Trends in Social Sciences and Humanities Research
ISSN: 2959-9938
DOI: https://doi.org/10.61784/tsshr3083

TYPICAL CASES OF ONLINE TEACHING QUALITY EVALUATION BASED ON MULTIMODAL AFFECTIVE STATE ANALYSIS

Jin Lu
Guangdong Key Laboratory of Big Data Intelligence for Vocational Education, Shenzhen Polytechnic University, Shenzhen 518000, Guangdong, China.
Corresponding Email: lujin0808@szpu.edu.cn

Abstract: With the popularity of online education, how to evaluate online teaching quality scientifically and comprehensively has become an urgent problem. Traditional teaching quality evaluation methods often rely on single text data (e.g., student feedback, teacher self-assessment, etc.), which has shortcomings such as strong subjectivity and incomplete information. The multimodal affective state analysis technology can capture multiple affective states of students in the learning process (e.g., facial expression, voice tone, body posture, etc.), thus providing a more comprehensive and objective basis for teaching quality evaluation. This paper proposes an intelligent modern quality evaluation scheme, using multimodal machine learning technology, integrating multidimensional information to comprehensively assess students, and realizing the organic combination of process evaluation, comprehensive ability evaluation and dynamic evaluation. The scheme proposed in this paper can achieve intelligent online teaching evaluation and establish an accurate portrait of teachers and students' learning, and gradually realize trace-free and accompanying teaching evaluation.The experimental results show that the multimodal emotion recognition method for online learning using fused video semantic information in this paper is able to increase the accuracy of emotion recognition by 6% in practical applications. This indicates that the method has great potential in online teaching and can provide teachers with more accurate feedback on students' affective states, so as to better adjust teaching strategies and improve teaching effectiveness.
Keywords: Multimodal emotional analysis; Online teaching quality evaluation; Adaptive teaching system; Video semantic information

1 INTRODUCTION

With the rapid development of information technology, the online teaching mode is becoming increasingly popular. The online education system breaks through the geographical and time constraints of traditional education, allowing learners to study anytime, anywhere, with a rich resource base that meets personalized learning needs, while also providing more opportunities for interaction. However, online teaching also faces some challenges, such as the high self-regulation requirements of online learning and the uneven quality of resources[1]. Therefore, scientific and effective evaluation of online teaching quality is crucial. As an emerging method, multimodal sentiment analysis provides new ideas for online teaching quality evaluation. Traditional sentiment analysis methods mainly focus on the sentiment information in text or speech, while ignoring other modalities. Multimodal sentiment analysis fuses multiple modal information, such as text, voice, image, video, etc., which can provide a more comprehensive understanding of human emotions. In online teaching, through multimodal sentiment analysis, students' various performances in the learning process can be considered comprehensively, such as the emotional expression in text communication, the emotional state in voice feedback, and the change of expression in video interaction, etc., so as to evaluate the teaching quality more accurately. According to statistics, the current online education market is expanding and the number of users continues to grow. However, the problem of uneven quality of online teaching is also increasingly prominent[2-3]. The introduction of multimodal sentiment analysis technology is expected to provide strong support to solve this problem. By analyzing students multimodal affective states, we can understand students learning situation and needs in a timely manner, which provides a basis for teachers to adjust their teaching strategies, thus improving the quality of online teaching[4].

The structure of this paper is as follows. Chapter 1 firstly describes the background and purpose of the study, pointing out that multimodal sentiment analysis provides new ideas for online teaching quality evaluation, aiming to improve the accuracy and comprehensiveness of evaluation. Chapter 2 introduces the theoretical basis of multimodal sentiment analysis, including the concepts and principles, data fusion technology, sentiment feature extraction technology, and the theoretical basis of online teaching quality evaluation, such as the principle of constructing the evaluation index system and the diversified selection of evaluation methods. Chapter 3 demonstrates the role of multimodal sentiment analysis in improving the accuracy of sentiment recognition, learners' emotional state and teaching effect through the application cases of multimodal sentiment analysis in online teaching, such as adaptive teaching system based on multimodal sentiment recognition and online learning sentiment recognition by fusing video semantic information. Chapter 4 proposes specific strategies for multimodal sentiment analysis for online teaching quality evaluation, including constructing multimodal sentiment indicators and analyzing the association between sentiment state and teaching

© By the Author(s) 2024, under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0).

quality. Finally, the research conclusions are summarized, pointing out the application value and deficiencies of multimodal sentiment analysis in online teaching quality evaluation, and looking forward to future research directions, including technological innovation, data collection and processing, and application expansion.

2 BASIC THEORY

2.1 Concepts and Principles of Multimodal Sentiment Analysis

Multimodal Sentiment Analysis is a method that integrates multiple modal information to analyse and understand human emotions. It does this by fusing data from different modalities such as text, speech, and images to obtain more comprehensive and accurate sentiment information. Multimodal data fusion technology aims to effectively fuse data information from different modalities to improve the accuracy and efficiency of target tracking and analysis. Currently, there are three main fusion methods for multimodal data fusion: front-end fusion, back-end fusion and intermediate fusion. The details are shown in Table 1.

Table 1 Classification of Multimodal Data Fusion Techniques

Integration Approach	Mode of Operation
Front-end Fusion[5]	By fusing multiple independent datasets into a single feature vector, which is then fed into a machine learning classifier. However, this approach often fails to take full advantage of the compartmentalizes between multiple modal data and the raw data usually contains a lot of redundant information. For example, methods such as Principal Component Analysis (PCA), Maximum Relevance Minimum Redundancy (mRMR) algorithm, and Auto-encoders are often combined with front-end fusion to remove redundant information
Back-end Fusion[6]	By transforming different modal data into high-dimensional feature representations first, and then fusing them in the intermediate layer of the model. A major advantage of the intermediate fusion method is the flexibility of choosing the fusion location
Intermediate Fusion[7]	Fusion is performed by scoring the outputs of classifiers trained separately for different modal data. The advantage of this approach is that the errors of the fused models come from different classifiers, and the errors from different classifiers tend to be uncorrelated and unaffected by each other, without causing further accumulation of errors. Common back-end fusion methods include maximum fusion, mean fusion, Bayes rule fusion, and integrated learning. Among them, integrated learning, as a typical representative of back-end fusion methods, is widely used in communication, computer recognition, speech recognition and other research fields

The study of online teaching quality evaluation method based on multimodal affective state analysis is an interdisciplinary research method that combines knowledge from the fields of psychology, education and computer science. This method aims to assess students' affective states by analyzing their multiple modalities (e.g. speech, facial expressions, body language, etc.) and evaluate the quality of online teaching accordingly.Sentiment feature extraction is one of the key aspects of multimodal sentiment analysis. Common sentiment feature extraction methods are shown in Table 2.

Table 2 Summary of Emotion Feature Extraction Techniques

Data Type	Extraction Method	Application Channel
Text Data[8]	The common methods used for sentiment feature extraction from text data are bag-of-words model, TF-IDF (Term Frequency-Inverse Document Frequency), and word embedding. Bag-of-words model takes each word in the text as a feature and counts the word frequency; TF-IDF combines the word frequency with penalties of other words in the document to consider the importance of the word. Word embedding use neural networks to learn semantic relationships between words, generating continuous vector representations	Emotional analysis of students' language in texts such as discussion boards and homework feedback to determine their emotional attitudes.
Speech Data[9]	For speech data, per-emphasis is applied to the speech sequence and then a discrete FFT transform is performed, and the MFCC parameters are calculated step by step. Specific steps include applying per-emphasis to the speech sequence and then doing the discrete FFT transform; sampling a frame in the speech, adding the Hamming window and then doing the FFT of M points to get the discrete power spectrum. Calculate the power value, natural logarithm, discrete cosine transform, etc. Finally, remove the DC component and take the specific parameters as MFCC parameters	Analyzing students' voice intonation characteristics such as rate of speech, volume, and pitch changes to determine their affective tendencies, e.g. excitement, frustration, or preoccupation.

Image Data[10]	The extraction of emotional features from image data is relatively complex, and emotional features in images can be extracted by deep learning models such as conventional neural networks, for example, capturing emotion-related information such as expressions and colour in images.	Students' facial expressions are captured by the camera, and computer vision techniques are used to recognize different types of expressions, such as happy, surprised, confused, etc., and then infer the students' emotional state.

2.2 Theoretical Basis of Online Teaching Quality Evaluation

The construction of the evaluation index system should follow the principles of scientifically and repeatability, which have important applications in multimodal evaluation. The principle of scientifically requires that the evaluation indexes can objectively and accurately reflect the quality of online teaching and the emotional state of students. Multimodal sentiment analysis can comprehensively assess students' learning process from different perspectives by integrating text, voice, image and other modal information, which provides strong support for the principle of scientifically[11-12]. For example, in text communication, students' questioning and discussion content can be analyzed to determine their understanding of knowledge and learning attitude. In voice feedback, students' emotional state and engagement are understood by analyzing features such as intonation and speed of speech. In video interaction, observe the change of students' expression to infer their learning interest and concentration. The principle of portability emphasizes that evaluation indicators are easy to measure and access. Multimodal sentiment analysis technology can use existing sensors and software tools to collect and analyse students' multimodal information. For example, voice data can be captured through microphones, image information can be captured using cameras, and text analysis tools can be used to process the content of students' online communication. Meanwhile, with the development of artificial intelligence and big data technology, the algorithms and models of multimodal sentiment analysis are continuously optimized, which makes the evaluation process more efficient and accurate and meets the requirements of the principle of repeatability, which is as shown in Figure 1.

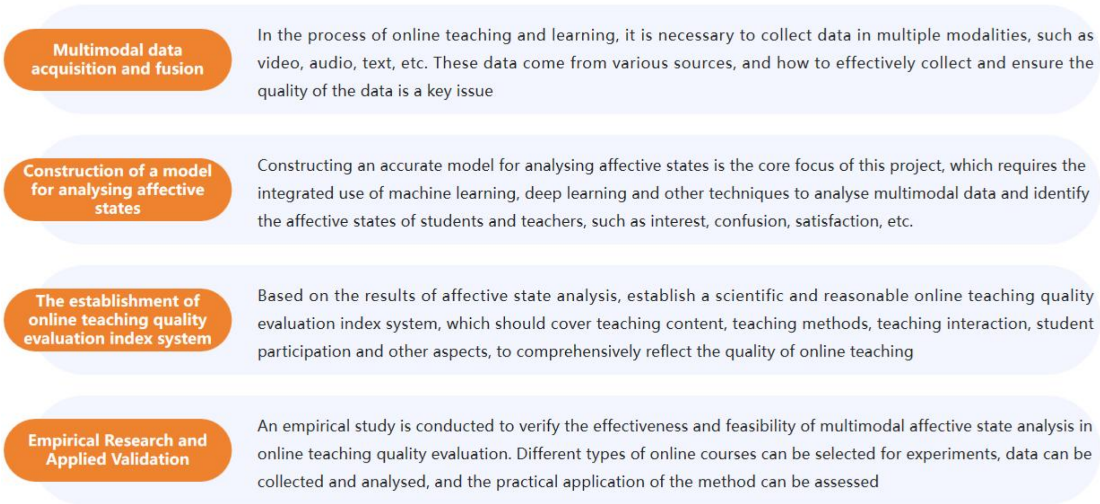

Figure 1 Difficulties in Multimodal-based Emotion Recognition for Teaching and Learning

In multimodal scenarios, methods such as comprehensive evaluation and self-evaluation have an important use value. The comprehensive evaluation method can combine the results of multimodal sentiment analysis to comprehensively assess the quality of online teaching. For example, the hierarchical analysis method is used to determine the weights of each modality information, and the results of the sentiment analysis of different modalities, such as text, speech, and image, are weighted and summed to obtain a comprehensive evaluation score. At the same time, statistical methods such as principal component analysis can also be used to extract the main features in the multimodal information, reduce the data dimensions, and improve the evaluation efficiency. Self-evaluation methods also have positive significance in multimodal online teaching. Students can self-evaluate their learning attitude, emotional state and engagement by reviewing their text communication, voice feedback and video interaction performance in the learning process. Teachers can guide students in self-evaluation to help them better understand their own learning and improve their self-management and learning ability. In addition, self-evaluation can be combined with teacher evaluation, peer evaluation, etc. to form a diversified evaluation system and provide more comprehensive feedback for the improvement of online teaching quality. Therefore, in online teaching quality evaluation, combining the theoretical basis of multimodal sentiment analysis, following the principles of evaluation index system construction, and choosing diversified evaluation methods can more accurately and comprehensively assess the quality of online teaching and provide powerful support for improving teaching effectiveness and students' learning experience.

3 TYPICAL CASE

With the continuous development of multimodal sentiment analysis technology, more and more adaptive teaching systems are beginning to apply this technology, which improves the accuracy and robustness of sentiment classification and detection by obtaining the user's expression during the virtual human's dialogue with the user and the textual content of the dialogue in real time in order to perform multimodal sentiment recognition.

3.1 Adaptive Teaching System Based on Multimodal Emotion Recognition

Video interaction behaviour and learner portrait features are important multimodal feature data in online teaching[13]. By analyzing learners' video interaction behaviors, such as eye contact and body movements, we can understand their level of engagement and concentration. Meanwhile, learners' portrait features, such as facial expressions and eyes, can also reflect their emotional state. For example, when learners are smiling and have focused eyes, it may indicate that they are interested in the content. And when learners frown and have wandering eyes, it may indicate that they are experiencing difficulties or feeling confused. Related studies have shown that multimodal sentiment analysis can improve the accuracy of sentiment recognition by feeding speech data and text data of the audio to be recognized into speech encoder and text encoder respectively. This multimodal emotion recognition technique can help teachers to better understand the emotional state of their students, so as to adjust online teaching strategies and improve teaching effectiveness.The specific architecture is shown in Figure 2.

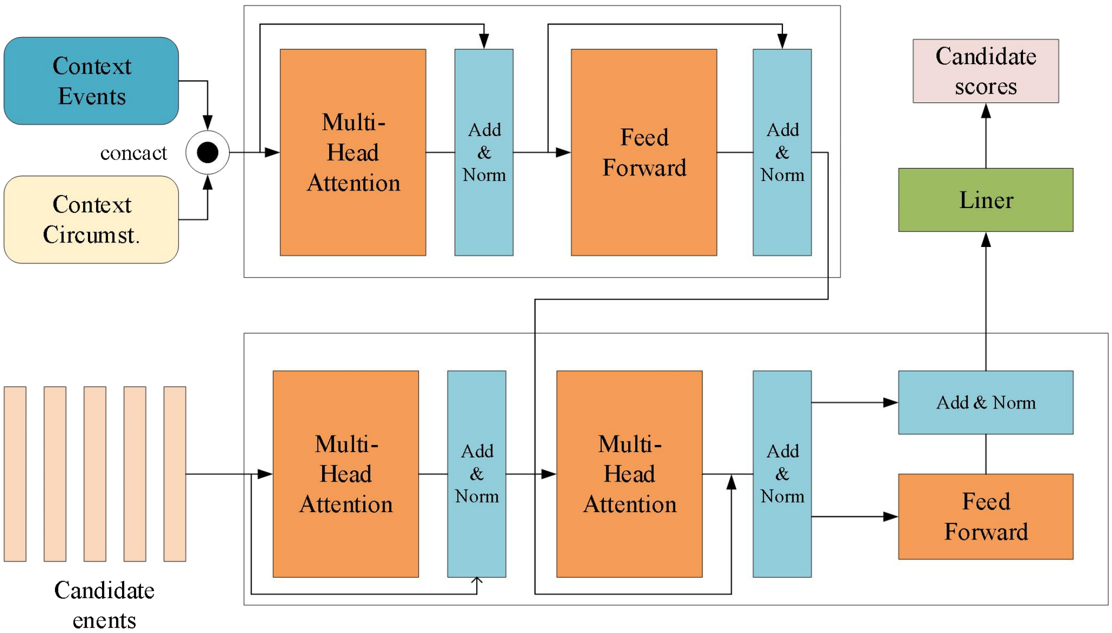

Figure 2 Video Interaction Behaviour and Learner Portrait Feature Application Extraction Architecture

As shown in Figure 2, adaptive teaching system based on multimodal emotion recognition has practical implications for online teaching. Firstly, it can enhance the learners' emotional state. By monitoring learners' affective state in real time, the system can give timely feedback and support, such as encouragement and reminders, to help learners maintain a positive learning attitude[14]. When the system detects that learners are confused, it can provide additional explanations and examples; when learners make progress, it can give praise and rewards. Secondly, this system can improve teaching effectiveness. Teachers can understand the learning situation and needs of students and adjust the teaching content and methods based on the results of multimodal sentiment analysis provided by the system. If most students are confused about a certain knowledge point, the teacher can slow down the teaching progress and add more explanations and exercises. If students are interested in a certain topic, the teacher can guide students to have an in-depth discussion and exploration. Adaptive teaching system based on multimodal emotion recognition has an important application value in online teaching, which can improve the accuracy of emotion recognition and enhance the learners emotional state and teaching effect by analyzing multimodal data such as video interaction behaviour and portrait characteristics of learners.

3.2 Fusing Video Semantic Information for Online Learning Emotion Recognition

The multimodal emotion recognition method fusing video semantic information is of great significance in online learning, which will be introduced in detail in the following with specific cases, as shown in Figure 3.

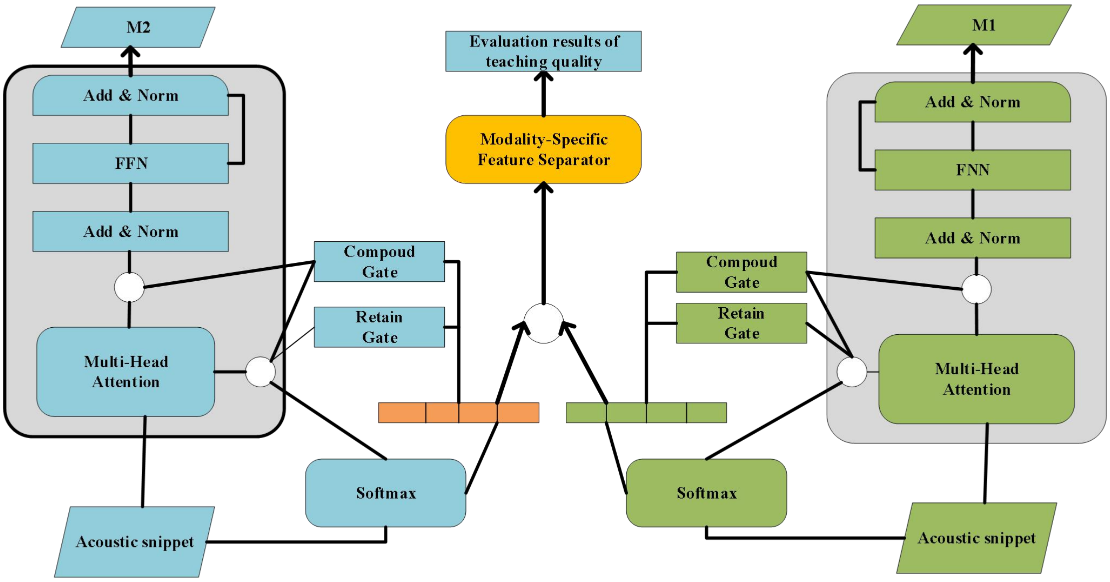

Figure 3 Video Interaction Behaviour and Learner Portrait Feature Application Extraction Architecture

In terms of data collection, this was done in three main ways. Firstly, a flat-panel eye-tracking device was used to collect data on gaze, eye jumps, pupil size, and eye-movement trajectories generated by the learners during the learning process[15]. This process captures learners' visual focus and attention changes while watching instructional videos. Second, wearable ear clip sensors are used to capture the PPG signals generated by the learners during the learning process.The PPG signals can reflect the changes in the physiological state of the learners and provide an objective basis for sentiment analysis. Finally, for video semantic extraction, a video text description is generated to capture the rich semantic information of the video. This textual description includes detailed textual information describing the scenes, objects, actions, and episodes in the video, as well as expressing the emotion and emotional context of the video content.
In the data prepossessing stage, different processing methods are adopted for different types of data. For eye-movement data, since there may be missing data due to subjects blinking, closing their eyes or looking down during the experiment, a linear interpolation method was used to fill in the missing values of the eye-movement data. These filled eye movement data were then baseline corrected to exclude variability between subjects and ensure data consistency. For the PPG signals, the raw PPG signals may be affected by a variety of interfering factors, including motion, light variations, noise, and electromagnetic interference, which may lead to artifacts. Firstly, a filter is used to noise the signal to reduce the effect of high-frequency noise and improve the quality of the PPG signal. Subsequently, baseline correction of the PPG signals was performed to exclude variability between subjects and ensure data consistency.
As shown in Figure 3, in the feature extraction stage, the pee-processed data are subjected to multimodal feature extraction. By comprehensively analyzing the eye movement, PPG signal and video semantic information, features that can reflect the learner's emotional state are extracted. The eye movement features can include gaze time, eye beat frequency, etc. The PPG signal features can include heart rate variation, pulse wave features, etc. Video semantic features can include emotional tendency of teaching content, plot tension, etc. This multimodal feature extraction method can significantly improve the accuracy of emotion recognition. On the one hand, data from different modalities complement each other and provide more comprehensive emotion information. The eye movement data can reflect the learner's attention allocation, the PPG signal can reflect the changes in physiological state, and the video semantic information can provide the impact of the teaching content on the learner's emotion. On the other hand, multimodal feature extraction can reduce the noise interference of single modal data. Relying on eye movement data alone may be affected by individual differences and environmental factors, while combining PPG signals and video semantic information can improve the stability and reliability of emotion recognition.

4 MULTIMODAL SENTIMENT ANALYSIS FOR EVALUATING ONLINE TEACHING QUALITY

In this paper, we intend to explore the optimal strategy for modelling teaching behaviors through online teaching multimodal collaborative data collection, modelling and correlation analysis, and at the same time, analyse the actual development of teaching quality evaluation in China, to develop an adaptive online teaching quality evaluation system based on multimodal data collaboration, and to carry out experimental testing of the functionality through a real teaching platform, with the specific architecture as shown in Figure 4.

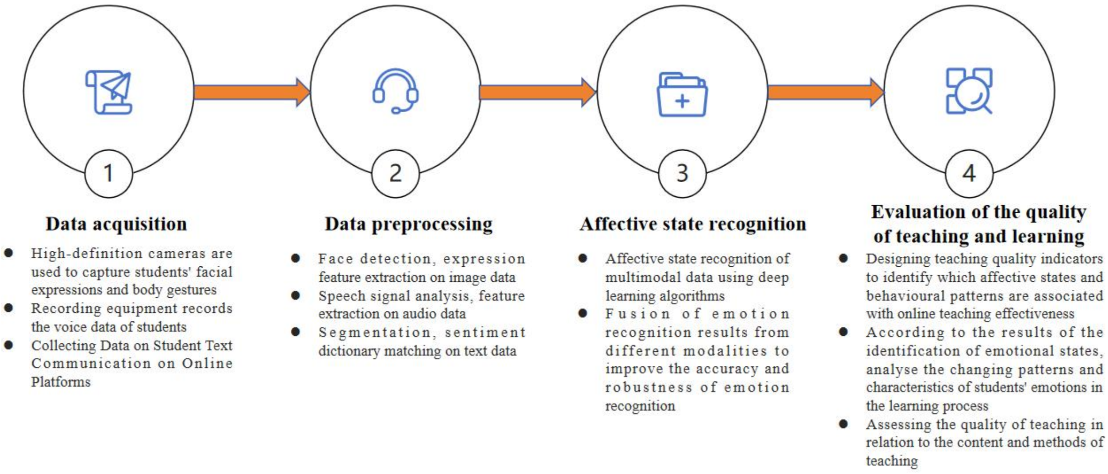

Figure 4 Overall Architecture

4.1 Data Acquisition

Facial Expression Capture, which captures students' facial expressions in real time via a high-definition camera. The camera needs to be placed in a suitable position to ensure that the details of students' facial expressions can be clearly captured. Voice tone capture, using professional recording devices or software, such as microphones or voice recorders, to record students' voice data in the online classroom. This data will be used for subsequent tone of voice analysis. Text content collection, through the online teaching platform or chat tool, collects students' content in text communication areas such as discussion forums and chat rooms. These data will be used for text content analysis.

4.2 Data Prepossessing

Image per-processing, which is performed on the captured facial expression images for face detection, face alignment, expression feature extraction, etc. This usually involves algorithms and techniques in the field of computer vision, such as Haar features, HOG features, Convolution Neural Networks (CNN), etc. Speech processioning, speech signal analysis is performed on the recorded data, including processing such as noise reduction, reverberation, and endpoint detection. Then extract the features of speech such as intonation, speech rate, volume, etc., which will be used for subsequent speech sentiment analysis. Text processioning, the collected text data are processed with word splitting, duplication, lexical labeling and so on. Then tools such as sentiment dictionary and sentiment analysis model are used to judge and classify the sentiment tendency of the text content.

4.3 Affective State Recognition

Multimodal fusion, fuses prepossessed image, speech and text data. This can be achieved by either feature fusion or model fusion. Feature fusion involves splicing or weighted summation of feature vectors from different modalities to obtain a comprehensive feature representation, while model fusion integrates the prediction results from different modalities to improve the accuracy of emotion recognition. Deep learning algorithms, which use deep learning algorithms for emotion state recognition on fused multimodal data. These algorithms can automatically learn complex feature representations in the data and perform sentiment classification or regression based on these features. Sentiment annotation and evaluation, where the recognized sentiment states are annotated and evaluated to ensure the accuracy and reliability of sentiment recognition. This usually involves comparison and validation with manual labeling results.

4.4 Evaluation of the Quality of Teaching and Learning

Affective change analysis, based on the identified affective state data, analyses the changing patterns and characteristics of students' affective changes in the learning process. This helps to understand the students' acceptance of the teaching content, the degree of interest, and the changes in the learning state. Teaching quality assessment, combining teaching content, teaching methods and other factors. By comparing the changes in students' affective state under different teaching sessions or teaching strategies, the quality of teaching can be assessed and problems and directions for improvement can be identified. Feedback and suggestions, based on the results of teaching quality assessment, provide teachers with targeted feedback and suggestions. These feedback and suggestions can help teachers optimize teaching strategies and improve teaching methods, thus enhancing the quality of online teaching.

In summary, the online teaching quality evaluation method based on multimodal affective state analysis achieves the comprehensive capture and analysis of students' affective states through the steps of data collection, data prepossessing, affective state recognition and teaching quality evaluation, which provides a more objective and comprehensive basis for teaching quality evaluation.

5 CONCLUSION

In this study, the online teaching quality evaluation method based on multimodal sentiment state analysis is deeply explored, and the following main results are achieved. First, multimodal sentiment analysis provides new ideas and methods for online teaching quality evaluation. By integrating multimodal information such as text, voice, image, etc., it is able to capture students' affective states in the online learning process more comprehensively and accurately, overcoming the limitations of traditional evaluation methods. In terms of theoretical foundation, the concepts and principles of multimodal sentiment analysis are clarified, including multimodal data fusion technology and sentiment feature extraction technology. Meanwhile, an online teaching quality evaluation index system based on the principles of scientifically and operability is constructed, and diversified evaluation methods are selected to provide theoretical support for the application of multimodal sentiment analysis in online teaching. In terms of application cases, the adaptive teaching system based on multimodal emotion recognition and the online learning emotion recognition method that integrates video semantic information improve the accuracy of emotion recognition and enhance the emotional state of learners and the teaching effect by analyzing the multimodal data such as video interaction behaviors, learner's portrait features, eye movement data, PPG signals and video semantic information. In terms of specific strategies, multimodal emotion indicators were constructed, including textual modal emotion indicators and image and sound modal emotion indicators, which provide multidimensional information for teaching quality evaluation. Meanwhile, the association between multimodal sentiment analysis and teaching quality is analyzed, and the influence of affective state on learning engagement and learning effect is clarified. Multimodal sentiment analysis has an important application value in online teaching quality evaluation and can provide powerful support for improving online teaching quality. However, there are some shortcomings in this study, such as the complexity of multimodal data fusion technology and the accuracy of sentiment feature extraction, which need to be further explored and improved in future research.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This work is partially supported by 2024 Guangdong Province Education Science Planning Project (Higher Education Special Project) "Research on Smart Classroom Teaching Behavior Analysis Methods Based on Scene Semantic Understanding and Deep Learning Characteristic Representation (2024GXJK766)", 2023 Guangdong Provincial Higher Vocational Education Teaching Reform Research and Practice Project "Research on Online Teaching Quality Evaluation Method Based on Multimodal Affective State Analysis (2023JG277)", 2024 Shenzhen Polytechnic University Quality Engineering Project "Research on Classroom Scene Understanding and Behavior Analysis Method Based on Multimodal Attention Mechanisms (7024310268)", 2024 Higher Education Scientific Research Planning Project of the Chinese Society of Higher Education "Research on the Analysis of Teaching and Learning Deep Interaction Characteristics in Smart Classroom Environment Supported by Multimodal Data (24XH0407)".

REFERENCES

[1] Shuyan W. Multi-attribute decision-making methods based on Aczel-Alsina power aggregation operators for English blended teaching quality evaluation under single-valued neutrosophic sets. International Journal of Knowledge-based and Intelligent Engineering Systems, 2024, 28(3): 539-552.

[2] Arega B, Mengistu M, Mersha A, et al. Evaluation of hospital quality of care outcomes in a teaching hospital in Ethiopia: a retrospective database study. BMJ open, 2024, 14(9): e082908.

[3] Guo S, Yang X, Farizan H N, et al. The analysis of teaching quality evaluation for the college sports dance by convolutional neural network model and deep learning. Heliyon, 2024, 10(16): e36067-e36067.

[4] Sharma A ,Nishadham V, Gupta P, et al. Evaluation of Turnaround Times of Diagnostic Biopsies: A Metric of Quality in Surgical Pathology. International journal of surgical pathology, 2024, 10668969241261561.

[5] Liu M. A combined exponential TODIM-GRA framework for multiple-attribute group decision-making under 2-tuple linguistic Pythagorean fuzzy sets and applications to art teaching quality evaluation in higher education institutions. Soft Computing, 2024, 28(17-18): 10317-10330.

[6] Ma W, Chu N. Evaluation of online teaching quality in colleges and universities based on digital monitoring technology. Journal of Intelligent Systems, 2024, 33(1).

[7] Camilo S, Edwin M, Jose A. Analysis of different affective state multimodal recognition approaches with missing data-oriented to virtual learning environments. Heliyon, 2021, 7(6): e07253-e07253.

[8] Caridakis G, Karpouzis K, Wallace M, et al. Multimodal user's affective state analysis in naturalistic interaction. Journal on Multimodal User Interfaces, 2010, 3(1-2): 49-66.

[9] Johnstone E, Liao H M, Strani K. How does it feel? Space and text as affective affordances in the multimodal museum. Multimodality & Society, 2024, 4(3): 314-329.

[10] VillenaGonzalez M. Caresses, whispers and affective faces: A theoretical framework for a multimodal interoceptive mechanism underlying ASMR and affective touch: An evolutionary and developmental perspective for understanding ASMR and affective touch as complementary processes within affiliative interactions. BioEssays, 2023, 45(12): e2300095-e2300095.

[11] Xiaoshuang G. Multimodality in language education: implications of a multimodal affective perspective in foreign language teaching. Frontiers in Psychology, 2023, 141283625-1283625.
[12] Olsen D. Media Arts Education:Transforming Education Through Multimodal Cognition, Holistic Learning, and Techno-Embodiment. Taylor & Francis: 2024-09-25. DOI:10.4324/9781003430834.
[13] Kewalramani S, Aranda G, Sun J, et al. A Systematic Review of the Role of Multimodal Resources for Inclusive STEM Engagement in Early-Childhood Education. Education Sciences, 2024, 14(6): 604-604.
[14] Anne-Marie H, M S M, P T H, et al. Falls After Hospital Discharge: A Randomized Clinical Trial of Individualized Multimodal Falls Prevention Education. The journals of gerontology. Series A, Biological sciences and medical sciences, 2019, 74(9): 1511-1517.
[15] Giannakos N M, Lee-Cultura S, Sharma K. Sensing-Based Analytics in Education: The Rise of Multimodal Data Enabled Learning Systems. IT Professional Magazine, 2021, 23(6): 31-38.