strengths and weaknesses be evaluated by identifying the teaching staff, curriculum, platform, and learner experience?

2 STATE OF THE ART

2.1 Analysis models

In recent years, sentiment analysis has been utilized in online course evaluation by analyzing course review comments. Various approaches have been employed, including document-level sentiment analysis, sentence-based sentiment analysis (SBSA), and aspect-based sentiment analysis (ABSA) [10]. These techniques enable a comprehensive assessment of the sentiment expressed in course feedback.

SBSA offers a holistic evaluation of the sentiment polarity in review comments. For instance, Sudhakar et al. [11] utilized a fuzzy neural network to classify sentences based on their emotions, which led to the development of an expressive text-to-speech system. Hayashi et al. [12] proposed an SBSA method that considers word importance based on word embeddings. However, this approach does not adequately capture the aspect-based sentiment of review comments, thus limiting its effectiveness in assessing specific aspects of a course.

Relative to SBSA, ABSA is a more detailed approach for the sentiment classification of objects or entities within a corpus, thus yielding more specific analysis results [13]. For instance, Liang et al. [14] proposed an ABSA method that detects sentiment polarity toward a given aspect via affective knowledge-enhanced graph convolutional networks. ABSA includes two tasks—namely, aspect term extraction (ATE) [15] and aspect sentiment classification (ASC)—to determine the sentiment associated with particular aspects in review comments. However, ABSA may not be suitable for analyzing the overall sentiment polarity of a sentence and may not cater to users seeking a quick understanding of the sentiment expressed in review comments.

Combining both SBSA and ABSA yields highly comprehensive analysis outcomes for overall and aspect-based sentiment [16]. However, existing research predominantly uses either SBSA or ABSA individually to evaluate course review comments, which restricts their ability to swiftly, comprehensively, and deeply understand the results.

2.2 Analysis method

Machine learning methods for the sentiment analysis of online course review comments can be categorized into traditional machine learning and deep learning [17].

Traditional machine learning methods are often combined with sentiment dictionary approaches to address certain challenges, such as the model portability and short-text sentiment analysis of online courses. These methods typically utilize algorithms, such as support vector machine [18], maximum entropy, and Naive Bayes, to enhance model generalization and improve classification performance on short texts. For instance, Osmanolu et al. [19] used the triple Likert method and machine learning techniques to conduct a sentiment analysis of online course feedback, while Sudhir et al. [20] used a synthetic dictionary and ambiguity mitigation for sentiment analysis. However, feature engineering, which plays a crucial role in ensuring the effectiveness of machine learning methods for sentiment analysis, often requires substantial manual effort.

Deep learning has emerged as a prominent approach for the sentiment analysis of course review comments primarily due to its ability to effectively process nonlinear information hierarchically, thereby eliminating the need for feature engineering. Compared with traditional machine learning, deep learning exhibits greater model generalization capabilities. In his empirical analysis, Onan [21] demonstrated that in the sentiment analysis models for educational data mining, deep-learning-based architectures outperform integrated learning methods and supervised learning approaches. Wang et al. [7] specifically designed the ALBERT-BiLSTM model for MOOC courses that effectively addressed the low accuracy in handling polysemous words and familiar words with contextualized meanings. However, the current research models in ABSA often employ separate sequential steps for ATE and ASC, thereby creating a gap in analyzing the emotional polarity of aspect words and providing a comprehensive summarization [22]. Moreover, these models have not been applied in analyzing course review comments, thereby underscoring the need for further research.

2.3 Evaluation indicators

Most of the existing online course evaluation indicators are created from the perspectives of teachers (or course developers) and learners.

Several evaluation standards have been devised from the standpoint of teachers or course developers [23, 24]. For example, the UOOC Alliance issued the *MOOC Quality Evaluation Form* that encompasses several aspects, such as teachers, course content, and platform. Similarly, the *Evaluation Criteria of Online Courses in Chinese Universities* comprises four dimensions; namely, teachers, course teaching, system support, and learning effect. However, these evaluation systems often overlook the requirements of learners.

To address this concern, several indicators derived from the learners' perspectives and experiences have been investigated with the aim to develop learner-centered evaluation metrics [25]. The crucial factors that learners consider when evaluating online courses include teachers, course materials, and layout [26, 27]. Ardiasih et al. [28] identified four dimensions of course quality based on the learners' viewpoints; namely, course content, teacher characteristics, video quality, and instructional design. Li et al. [5] found that learners' emotions influence course evaluation outcomes and then proposed a method for extracting learners' opinions and suggestions using a part-of-speech combination pattern for course feature-opinion words. Fan et al. [29] and Douglas et al. [30] analyzed learners' behaviors and performance to provide recommendations for MOOCs. Despite emphasizing the significance of incorporating learners' emotions in online courses evaluation, these indicators have not achieved widespread adoption compared with the aforementioned evaluation criteria. Therefore, a comprehensive and standardized online course evaluation system is still lacking.

To address these shortcomings, this study proposes an online course evaluation framework that incorporates both SBSA and ABSA. This framework facilitates a simultaneous analysis of overall sentiment and aspect-based sentiment, thus ensuring a comprehensive and in-depth evaluation. By leveraging deep learning techniques, this framework enhances the efficiency of sentiment analysis for a large volume of course review comments while minimizing labor input. Furthermore, this framework utilizes the local context focus–aspect term extraction and polarity classification (LCF-ATEPC) model to extract aspect words and their corresponding sentiment polarity simultaneously, thereby improving the efficiency of ABSA. Building upon existing online course evaluation standards, this study combines the outcomes of SBSA and ABSA in categorizing aspect words and then analyzes their sentiment while taking into account the authority and learner relevance of online course evaluation systems. This integrated approach enhances the credibility and applicability of the evaluation results.

3 FRAMEWORK FOR ONLINE COURSE EVALUATION

In the proposed framework, course review comment aspects are classified by integrating the sentiment analysis results with the evaluation indicators derived from the perspectives of learners and course developers. To analyze online course satisfaction, attention, and characteristics, the length of review comments and the sentiment polarity of the identified aspects are considered. By leveraging large-scale data and pedagogical theory, the proposed approach enhances the efficiency, accuracy, and validity of online course evaluations.

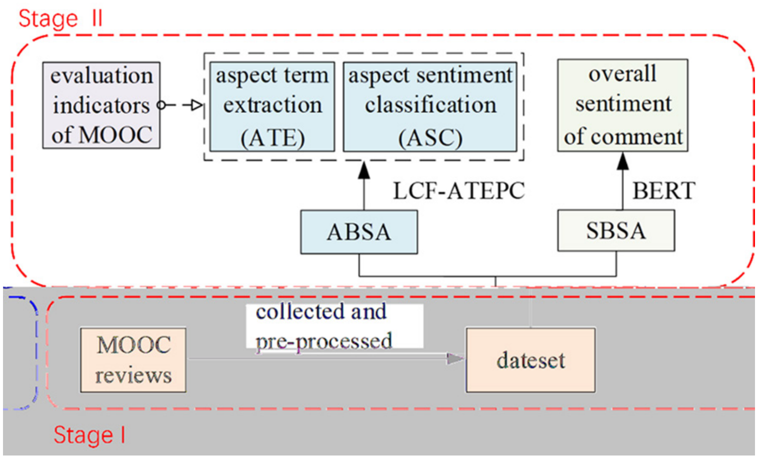

Fig. 1. The framework of the online course evaluation

The framework for the online course evaluation method based on sentiment analysis is illustrated in Figure 1. This framework comprises two stages; namely, 1) collection of MOOC course review comments using a web crawler and preprocessing of the data for the subsequent stages, and 2) training of sentiment analysis models using deep learning techniques to obtain the SBSA and ABSA results.

This study focuses on obtaining online course evaluation results from three perspectives; namely, overall course satisfaction, aspect-based attention and satisfaction, and specific course characteristics. By drawing from existing research on online course evaluation and considering the learners' needs, the main evaluation aspects encompass teachers, curriculum teaching, platform, and learner experience [31, 32].

3.1 Collection of course review comments

Experiments were conducted using a dataset comprising review comments from Chinese university MOOC courses to investigate the effectiveness of our approach. Given the variability of sentiment corpus across different domains, multiple types of course review comments were combined to create a diverse dataset, enabling the trained model to adapt to various course domains. The review comment data encompassed 13 major course categories, including intellectual property, computer science, foreign languages, natural science, engineering, and law. After careful screening to remove invalid data, such as comments that consist solely of symbols or numeric grades without any accompanying text, a dataset of over 210,000 review comments was obtained. The experiments were conducted using PyTorch 1.2.0, implemented in Python 3.6.8 programming language, and developed in a Linux operating system environment using PyCharm.

Dataset description. The length of review comment texts in the dataset was analyzed, and the findings are illustrated in Figure 2. The majority of the comments (approximately 63.1%) contained no more than 15 words, with a significant concentration of texts within the 0- to 200-word range. The most common length of review comments was around 15 words, and these comments consisted of 223 words on average. A considerable proportion of samples (137,479, 63.1%) had comment lengths ranging from 0 to 15 words. Overall, these MOOC review comments tend to be short.

Due to the lack of alignment between the original grades and the sentiment polarity of review comments, a significant majority of the comments (up to 95%) were categorized as "positive" based on the sentiment polarity associated with the original grades ranging from 1 to 5, where scores of 4 and 5 were considered "positive." This categorization led to a substantial imbalance between the proportions of positive and negative review comments, which could introduce bias into the models. Therefore, the sentiment polarity of the review comments needed to be manually reviewed and assessed. However, manually proofreading over 200,000 data entries proved to be a daunting task. To address the issues of data imbalance and the large volume of data, a relatively balanced subset was extracted from the complete dataset for model construction and training, resulting in separate datasets for SBSA and ABSA.

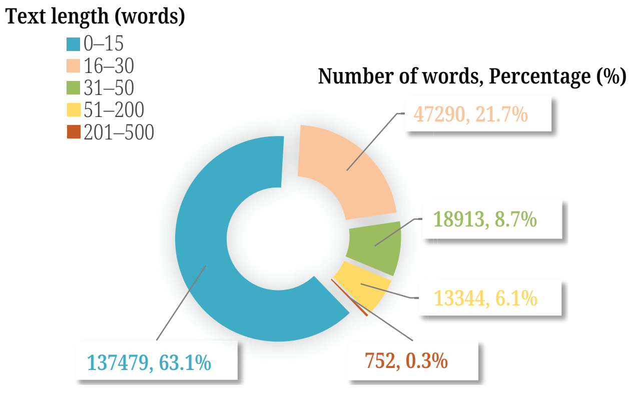

Fig. 2. Pie chart of different review comment text lengths

Dataset for SBSA. An SBSA benchmark was trained after meticulously proofreading 7,815 comments, which comprised 4,000 positive and 3,815 negative samples. The training set encompassed 80% of the data, while the remaining 20% was allocated to the test and validation sets. The trained benchmark was employed to process the complete dataset of 217,778 entries. The resulting SBSA model demonstrated its efficacy in analyzing review comment texts within the education domain. The sentiment polarity of the processed dataset, as determined by the benchmark, is presented in Table 1. By leveraging the benchmark to process over 200,000 entries, the manual workload was successfully alleviated while preserving the integrity of the data characteristics.

Table 1. The sentiment polarity of the entire dataset after being processed by the benchmark

Sentiment Polarity	Train	Test	Val	Total
Negative	9010	500	500	10010
Positive	206768	500	500	207768
Total	215778	1000	1000	217778

Dataset for SBSA. The review comment text was further annotated using the entire dataset. For ABSA, the aspect terms within the sentences were extracted. The sentences were initially segmented using punctuation marks, such as "?", "!", ".", ",", "…", " ", and "、", according to a predefined rule. Subsequently, the aspect terms and their corresponding sentiment polarity were manually annotated as presented in Table 2.

Table 2. The aspect terms and their sentiment polarity after segmentation

Number	Review Comment Data after Processing	Words	Sentiment Polarity
1	The content of the class is abundant	content	Positive
2	Mr. Li's lecture ideas are very clear	lecture ideas	Positive
3	Poor course experience	course experience	Negative
4	The teacher's case is very typical and realistic	case	Positive
5	A little too much lecture	lecture	Negative
6	Video playback failed	video playback	Negative
7	The information is very complete	information	Positive
8	The content is concise and interesting	content	Positive
9	The teaching form is very novel and interesting	teaching form	Positive
10	The scoring standard is unreasonable	scoring standard	Negative
…	…	…	…

The ABSA dataset consisted of 11,969 data points, of which 8,703 were positive and 3,266 were negative. The dataset was constructed by manually annotating the polarity of entities within 4,161 short sentences. The model was trained using this annotated dataset to automatically extract all aspects of information and predict the extreme sentiment polarity. This approach enables the model to be adaptable to large-scale datasets encompassing the aspects and emotions of students' feedback.

3.2 Sentiment analysis models for course review comments

The sentiment analysis models utilized in this study include SBSA, which leverages the BERT model, and ABSA, which employs the LCF-ATEPC model.

SBSA: BERT model. BERT, a pre-training model for text, has demonstrated its effectiveness in text classification [33]. In contrast to conventional convolutional and recurrent neural network architectures, BERT adopts a Transformer structure with a multi-head self-attention mechanism to construct its network model. This design enables an improved contextual comprehension of text and enhances prediction accuracy.

The pre-trained BERT model can be fine-tuned to enhance its performance in natural language processing tasks. In our study, the BERT model is leveraged for feature extraction and sentiment polarity determination.

The BERT model utilizes the encoder component of the Transformer, which is denoted by T_r in Figure 3. The Transformer Compiler, a sequence-to-sequence (Seq2Seq) model, incorporates a self-attention mechanism. The model comprises a stack of multiple encoders and decoders, where the input is processed by the left encoder and the output is generated by the right decoder, as depicted in Figure 4. By employing a multi-head self-attention mechanism, the Transformer captures long-range dependencies and captures additional contextual information. Compared with traditional recurrent neural networks, the parallel training approach of the Transformer enables a more efficient computation by simultaneously training all words.

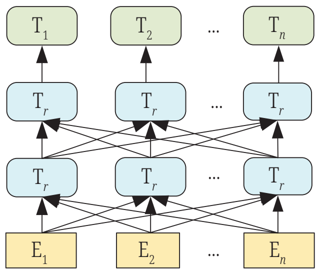

Fig. 3. BERT model

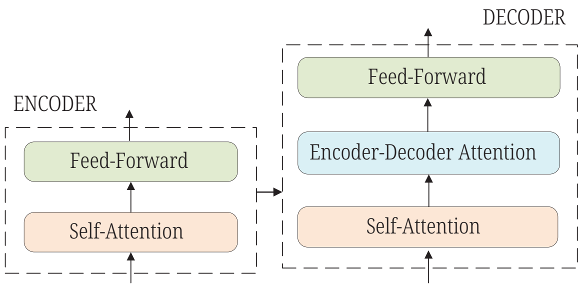

Fig. 4. Encoder–decoder architecture diagram

ABSA: LCF-ATEPC model. ABSA typically encompasses ASC, which is typically considered a distinct task. We employed the Chinese-oriented multi-task learning model LCF-ATEPC proposed by Yang and Zeng [22] for analyzing ATE and ASC within ABSA. This model facilitates a simultaneous extraction of aspect terms and the classification of their sentiment polarity, thereby capturing a wide range of pertinent textual features.

As depicted in Figure 5, the LCF-ATEPC model comprises a local context feature generator (LCFG) on the left and a global context feature generator (GCFG) on the right. The LCFG leverages a BERT model, a multi-head self-attention mechanism, and context dynamic masking (CDM) or context dynamic weighting (CDW) techniques to capture comprehensive feature information. The BERT model is utilized for primary feature extraction, and its multi-head self-attention mechanism is leveraged to address the limitations of long-distance context dependencies and obtain a comprehensive set of features. Specifically, the BERT layer learns non-local context features, which can be further enhanced through CDM, CDW, or weighted masking to acquire highly precise feature information. Meanwhile, the GCFG primarily applies the BERT model to obtain global features. The aspect extractor (AE) employs the global context feature generator to extract aspect terms, while the polarity extractor (PE) combines the information obtained from both LCFG and GCFG to determine the sentiment polarity judgment outcome. This model is capable of adapting to large-scale datasets, thereby reducing the manual annotation workload.

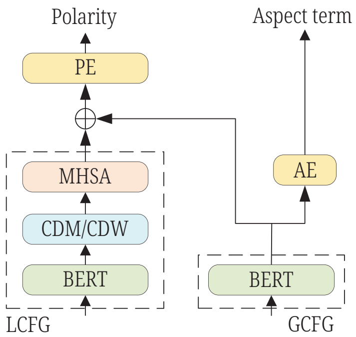

Fig. 5. LCF-ATEPC model

4 RESULTS AND DISCUSSION

4.1 Evaluation index

The effectiveness of the sentiment analysis model was assessed using four evaluation metrics derived from the confusion matrix; namely, *Accuracy*, *Precision*, *Recall*, and *F-measure*. These metrics were employed to analyze and interpret the classification results obtained from the model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F\text{-}measure = \frac{2 \cdot P \cdot R}{P + R} \tag{4}$$

Accuracy is utilized to evaluate the classifier's performance by measuring the proportion of correctly classified samples out of the total number of samples in a given test dataset. In this study, *Accuracy* is employed as an evaluation criterion to assess the model's efficacy in determining the aspect sentiment polarity. This metric represents the ratio of review comments with the correct sentiment polarity determined by the model to the total number of review comments within the sample.

Precision quantifies the proportion of true positive predictions and represents the review comments that are correctly identified as having a positive polarity among all review comments predicted as positive by the model. This metric is computed as the ratio of true positives (*TP*) to the sum of true positives (*TP*) and false positives (*FP*). *Precision* serves as an indicator of the model's accuracy in summarizing aspects and sentiments. In this context, *TP* denotes the number of negative samples correctly predicted as negative, *TN* denotes to the number of negative samples correctly predicted as negative, *FP* refers to the number of negative samples incorrectly predicted as positive, and *FN* represents the number of positive samples incorrectly predicted as negative.

P represents *Precision*, and *R* represents *Recall*. Both of these metrics are often inversely proportional, meaning that an increase in one metric may result in a decrease in the other metric. A high *Precision* indicates that the model accurately identifies a significant proportion of positive samples but may overlook some positive samples, resulting in a low *Recall*. Conversely, a high *Recall* suggests that the model identifies a substantial number of positive samples but may mistakenly classify some negative samples as positive, leading to a low *Precision*. To strike a balance between *Precision* and *Recall*, the *F-measure* is employed as an evaluation metric. The *F-measure* is calculated as the harmonic mean of *Precision* and *Recall*, and a high *F1* value signifies the superior performance of the model.

4.2 Model validation

Table 3 presents the *Accuracy*, *Precision*, *Recall*, and *F1* values of the BERT and LCF-ATEPC models on the dataset. The BERT model achieves an accuracy of 93.0% on the dataset, with the aforementioned metrics for ASC and ATE in ABSA exceeding 90%. These results prove the strong performance of the LCF-ATEPC model in text sentiment analysis.

Table 3. Performance metrics for BERT and LCF-ATEPC models

Model	Evaluation Indicators (%)			
	Accuracy	Precision	Recall	F-Measure
BERT	93	88.2	99.2	93.4
LCF-ATEPC (ASC)	91.5	90	93.1	91.5
LCF-ATEPC (ATE)	96.9	91.8	94.9	93.3

4.3 Cases in MOOC

This section presents the application of the proposed framework in evaluating online courses in Chinese universities. To illustrate its effectiveness, the course was taken as a case study to demonstrate how the outcomes of SBSA and ABSA can be leveraged to evaluate both the overall satisfaction with the course and its specific aspects.

Course description. Review comment data were gathered from five distinct courses offered on the MOOC platform of Chinese universities. These courses spanned various disciplines, including computer science, foreign languages, natural science, law, and education and teaching. These courses were not part of the original research dataset and were deliberately selected from diverse categories to demonstrate the model's generalizability and capability to analyze course review comments from various categories. Table 4 provides an overview of the selected courses and their key characteristics. All courses were highly regarded, with enrollment numbers ranging from 6 to 11 and total student counts placing them among the top three courses in their respective categories. The student enrollment for these courses varied from 3,000 to 300,000, indicating a diverse and substantial dataset.

Table 4. Course information in use case

Course No.	Category	Title	Starting Unit	Participants	Classes	Comments
A	Computer science	Artificial intelligence and information society	Peking University	104788	7	814
B	natural science	Advanced mathematics	Xi'an Jiaotong University	304054	6	605
C	foreign language	English listening skills and practice	JiMei University	158309	7	817
D	education	Teaching application of mind map	Love Course	167709	11	3085
E	law	Intellectual property law	Zhongnan University	3762	11	467

SBSA. After applying SBSA to the review comments from the five selected courses, Figure 6 displays the overall polarity of sentiment polarity across these courses. The figure illustrates the original sentiment polarity of the review comments as identified by SBSA. Overall, these courses received predominantly positive review comments from students. However, the SBSA results reveal a lower proportion of positive sentiment and a higher proportion of negative sentiment compared with the learners' direct grading. This discrepancy may be attributed to the learners' grading tendencies because they often adhere to a binary approach of "high or low" when providing comments and assigning grades. Consequently, negative emotions are expressed by some students in their review comments despite the relatively high final scores.

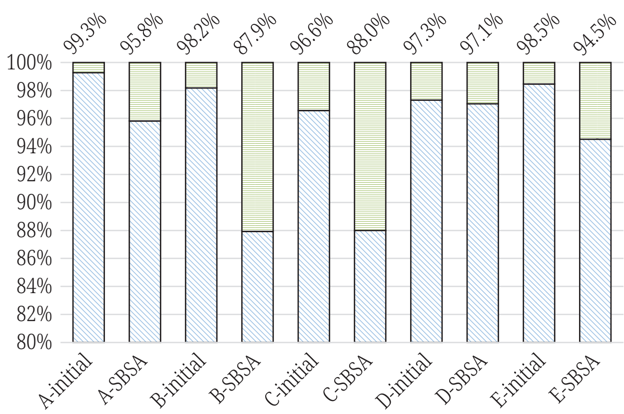

Fig. 6. Overall sentiment polarity of the courses

In practical applications, the ranking results obtained by SBSA, which may differ from the original grades, can offer new perspectives for learners in their course selection. As depicted in Figure 6, after applying SBSA, the rankings of courses A and B are lower compared with that of course D, which indicate that the learner satisfaction for course D surpasses that of other courses, thus suggesting that SBSA can provide a highly precise and comprehensive evaluation of learner sentiment toward a particular course. These insights can aid learners in making informed decisions when selecting courses.

The prevalence of positive polarity comments from past learners may diminish in relation to the course's level of popularity. Specifically, courses B (advanced mathematics) and C (English listening skills and practice) exhibit a greater reduction in positive polarity following the application of SBSA compared with the three other courses. This trend could be attributed to the tendency of learners in popular courses to provide highly detailed and specific review comments. Similarly, course E (intellectual property law), which attracts a professional student cohort, receives a comparatively smaller number of targeted and specific review comments compared with the other courses.

ABSA. ABSA was utilized to analyze the review comments provided by the learners, aiming to identify aspect terms relevant to the course. These terms were then categorized into four aspects; namely, teachers, curriculum teaching, platform guarantee, and learning effect and learner experience, based on the online course evaluation system, which reflects the learners' requirements. Through the systematic arrangement of these aspect terms alongside their corresponding sentiment scores, valuable insights were gained regarding the learners' preferences toward different aspects of the course and their overall feedback attitudes. This information is shown in Figures 7 and 8. By examining the collected data, a comprehensive list of features related to the online course pertaining to each of the four aspects was compiled and presented in Table 5.

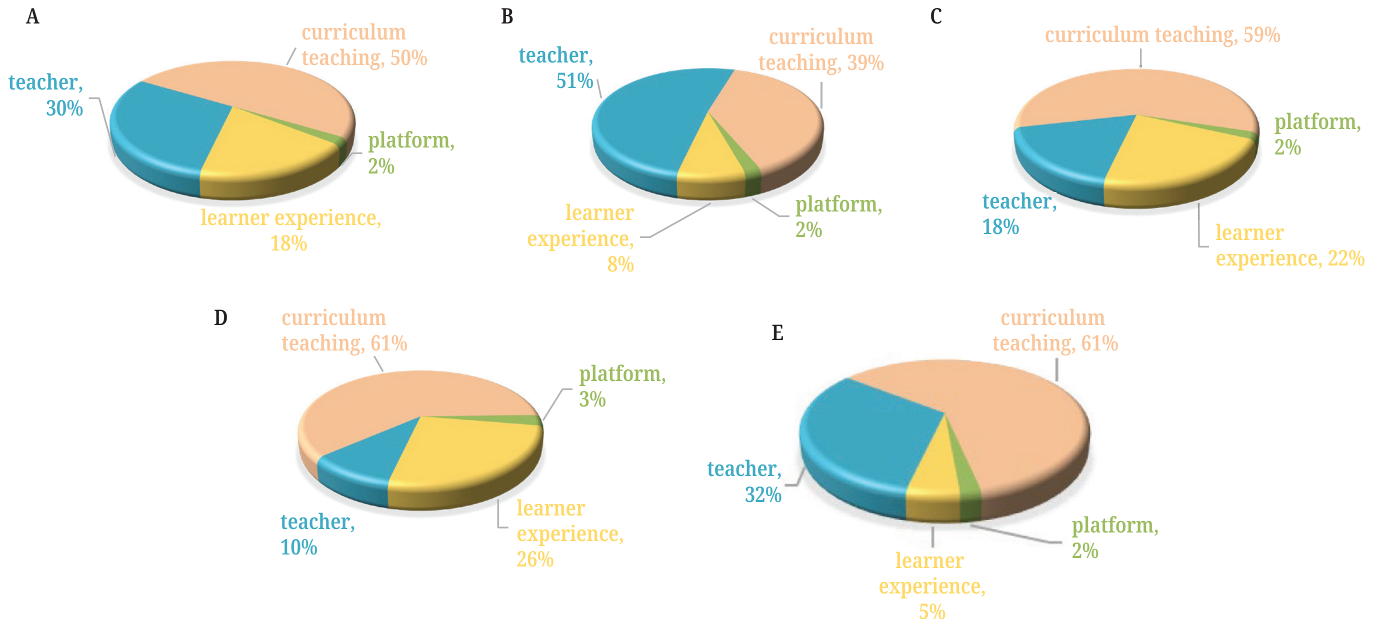

Fig. 7. Size of review comments on each aspect

ATE of ABSA. In general, learners prominently focus on the course teaching and teacher aspects, as shown in Figure 7. These aspects receive the highest degree of attention among learners, followed by learner experience and platform aspects. Among the five courses analyzed, course D attracts considerable interest in course teaching, accounting for 61% of the total attention. Conversely, learners enrolled in course B demonstrate a heightened interest in the teacher aspect, implying a high inclination to provide feedback or suggestions to teachers. The aspects of learning effect and learner experience receive moderate attention across all five courses, with course D showing the highest proportion. Meanwhile, the platform aspect receives relatively less attention and suggestions compared with the other aspects, thereby suggesting a stronger emphasis on the quality of the course content than the platform itself.

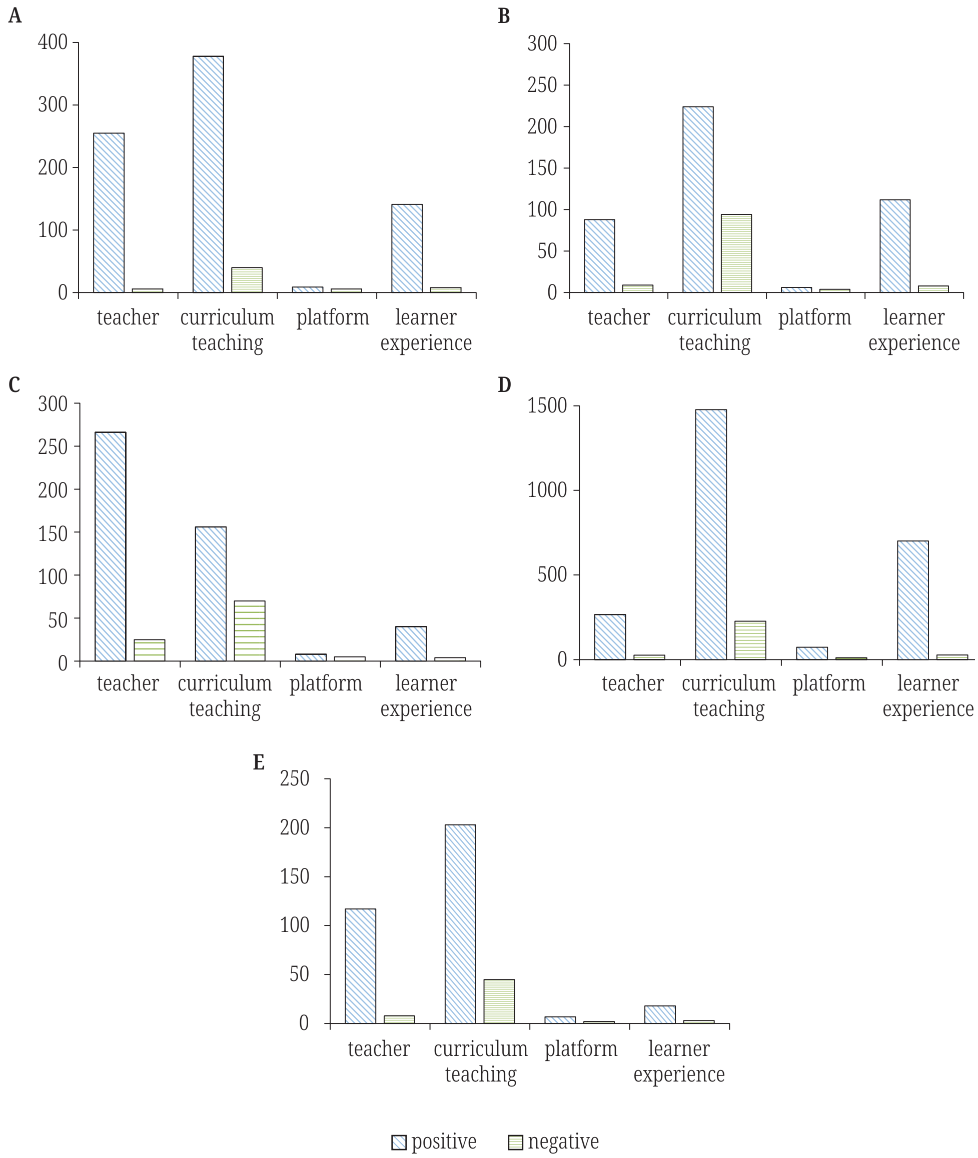

Fig. 8. Sentiment polarity at all aspects for courses A–E

ASC of ABSA. The polarity of different aspects can be analyzed by counting the occurrence of the aspect terms. As shown in Figure 8, learners exhibit a predominantly positive attitude toward teachers, course teaching, learning effects, and learners' experience. However, a relatively high proportion of negative feedback is observed for teaching in courses B and C, indicating that students may have concerns regarding course organization, schedule, and teaching methods.

The aspects that receive the highest levels of satisfaction and dissatisfaction from learners generally align in courses A, B, D, and E. However, this pattern is inconsistent. For instance, in course C, the teacher aspect receives notably more positive feedback compared with the other aspects despite not having the highest degree of negative feedback. Learners who provide review comments exhibit two distinct types of behavior. Some learners start by expressing positive remarks about the aspects they find satisfactory before mentioning areas for improvement, while others directly express their positive opinions about satisfactory aspects and negative sentiments toward unsatisfactory aspects. Upon analyzing the review comments for the five courses, the former behavioral pattern appears more prevalent than the latter. This finding is further supported by the ASC results, which revealed that certain positive and negative feedbacks were associated with the same aspect terms.

Adjustments were made to the positive polarity feedback on all courses. The final polarity tendency for each course aspect was determined by subtracting the number of negative emotion review comments from the number of positive emotion review comments. For example, in course C, the corrected positive polarity distribution value of the curriculum teaching aspect was obtained by subtracting the number of negative review comments from the number of positive review comments. By integrating the findings from Figures 7 and 8, we can infer the learners' concerns, satisfaction, and dissatisfaction with all aspects of the course. This information can be valuable in summarizing the course characteristics and formulating improvement strategies. In course B, the adjusted polarity distribution results reveal that learners are equally satisfied with teachers, curriculum teaching, and learners' experience but are highly dissatisfied with teaching. The flaws in the course's curriculum teaching, which were not apparent in the original analysis results, can now be identified by those involved in course design and development.

Course characteristics from aspect terms. The evaluation process for an online course based on the outcomes of SBSA and ABSA can be delineated into the following steps. First, the overall satisfaction with the course is evaluated by taking into account both the positive and negative aspects identified in the course review comments through SBSA and ABSA. Second, the course characteristics are elucidated by describing the aspect terms extracted by the model.

A sentiment analysis was conducted in this study to evaluate the online course on intellectual property law. The analysis of the learners' review comments reveals an overall satisfaction rate ranging from 90% to 95%, with the majority of the students showing their highest interest in the curriculum teaching and teacher aspects, followed by learner experience. Meanwhile, the platform aspect received the least attention. The sentiment polarity distribution statistics also reveal the high level of satisfaction of existing learners with the course's teachers and curriculum teaching. The aspect terms of the course review comments, along with their frequencies, are presented in Table 5.

Table 5. Aspect terms of the intellectual property law course

Polarity	Aspects	Aspect Terms and Frequencies
positive	teachers	teacher (32), speaking (27), explaining (28), lecturing (11), teacher lecturing (4), giving lessons (4), giving lessons (4), telling (2), teacher speaking (2), telling (2), professor (1), faculty (1), teacher giving lessons (1), teacher speaking (1), speaking speed (1), guiding (1), teacher explaining (1)
	curriculum teaching	content (67), course (45), lesson (41), knowledge (24), experienced (7), case (7), material (5), video (2), plate (2), material (1), teaching (1), law (1), teaching (1), answer (1), exercise (1), resource (1), classroom (1), courseware (1), example (1), teaching (1), method (1), difficulty (1), example (1), material (1)
	platform	MOOC (5), online class (2)
	learner experience	learn (10), help (4), gain (2), way (2), useful (1)
negative	teachers	teacher (32), speaking (5), explain (3), interacting (1), make a speech (1)
	curriculum teaching	content (12), lesson (8), course (5), problem (3), case (2), question (2), material (2), old (2), document (1), courseware (1), starting course (1), book (1)
	platform	MOOC (1), platform (1)
	learner experience	meaning (1), feeling (1)

Table 5 provides an overview of the aspect terms related to the intellectual property law course along with their corresponding sentiment polarities. The aspects covered in the table include teachers, curriculum teaching, platform, and learner experience. The frequency of each aspect term is enclosed in parentheses on the right. The sentiment polarity of these aspects can be either positive or negative. For instance, the most frequently mentioned words under the teachers aspect include "teacher," "speaking," "explaining," and "lecturing." Similarly, the most commonly mentioned words under the curriculum teaching aspect are "content," "course," "lesson," and "knowledge." The platform aspect has only a few mentions, and its sentiment polarity is neutral.

Table 5 shows that learners express high levels of satisfaction with various aspects of their teacher's performance, including his/her explanations, teaching methods, teaching style, expression, and speaking speed. In other words, the course teacher delivers engaging lectures and presents the complex subject matter of intellectual property law in a comprehensible manner. The negative review comments about the teacher accounted for 18% of all negative feedback, with most of them focusing on the teacher's explanations and interaction. The frequency ratio of the aspect terms related to the teacher's explanations indicates a satisfactory-to-unsatisfactory ratio of 7:3, thereby suggesting an appropriate level of course difficulty that caters to both beginners and interested learners. However, compared with offline courses, the level of interaction between teachers and learners in online courses is limited, which may pose some inconvenience for learners.

Based on their review comments, the learners are overall satisfied with the course design, course materials, and course difficulty. This finding is supported by the frequent occurrence of certain words, such as "course," "lesson," "knowledge," "case," "material," "video," and "plate." As a well-established course, intellectual property law is likely to have amassed extensive teaching experience, comprehensive content, in-depth knowledge points, relevant case studies, and supplementary materials.

However, the analysis also reveals that 78% of the unsatisfactory review comments are associated with the curriculum teaching aspect, with learners expressing their dissatisfaction with the outdated course content, cases, and materials. Students should complement their learning by taking other courses that offer current and up-to-date knowledge points.

The findings and interpretations of this study offer valuable insights for curriculum developers aiming to enhance the quality of the intellectual property law course. The identified areas for improvement can be prioritized based on the attention and satisfaction levels of the learners. The level of attention and satisfaction for each aspect remains consistent across the course review comments. Therefore, course developers may consider enhancing their online communication with learners to further improve teaching effectiveness. These recommendations can also guide curriculum developers in making informed decisions to elevate the overall learning experience of students.

5 CONCLUSION

This study develops several online course evaluation methods tailored to learners' requirements by applying SBSA and ABSA on MOOC review comments. An LCF-ATEPC model is trained to perform sentiment analysis on these review comments. As its primary contribution, this study provides learners with comprehensive course evaluation results that encompass teachers, courses, platforms, and learning effectiveness based on a large corpus of course review comments. The recommendations derived from this evaluation process can significantly enhance the effectiveness of course selection. Insights into the overall satisfaction of past students with any online course can be gained by learners through the proposed evaluation method along with their satisfaction with teachers, course schedules, and learning experiences. The evaluation results also offer valuable insights into various aspects of the course, such as teacher suitability, course difficulty, and learning effectiveness. Course developers may use these findings to formulate targeted improvement strategies for their curriculum and teaching quality.

The proposed course evaluation method may yield lower overall course satisfaction grades compared with direct manual grading. This method may also produce different course rankings, as evidenced in the illustrated cases. The utilization of SBSA and ABSA techniques in the evaluation process facilitates the effective extraction of course characteristics and difficulty levels.

Future research may enhance the proposed model to recognize emotions other than negative and positive ones, including neutral emotions, to address the limitation of SBSA in identifying emotions. Furthermore, given the suboptimal accuracy of ABSA in extracting aspect terms, modifications should be made to the model to improve its aspect term extraction precision and entity recognition efficiency.

6 ACKNOWLEDGMENT

This study was supported by the University-Industry Cooperation of Fujian Province under Grant (NO. 2021H6026); the Intelligent Computing and Application Research Team of Concord University College (NO. 2020-TD-001); Fujian Provincial Social Science Fund Project (NO. FJ2020C046); Information Project for Middle and Young People in Fujian Province's Universities (NO. JZ180200); Innovative Project of

Fujian Agriculture and Forestry University (NO. CXZX2021035); Scientific Research Project for Outstanding Young Scholars of Fujian Agriculture and Forestry University (NO. XJQ2021S3).

7 REFERENCES

[1] Abdullah, N. A. (2020). Evaluating pre-service teaching practice for online and distance education students in Pakistan. International Review of Research in Open and Distance Learning, 21(2): 81–97. https://doi.org/10.19173/irrodl.v21i2.4606

[2] Levitskyi, A., Benzar, I., Vitiaz, V., Golovatiuk, D., Karabenyuk, O. (2021). Teaching pediatric surgery in the online environment: challenges and opportunities. International Journal of Online and Biomedical Engineering, 17(5): 119–132. https://doi.org/10.3991/ijoe.v17i05.22051

[3] Du, Z., Wang, F., Wang, S., Xiao, X. (2022). Enhancing learner participation in online discussion forums in massive open online courses: the role of mandatory participation. Frontiers in Psychology, 13: 819640. https://doi.org/10.3389/fpsyg.2022.819640

[4] Wang, S., Song, F., Liu, Y., Song F., Xie, X., Dong Y. (2021). Research on evaluation system of user experience with online live course platform. IEEE Access, 9: 23863–23875. https://doi.org/10.1109/ACCESS.2021.3054047

[5] Li, H. J., Peng, M. (2019). Online course learning outcome evaluation method based on big data analysis. International Journal of Continuing Engineering Education and Lifelong Learning, 29(4): 349–361. https://doi.org/10.1504/IJCEELL.2019.10023378

[6] Dina, N. Z., Yunardi, R. T., Firdaus, A. A. (2021). Utilizing text mining and feature-sentiment-pairs to support data-driven design automation massive open online course. International Journal of Emerging Technologies in Learning, 16(1): 134–151. https://doi.org/10.3991/ijet.v16i01.17095

[7] Wang, C., Huang, S., Zhou, Y. (2021). Sentiment analysis of MOOC reviews via ALBERT-BiLSTM model. MATEC Web of Conferences, 336: 05008. https://doi.org/10.1051/matecconf/202133605008

[8] El-Demerdash, A. A., Hussein, S. E., Zaki, J. F. (2022). Course evaluation based on deep learning and SSA hyperparameters optimization. Computers, Materials & Continua, 71(1): 941–959. https://doi.org/10.32604/cmc.2022.021839

[9] Qi, C., Liu, S. (2021). Evaluating on-line courses via reviews mining. IEEE Access, 9: 35439–35451. https://doi.org/10.1109/ACCESS.2021.3062052

[10] Fan, S., Yao, J., Sun, Y., Zhan Y. (2020). A summary of aspect-based sentiment analysis. Journal of Physics: Conference Series, 1624(2): 022051. https://doi.org/10.1088/1742-6596/1624/2/022051

[11] Sudhakar, B., Bensraj, R. (2014). An efficient sentence-based sentiment analysis for expressive text-to-speech using Fuzzy Neural Network. Research Journal of Applied Sciences Engineering & Technology, 8(3): 378–386. https://doi.org/10.19026/rjaset.8.983

[12] Hayashi, T., Fujita, H. (2019). Word embeddings-based sentence-level sentiment analysis considering word importance. Acta Polytechnica Hungarica, 16(7): 7–24. https://doi.org/10.12700/APH.16.7.2019.7.1

[13] Fink, C. R., Chou, D. S., Kopecky, J. J., Llorens, A. J. (2011). Coarse- and fine-grained sentiment analysis of social media text. Johns Hopkins Apl Technical Digest, 30(1): 22–30. https://doi.org/10.1016/j.isatra.2010.08.002

[14] Liang, B., Su, H., Gui, L., Cambria, E., Xu, R. (2022). Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks. Knowledge-Based Systems, 235: 107643. https://doi.org/10.1016/j.knosys.2021.107643

[15] Wei, Y., Zhang, H., Fang, J., Wen, J., Ma, J., Zhang, G. (2021). Joint aspect terms extraction and aspect categories detection via multi-task learning. Expert Systems with Applications, 174: 114688. https://doi.org/10.1016/j.eswa.2021.114688

[16] Thin, D. V., Nguyen, N. L., Truong, T. M., Le, L. S., Vo, D. T. (2021). Two new large corpora for Vietnamese aspect-based sentiment analysis at sentence level. ACM Transactions on Asian and Low-Resource Language Information Processing, 20(4): 1–22. https://doi.org/10.1145/3446678

[17] Depeursinge, A., Al-Kadi, O. S., Mitchell, J. R. (2017). Biomedical texture analysis: Fundamentals, tools and challenges. Academic Press.

[18] Piryani, R., Madhavi, D., Singh, V. K. (2016). Analytical mapping of opinion mining and sentiment analysis research during 2000–2015. Information Processing & Management, 53(1): 122–150. https://doi.org/10.1016/j.ipm.2016.07.001

[19] Osmanolu, U. M., Atak, O. N., Caglar, K., Kayhan, H., Can, T. (2020). Sentiment analysis for distance education course materials: a machine learning approach. Journal of Educational Technology and Online Learning, 3(1): 31–48. https://doi.org/10.31681/jetol.663733

[20] Sudhir, G. I., Dang, P. M., Sanjay, A. S., Kantrao, M. R. (2018). A survey on product review sentiment analysis. International Journal of Advance Research and Innovative Ideas in Education, 4(2): 1848–1852.

[21] Onan, A. (2021). Sentiment analysis on massive open online course evaluations: a text mining and deep learning approach. Computer Applications in Engineering Education, 29(3): 572–589. https://doi.org/10.1002/cae.22253

[22] Yang, H., Zeng, B., Yang, J., Song, Y., Xu, R. (2021). A multi-task learning model for Chinese-oriented aspect polarity classification and aspect term extraction. Neurocomputing, 419: 344–356. https://doi.org/10.1016/j.neucom.2020.08.001

[23] Gomez-Rey, P., Fernandez-Navarro, F., Barbera, E., Carbonero-Ruz, M. (2018). Understanding student evaluations of teaching in online learning. Assessment & Evaluation in Higher Education, 43(8): 1272–1285. https://doi.org/10.1080/02602938.2018.1451483

[24] Sood, C., Calderon, O. (2020). Evaluating learning outcomes of an asynchronous online discussion assignment: a post-priori content analysis. Interactive Learning Environments, 28(1): 3–17. https://doi.org/10.1080/10494820.2018.1510421

[25] Barteit, S., Guzek, D., Jahn, A., Bärnighausen, T., Jorge, M. M., Neuhann, F. (2020). Evaluation of e-learning for medical education in low- and middle-income countries: a systematic review. Computers & Education, 145: 103726. https://doi.org/10.1016/j.compedu.2019.103726

[26] Vaz-Fernandes, P., Caeiro, S. (2019). Students' perceptions of a food safety and quality e-learning course: a CASE study for a MSC in food consumption. International Journal of Educational Technology in Higher Education, 16: 37. https://doi.org/10.1186/s41239-019-0168-8

[27] Barros, J. P. (2014). On the description of individual course units for programme assessment and better syllabuses. International Journal of Engineering Pedagogy, 4(5): 28–33. https://doi.org/10.3991/ijep.v4i5.3536

[28] Ardiasih, L. S., Emzir, E., Rasyid, Y. (2018). Implementing constructivism approach in online writing course. English Review Journal of English Education, 6(2): 127–138. https://doi.org/10.25134/erjee.v6i2.1261

[29] Fan, J., Jiang, Y., Liu, Y., Zhou, Y. (2022). Interpretable MOOC recommendation: a multi-attention network for personalized learning behavior analysis. Internet Research: Electronic Networking Applications and Policy, 32(2): 588–605. https://doi.org/10.1108/INTR-08-2020-0477

[30] Douglas, K. A., Bermel, P., Alam, M. M. (2016). Big data characterization of learner behaviour in a highly technical MOOC engineering course. Journal of Learning Analytics, 3(3): 170–192. https://doi.org/10.18608/jla.2016.33.9

[31] Tran, T., Ma, Q. (2021). Using formative assessment in a blended EFL listening course: student perceptions of effectiveness and challenges. International Journal of Computer-Assisted Language Learning and Teaching, 11(3): 17–38. https://doi.org/10.4018/IJCALLT.2021070102

[32] Rodríguez, S., González-Gómez, F., Guardiola, J. (2019). Do course evaluation systems have an influence on e-learning student satisfaction? Higher Education Evaluation and Development, 13(1), 18–32. https://doi.org/10.1108/HEED-09-2018-0022

[33] Su, P., Vijay-Shanker, K. Investigation of improving the pre-training and fine-tuning of BERT model for biomedical relation extraction. BMC Bioinformatics, 23(1), 120. https://doi.org/10.21203/rs.3.rs-640112/v1

8 AUTHORS

Jinjing Zeng, PhD degree, is a librarian at Fujian Agriculture and Forestry University. Her research interests focus on intellectual property and information resource management (email: carrajue@163.com).

Kailing Luo, Master's degree, in College of Education, Fujian Normal University, is a teacher at Fuzhou Gulou Experimental Primary School. Her research interests focus on educational technology (wmail: 525101448@qq.com).

Yu Lu, Master's degree, is a professor at Concord University College, Fujian Normal University. His research interests focus on computational intelligence and educational technology (email: media@fjnu.edu.cn).

Mingfen Wang, Master's degree, is an associate professor at the Concord University College, Fujian Normal University. Her research interests focus on educational technology (email: wmf_0318@163.com).