Optimization Strategies for the Construction and Application of Artificial Intelligence Education Evaluation System Based on Multimodal Data Fusion

Hui Wu[1,*], Xiya Wang[2]

[1]Medical School, Shaoguan University, Shaoguan, China

[2]College of Biological Engineering, Xinxiang University, Xinxiang, China

*Corresponding author: 1825949137@qq.com

Keywords: AI; Education Assessment; Multimodal Data; Personalized Evaluation; Learning Analytics

Abstract: This paper focuses on the innovative development of education assessment. In response to the limitations of traditional education assessment methods, this paper conducts in-depth research on an artificial intelligence-based education assessment system utilizing multimodal data fusion. By integrating multi-source data, including text, images, audio, and video, and combining machine learning and deep learning technologies, a comprehensive, objective, and personalized education assessment system is constructed. This paper discusses the theoretical foundation, technical framework, and the specific processes involved in system construction. It analyzes application examples in various contexts, such as school education and online learning. Additionally, it proposes optimization strategies from the perspectives of technology, education, teaching, management, and policy. Research shows that the evaluation system effectively enhances the accuracy and efficiency of educational evaluation, provides strong support for improving educational quality, and is also of great significance in promoting educational equity and personalized development.

1. Introduction

In the dynamic development of the education system, education evaluation has always played a key role. It is not only important for measuring the quality of education but also a core driving force for promoting education reform and innovation. However, the traditional education evaluation model has long relied on single-dimensional evaluation methods such as standardized tests and written assignments. While the "result-oriented" evaluation mechanism can quantify students' knowledge mastery to some extent, it fails to fully capture the development trajectory of students' critical thinking and practical application skills throughout the learning process. In actual teaching, important qualities such as students' innovative thinking during classroom discussions, their spirit of exploration when facing difficult problems, and their communication skills in teamwork are often overlooked due to ineffective evaluation methods. This oversight leads to a disconnect between evaluation results and students' actual learning status, making it challenging to provide an accurate basis for personalized educational decisions.

With the accelerated advancement of digital transformation in education, the educational landscape is undergoing profound changes. The popularity of online education platforms and the widespread adoption of intelligent teaching equipment have led to a surge in educational data. At the same time, the vigorous development of multimodal data fusion and artificial intelligence technology has brought new ways to solve the difficulties in traditional education evaluation. Multimodal data, with its diverse sources, records students' learning behaviors and processes across multiple dimensions, including text, images, audio, and video. For example, videos can capture students' body language and expression changes to reflect their concentration. Additionally, voice interaction data can be used to analyze students' language expression and logical thinking. The text data generated by the learning platform can show the students' knowledge construction process. Moreover, the machine learning and deep learning algorithms in artificial intelligence technology possess powerful data analysis and pattern recognition capabilities, which can effectively mine and integrate complex

  DOI: 10.25236/gemmsd.2025.047

multimodal data, thereby enabling dynamic monitoring and accurate evaluation of students' learning conditions [1].

In this context, building an artificial intelligence education evaluation system based on multimodal data fusion is not only an inevitable trend in line with the development of digital education but also an urgent requirement to meet the needs of personalized education. This study aims to overcome the limitations of traditional evaluation models, integrate multimodal data resources and the advantages of artificial intelligence technology, develop a scientific, comprehensive, and dynamic education evaluation system, and thoroughly explore its optimization strategies in practical applications. It helps to improve the accuracy and effectiveness of educational evaluation, providing a reliable basis for teachers to adjust their teaching strategies and for students to optimize their learning methods. Additionally, it has far-reaching theoretical significance and practical value in promoting educational equity and realizing personalized educational development.

2. Theoretical Basis and Technical Framework

2.1 Principle of Multimodal Data Fusion

Multimodal data refers to different types of data obtained through multiple channels. In the field of education, text data includes students' homework and papers; image data comprises photos of students' expressions and actions recorded in class; audio data encompasses classroom conversations between teachers and students, as well as students' voices when answering questions; and video data records the complete classroom teaching process. These different modal data contain rich information and reflect students' learning situations from different angles. Multimodal data fusion mainly includes feature-level fusion and decision-level fusion. Feature-level fusion involves combining data at the feature level to extract and integrate the features of different modal data, thereby forming a more comprehensive and representative feature vector. Decision-level fusion involves analyzing and making decisions on each modal data separately and then synthesizing these decision results to reach a conclusion. Taking the evaluation of student's participation as an example, feature-level fusion is performed by extracting action features in the video and speech frequency in the audio, or decision-level fusion is performed after analyzing students' activity according to the video and students' speech according to the audio, to more accurately evaluate students' classroom participation.

2.2 Technology of AI Education Assessment

Machine learning and deep learning are the core technical support for AI education assessment. Machine learning algorithms, such as classification algorithms, can be used to determine students' mastery of knowledge, regression algorithms can predict students' future academic performance, and clustering algorithms can group students with similar learning behaviors [2]. Additionally, neural network models in deep learning, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), possess unique advantages in processing multimodal data. CNNs are well-suited for processing image data and can extract key features from images; RNNs are more suitable for processing sequence data, such as students' learning time series and text content sequences. Figures 1 and 2 are schematic diagrams of the structures of CNNs and RNNs. Through these algorithms and models, multimodal data can be thoroughly analyzed, which helps identify students' learning behaviors, such as determining whether students are focused in class and predicting learning outcomes.

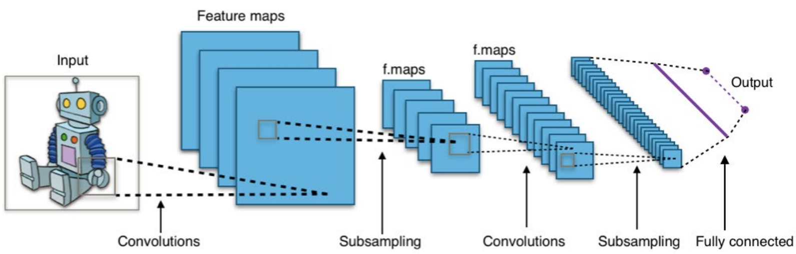

Fig. 1 Schematic diagram of CNN

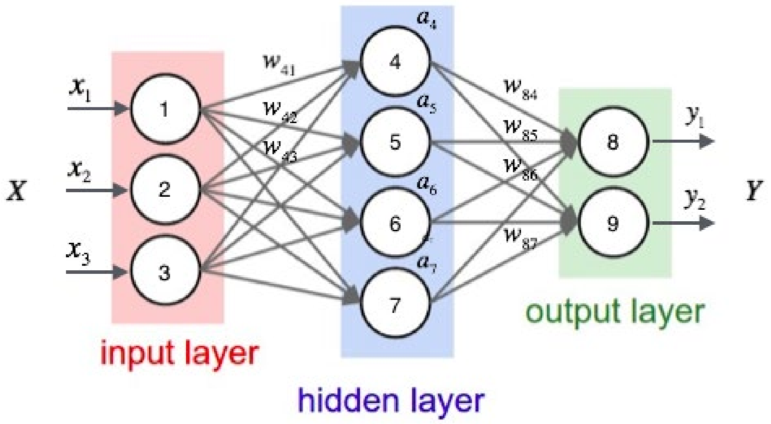

Fig. 2 Schematic diagram of RNN

2.3 Theoretical Basis of Educational Evaluation

Educational measurement theory provides a basis for the quantitative analysis of educational evaluation, emphasizing the quantitative evaluation of student's knowledge and skills through scientific measurement methods. It also provides theoretical guidance for designing indicators in the knowledge and skills dimension of the multimodal evaluation system. The theory of multiple intelligences posits that everyone possesses multiple intelligences, including linguistic intelligence, logical-mathematical intelligence, spatial intelligence, and others. This theory advocates for a shift in educational evaluation from a single dimension to multiple dimensions, providing an important basis for designing indicators that encompass dimensions such as emotional attitudes and learning processes within the evaluation system [3]. Based on these theories, we have developed an educational evaluation framework that integrates theory with practice to ensure that the evaluation system comprehensively and scientifically reflects students' learning situations.

3. Construction and Design of Evaluation System

3.1 Principles of System Construction

To build an AI education evaluation system based on multimodal data fusion, we must strictly follow the four core principles of comprehensiveness, objectivity, personalization, and dynamism. The principle of comprehensiveness requires the evaluation system to break through the limitations of traditional single dimensions and include students' knowledge and skills, learning process performance, and emotional attitude development within the evaluation scope. For example, in addition to considering test scores, we also need to pay attention to student's ability to solve practical problems in group projects and their psychological adjustment abilities when facing learning difficulties to ensure that the evaluation results reflect the students' comprehensive qualities. The principle of objectivity emphasizes that the evaluation process should minimize human interference and use multimodal data collection technology and scientific algorithms for analysis. Students learning behavior data is automatically recorded through smart devices, and machine learning algorithms are used to quantify these data, thereby avoiding the influence of teachers' subjective bias on the evaluation results and ensuring that the evaluation results are accurate and reliable [4].

The principle of personalization begins with the individual differences of students, examining each student's unique learning style, strengths, and weaknesses through the analysis of multimodal data. For example, targeted image and video resources can be provided for visual learners, while students with strong logical thinking skills are presented with more challenging reasoning questions. The principle of dynamism emphasizes the continuous monitoring of students' learning processes. It utilizes the real-time characteristics of multimodal data to capture changes in students during the learning process promptly. When it is found that students' learning enthusiasm has declined, the system automatically issues an early warning to help teachers adjust teaching strategies promptly and help students achieve better learning outcomes.

3.2 Data Collection and Processing

Data collection is primarily conducted through learning management systems, smart terminal devices, and sensors. Learning management systems can record various types of data, including students' homework submissions, online learning time, and test scores. Smart terminal devices, such as tablets and smartphones, can be used to collect students' learning video and audio data. Sensors can collect students' physiological data in class, such as heart rate and facial expressions. The collected data requires preprocessing, including cleaning, labeling, and normalization. Data cleaning is to remove duplicate, erroneous, or incomplete data; labeling is to add labels to the data for algorithm recognition and analysis; normalization is to unify data of different scales into the same range, improve the comparability of data and the accuracy of the algorithm, and ensure that the data quality meets the evaluation requirements [5].

3.3 Evaluation Indicators and Model Construction

The evaluation index system encompasses knowledge and skills, the learning process, and emotional attitude. The knowledge and skills dimension include indicators such as the mastery of subject knowledge and problem-solving ability. The learning process dimension encompasses aspects such as learning time allocation, learning resource usage, and problem-solving strategies. The emotional attitude dimension encompasses learning interest, learning motivation, class participation, and other related factors. For each indicator, we have clarified specific quantitative methods, such as quantifying the mastery of knowledge through homework and test scores and quantifying class participation through classroom video analysis. For the construction of the evaluation model, the researchers selected appropriate machine learning or deep learning models, such as building a comprehensive evaluation model based on neural networks. The model is trained using a large amount of labeled data, and its parameters are continuously adjusted to optimize performance. The model is then verified using verification data, and its reliability and effectiveness are evaluated using indicators such as accuracy and recall to ensure that it can accurately assess students' learning [6].

4. Application and Optimization Strategies

4.1 Technology Optimization

On a technical level, the continuous improvement of data fusion algorithms is key to enhancing the performance of the evaluation system. In the current multimodal data fusion process, the characteristics of different modal data are very different, and the information complementarity is complex. Therefore, it is difficult for traditional algorithms to fully tap the value of data. To this end, we can explore data fusion algorithms based on attention mechanisms, which can dynamically assign weights and focus on key data features according to the importance of different data modalities to the evaluation target. For example, when evaluating students' classroom learning concentration, the attention mechanism is used to automatically determine the weight of the facial expression data in the visual modality and the speech frequency data in the audio modality, thereby more accurately fusing the data. At the same time, it is crucial to establish a continuous optimization and update mechanism for artificial intelligence models. With the continuous changes in educational scenarios and the generation of new data, the model must adjust its structure and parameters in a timely manner to adapt to new needs. For example, the transfer learning method utilizes the training results of existing models on similar tasks to quickly adjust the model and adapt it to new educational evaluation tasks, thereby improving the model's generalization ability.

Data security and privacy protection are important links in technology optimization. In the education evaluation system, student data involves personal privacy and sensitive information, and once leaked, it will have serious consequences. It is recommended to use end-to-end encryption technology to ensure the security of data during collection, transmission, and storage, thereby preventing data from being stolen or tampered with. Differential privacy technology perturbs the original data to protect students' privacy without compromising the accuracy of data analysis. For example, when statistically analyzing the distribution of students' academic performance, data anonymization is achieved by adding controllable noise, thereby effectively addressing technical issues such as algorithm bias and data leakage while ensuring the stability and security of the evaluation system.

4.2 Measures to Optimize Education and Teaching

Formulating personalized teaching strategies based on evaluation results is the core of improving the quality of education and teaching. The multimodal data evaluation system can accurately analyze students' strengths and weaknesses in knowledge and skills, learning processes, and emotional attitudes [7]. For example, when the evaluation reveals that students have weaknesses in mathematical and logical reasoning but are more adept at graphical cognition, teachers can provide targeted videos explaining graphical and mathematical problem-solving and assign relevant exercises to guide students in transferring their strengths to weak areas. At the same time, according to the differences in students' learning styles, more image and video learning resources are provided for visual learners, while more audio explanations and interactive discussions are arranged for auditory learners, thereby achieving "teaching students by their aptitude" in the true sense.

Designing a systematic teacher training program is the key to ensuring that the evaluation system is effectively applied to teaching. As the implementers of education and teaching, teachers' understanding and application of evaluation tools directly impact the effectiveness of their teaching. The training content should cover the theoretical basis of the evaluation system, data interpretation methods, and how to transform evaluation results into teaching strategies. For example, case teaching training can be conducted to guide teachers in analyzing the student learning problems reflected in actual student evaluation data cases and formulating corresponding improvement measures. In addition, innovating methods for students to participate in evaluation and self-reflection will help cultivate students' autonomous learning abilities. It is necessary to introduce diverse evaluation methods, such as peer evaluation and self-evaluation, to guide students in participating in the evaluation process. For example, in group project learning, let students evaluate each other's teamwork abilities, task completion quality, etc., and write a learning summary in combination with self-reflection to encourage students to think deeply about their learning process and results, thereby achieving mutual benefit between teaching and learning.

4.3 Management and Policy Guarantee

Establishing and improving the management norms of the evaluation system is essential for ensuring that the evaluation process is both scientific and fair. Detailed operating guidelines need to be formulated to clarify each link of the evaluation process, from the standard requirements for data collection and the standard process for data processing to the generation and review of evaluation results. For example, managers specify the installation location of data collection equipment and the time range for data collection to ensure the objectivity and consistency of the data collection process. It is recommended to standardize the algorithm selection and parameter settings in data processing to ensure the scientific integrity of the data processing. Additionally, it is recommended to implement strict data usage rules to clarify access rights, the scope of use, and the retention period, thereby preventing data misuse. It is necessary to formulate application standards for evaluation results to guide teachers and education managers in using them reasonably and avoiding over-interpretation or incorrect application.

Promoting the formulation of relevant policy standards can provide macro guidance and policy support for the construction and application of the evaluation system. The education authorities should collaborate with scientific research institutions, schools, and enterprises to jointly study and formulate industry standards for evaluating multimodal data education, including data format standards, evaluation indicator system standards, and model evaluation standards. For example, unifying the storage format of multimodal data will facilitate data sharing and interaction between different educational platforms. Building a collaborative cooperation platform for schools, educational institutions, enterprises, and other stakeholders will help promote the sharing of resources and the exchange of experiences. Furthermore, it is essential to encourage schools to collaborate with technology companies in developing evaluation technologies and tools that are appropriate for educational settings. Educational institutions should be supported in organizing experience-sharing sessions focused on the use of evaluation systems. By promoting practical examples and leveraging the strengths of all parties involved, we can collaborate to improve and advance artificial intelligence education evaluation systems that utilize multimodal data fusion. This teamwork will also help ensure the sustainable application of these evaluation systems.

5. Conclusion

This study successfully developed an artificial intelligence education evaluation system based on multimodal data fusion and proposed a systematic optimization strategy for its application. By integrating multimodal data and artificial intelligence technology, the evaluation system can evaluate students' learning situations more comprehensively and accurately. The application of multiple scenarios, such as school education and online education, effectively enhances the efficiency and quality of educational evaluation, providing strong support for improving educational quality. However, the research also has certain limitations. In terms of data acquisition, the collection of some data may be subject to technical and ethical restrictions. The generalization ability of the models still requires significant improvement, and algorithmic bias needs to be addressed further. Future research should focus on exploring more efficient data collection technologies, optimizing algorithm models, and enhancing interdisciplinary collaboration. This approach can help refine the evaluation system, allowing it to better promote educational equity and facilitate personalized development, ultimately providing a strong foundation for the advancement of educational modernization.

References

[1] Tanberkan H, Zer M, Gelbal S. Impact of Artificial Intelligence on Assessment and Evaluation Approaches in Education[J]. International Journal of Educational Studies & Policy (IJESP), 2024, 5(2). DOI:10.5281/zenodo.14016103.

[2] Jin L, Zhang H. New Opportunities, Challenges, and Strategies for Educational Evaluation Reform in the Era of Artificial Intelligence[J]. Journal of Contemporary Educational Research, 2025, 9(1):66-73.

[3] Zhou J H, Zhang J, Li H. Exploring the Use of Artificial Intelligence in Teaching Management and Evaluation Based on Citation Space Analysis[J]. Journal of Education and Educational Research, 2023. DOI:10.54097/jeer.v3i2.9014.

[4] San F G, Zkaya A B, Tengiz F, et al. Current evaluation and recommendations for the use of artificial intelligence tools in education[J]. Turkish Journal of Biochemistry / Turk Biyokimya Dergisi, 2023, 48(6). DOI:10.1515/tjb-2023-0254.

[5] Bian Y, Lu Y, Li J. Research on an Artificial Intelligence-Based Professional Ability Evaluation System from the Perspective of Industry-Education Integration[J]. Scientific Programming, 2022.DOI:10.1155/2022/4478115.

[6] Rong Z, Gang Z. An artificial intelligence data mining technology based evaluation model of education on political and ideological strategy of students[J]. J. Intell. Fuzzy Syst. 2021, 40:3669-3680. DOI:10.3233/jifs-189401.

[7] Torres-Gutiérrez, Arturo, Alfredo Lino-Gamio J ,de la Cruz Díaz-Ledezma, José,et al. Evaluation of Experience in Higher Education, innovating with artificial intelligence in the implementation of predictive models[J]. Journal High School, 2024, 8(19). DOI:10.35429/JHS.2024.8.19.1.11.