Advances in Educational Technology and Psychology (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/aetp.2025.090605
ISSN 2371-9400 Vol. 9 Num. 6

Research on Personalized Teaching Recommendation Model Integrating Attention Mechanism and Cluster Analysis

Xinyi Chen

School Mathematical Sciences and Applications, Nanjing Normal University Taizhou College, Taizhou, Jiangsu, China
1910936803@qq.com

Keywords: Personalized Teaching; Recommendation System; Bi-LSTM; Attention Mechanism; K-Means Clustering; Learning Behavior Modeling

Abstract: To improve the accuracy and interpretability of personalized teaching recommendations in primary and secondary school settings, this paper proposes a teaching recommendation model that integrates behavioral modeling and cluster analysis. This model first uses a bidirectional long short-term memory (Bi-LSTM) network to model students' learning behavior sequences to capture the dynamic changes in their learning process. It then introduces an attention mechanism to focus on key behavioral segments and generate recognizable individual feature representations. Finally, K-Means clustering is used to construct student learning profiles, enabling hierarchical resource recommendations. Experimental results demonstrate that this model outperforms comparative methods across multiple evaluation metrics (Precision, Recall, and NDCG), demonstrating superior recommendation performance and clustering quality. This research provides effective insights for implementing lightweight and scalable intelligent teaching recommendation systems and lays the foundation for the implementation of AI applications in education.

1. Introduction

In recent years, with the continuous advancement of educational informatization and the rapid development of artificial intelligence (AI), personalized instruction has gradually become a key direction for reform in basic education [1]. Compared to traditional teaching methods that rely on a unified schedule and resources, personalized instruction emphasizes a student-centered approach, implementing differentiated teaching strategies based on their knowledge base, cognitive style, and learning trajectory [2]. Faced with the increasing abundance and diversity of student behavioral data, how to effectively leverage AI to mine individual characteristics and accurately match teaching resources has become a key topic in current educational technology research [3]. However, the practical application of existing personalized recommendation systems in basic education settings still faces many challenges, such as limited ability to model student behavioral characteristics, a lack of dynamic and interpretable recommendation results, and difficulty adapting to the complex demands of teaching scenarios [4].

To address these issues, this paper proposes a personalized teaching recommendation model that integrates an attention mechanism with cluster analysis. This model takes multidimensional student behavior sequences as input. It first uses a bidirectional long short-term memory (Bi-LSTM) network to model the temporal characteristics of students' learning processes, capturing their learning rhythm and behavioral evolution patterns. It then introduces an attention mechanism to weightedly extract representative key moments in the behavior sequence to construct more expressive individual embedding vectors. Furthermore, it uses a K-Means clustering algorithm to profile student groups, generating interpretable learner category labels that drive tiered recommendations of personalized teaching resources. The synergistic effect of the model's various modules not only improves the alignment of recommended content with student needs but also demonstrates strong scalability and practical adaptability.

The main contributions of this paper are as follows: First, we construct a teaching recommendation framework that integrates time series modeling, attention extraction, and behavior clustering, enabling automated modeling from learning behavior to teaching intervention. Second, we conduct an empirical analysis of the model on a public educational behavior dataset, verifying its performance advantages in recommendation accuracy and clustering effectiveness by comparing it with existing typical methods. Third, by combining the attention weights and cluster center characteristics of the recommendation process, we conduct a visual analysis of the recommendation logic, providing decision support for teachers and education administrators. This research is expected to provide a lightweight, explainable, and deployable algorithmic paradigm for the implementation of personalized teaching models in intelligent education scenarios.

2. Related Work

Personalized instructional recommendations, a key branch of intelligent education, have garnered widespread attention in recent years [5]. Early instructional recommendation systems were mostly designed based on rules or collaborative filtering, relying primarily on static student information (such as basic grades and course preferences) to match content [6]. While simple to implement, these approaches struggle to capture dynamic changes in students' learning processes, and their recommendations lack real-time performance and adaptability [7]. To improve the generalization and personalized responsiveness of these models, researchers have begun incorporating machine learning and data mining techniques to construct data-driven instructional recommendation models [8]. For example, some researchers have used matrix factorization to construct a student-resource rating matrix and latent factor modeling to implement recommendations [9]. However, these approaches often ignore the temporal dimension and behavioral sequence information, making them unable to address short-term interest fluctuations and the dynamic evolution of learning paths [10].

To more accurately characterize student learning behavior sequences, sequence modeling methods have been widely adopted in personalized instructional recommendation tasks in recent years [11]. Among them, LSTM, due to its excellent sequence memory capabilities, has become a mainstream tool for learning behavior modelling [12]. Compared to traditional RNN structures, LSTM is better at capturing long-range dependencies and is suitable for processing educational data such as question-answering sequences, homework trajectories, and resource click records [13]. Furthermore, Bi-LSTM can simultaneously consider forward and backward information flow during modeling, effectively improving sequence representation capabilities [14]. However, LSTM itself still suffers from information dilution when processing long sequences. Therefore, researchers have introduced an attention mechanism to enhance the model's focus on key behavioral segments. This attention mechanism not only improves the model's representational accuracy but also enhances its interpretability, facilitating the tracking of recommendation rationales during instructional interventions [15].

In addition to behavioral modeling and feature extraction, effective matching and distribution of personalized teaching resources is also a key research direction. Some current research attempts to construct student profiles using clustering algorithms, grouping students with similar behavioral characteristics into similar groups and recommending resources based on group center features. K-Means, a classic unsupervised clustering method, is widely used in educational profile modeling due to its high computational efficiency and strong interpretability. By inputting the learned behavior vectors after attention enhancement into K-Means clustering, classification quality can be improved to a certain extent and structured support can be provided for subsequent recommendations. In recent years, some research has used clustering results as a basis for teaching intervention, enabling layered content push and resource recommendations. However, most work remains focused on clustering itself, with relatively crude integration of recommendation modules. In summary, a recommendation framework that deeply integrates time series modeling, attention mechanisms, and cluster analysis is still lacking. Building on this foundation, this paper proposes a unified modeling approach to achieve more accurate, dynamic, and interpretable personalized teaching recommendation services.

3. Methods

This paper proposes a personalized teaching recommendation model that integrates behavior sequence modeling, attention mechanisms, and cluster analysis. The overall architecture consists of four main modules: a behavior data encoding layer, an attention feature extraction layer, a cluster modeling layer, and a recommendation generation layer. These modules work together to achieve accurate modeling of student behavior data and intelligent distribution of recommended resources. The model structure is shown in Figure 1.

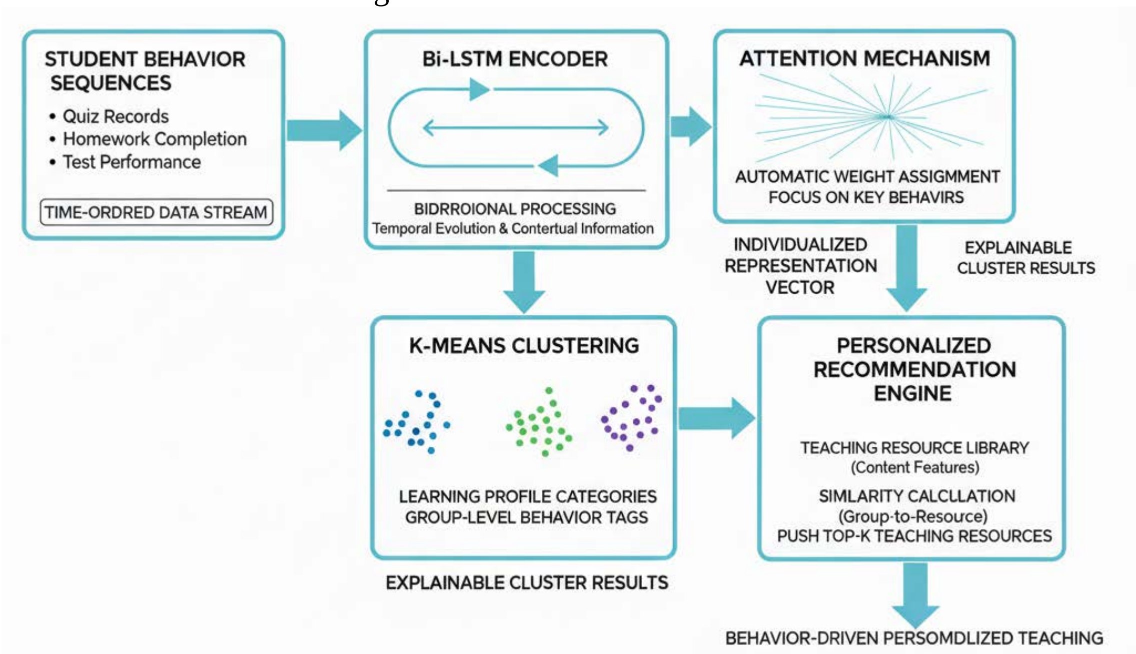

Figure 1: Overall methodological framework

3.1 Learning Behavior Sequence Modeling: Bi-LSTM

Let each student's learning behavior data be represented as a time series $\mathbf{X} = \{x_1, x_2, \dots, x_T\}$, where $x_t \in \mathbb{R}^d$ denotes the feature vector of learning behavior at time step t (such as question response time, correctness rate, video progress, etc.), and T is the sequence length. We use a Bi-LSTM to encode the behavioral sequence.

The forward and backward hidden states are denoted as $\vec{h}_t$ and $\overleftarrow{h}_t$, respectively. The final output at each time step is obtained by concatenating both directions:

$$h_t = [\vec{h}_t ; \overleftarrow{h}_t] \in \mathbb{R}^{2h} \tag{1}$$

Where [;] denotes the concatenation operation, and h is the hidden size of the LSTM. After encoding, we obtain the contextual representation of the entire sequence as $\mathbf{H} = \{h_1, h_2, \dots, h_T\}$.

3.2 Key Behavior Focusing: Attention Mechanism

Since not all time steps in a student's learning behavior sequence contribute equally to learning performance, an attention mechanism is introduced to assign different weights to significant behavioral moments. This mechanism enables the model to focus more on the critical interactions that reflect learning intent or cognitive change.

In this study, we adopt the additive attention mechanism (Bahdanau attention) to compute the importance score αt of each time step. The attention weights and aggregated representation are computed as follows:

$$e_t = v^\top \tanh(W_h h_t + b_h), \quad \alpha_t = \frac{\exp(e_t)}{\sum_{k=1}^{T} \exp(e_k)} \tag{2}$$

$$r = \sum_{t=1}^{T} \alpha_t h_t \tag{3}$$

Here, $W_h \in \mathbb{R}^{d_a \times 2h}$, $b_h \in \mathbb{R}^{d_a}$, and $v \in \mathbb{R}^{d_a}$ are trainable parameters; e_t represents the unnormalized attention energy of the t-th hidden state; α_t denotes its normalized attention weight; and $r \in \mathbb{R}^{2h}$ is the final aggregated representation vector that captures the overall behavioral embedding of the student.

Through this attention-based weighting, the model can emphasize meaningful behavioral events (e.g., conceptual transitions, task completion patterns) and suppress noise from redundant or less informative steps, thereby improving the interpretability and efficiency of the personalized recommendation model.

3.3 Student Group Modeling: K-Means Clustering

To enhance the interpretability and structural organization of personalized teaching recommendations, this study employs K-Means clustering on the attention-based representations of students. The goal is to identify latent learning patterns and group students with similar behavioral characteristics, enabling differentiated teaching resource recommendations across clusters.

Let $\{r_1, r_2, \dots, r_N\}$ denote the set of attention-enhanced embedding vectors obtained from the previous stage, where each $r_i \in \mathbb{R}^{2h}$ represents the behavioral embedding of the i-th student, and N is the total number of students. K-Means aims to partition these representations into K clusters $\{C_1, C_2, \dots, C_K\}$ by minimizing the intra-cluster distance while maximizing inter-cluster separation. The optimization objective is formulated as:

$$\mathcal{L}_{\text{cluster}} = \sum_{i=1}^{N} \left\| r_i - c_{z_i} \right\|^2 \tag{4}$$

Where c_{z_i} denotes the centroid of the cluster assigned to the i -th student, and $z_i \in \{1,2,\dots,K\}$ is the cluster label. The algorithm iteratively updates the cluster assignments and centroids until convergence, as shown below:

$$c_k = \frac{1}{|C_k|}\sum_{r_i \in C_k} r_i, \quad z_i = \arg\min_k \| r_i - c_k \|^2 \tag{5}$$

Through this clustering process, students exhibiting similar temporal learning dynamics and behavioral patterns are grouped together. The resulting clusters can serve as interpretable learner profiles that provide the basis for group-level personalized recommendation, where each cluster corresponds to a specific learning style or knowledge mastery level. Consequently, this approach bridges the gap between individual modeling and group-based teaching decision-making, supporting scalable deployment in intelligent education systems.

3.4 Personalized Teaching Recommendation Strategy

Based on the cluster labels obtained in the previous stage, this study designs a lightweight and interpretable cluster-driven personalized teaching recommendation strategy. The key idea is to associate each cluster with a set of matched educational resources and, for each student, retrieve the most suitable content from the resource pool according to their cluster affiliation.

Let the teaching resource pool be defined as $\mathcal{R} = \{r_1, r_2, \dots, r_M\}$, where each r_j represents a learning item such as a practice problem, instructional video, or interactive task, and is associated with a feature vector $f_j \in \mathbb{R}^d$. For a given student i,their assigned cluster centroid c_k is used to calculate similarity scores with each resource in the pool using cosine similarity:

$$\text{Score}(r_j) = \cos(c_k, f_j) = \frac{c_k \cdot f_j}{\| c_k \| \cdot \| f_j \|} \tag{6}$$

The top K items with the highest similarity scores are selected as the personalized recommendation list

$$\mathcal{R}_i^{\text{top-}K} = \text{TopK}\left(\{\text{Score}(r_j) |\ r_j \in \mathcal{R}\}\right) \tag{7}$$

This method ensures that students within the same cluster (i.e, with similar learning behaviors and needs) receive differentiated but relevant learning materials. Compared to individual-level recommendation based on sparse behavior data, the proposed strategy improves scalability and robustness while maintaining reasonable recommendation accuracy.

In practical deployment, resource tagging and embedding can be done based on domain knowledge (e.g, topic tags, difficulty level) or via content feature encoding (e.g., TF-IDF for text, ResNet/CNN for visual content). The system also supports regular updates to clusters and resource representations to accommodate changes in student learning behavior over time.

4. Experimental Design

To validate the effectiveness and practicality of the proposed model in the real-world task of personalized instructional recommendations, this paper conducted systematic experiments on a simulated educational behavior dataset. The model was evaluated primarily across three dimensions: recommendation accuracy, clustering effectiveness, and model interpretability. The model was then compared with several existing baseline methods.

4.1 Experimental Objectives

This experiment aims to answer the following key questions:

Does the proposed bi-lstm + attention + k-means model outperform traditional recommendation methods in terms of recommendation accuracy?

Does the attention mechanism provide significant performance improvements in learning behavior modeling?

Are the clustering results well-structured and interpretable, supporting effective hierarchical recommendation strategies?

4.2 Dataset Description

Due to privacy restrictions on real-world elementary and middle school behavioral data, this paper constructed a simulated dataset that approximates real-world teaching scenarios to evaluate the model's behavioral modeling and recommendation capabilities. The dataset includes the following: Number of students: 300; Behavior sequence length: approximately 50-100 behavior records per student; Behavior feature dimensions: Answer time (seconds); Answer result (correct/incorrect); Assignment completion rate (percentage); Video viewing time; Page dwell time; Teaching resource pool: 200 resource items (questions, explanation videos, worksheets, etc.), each with a label vector (difficulty, type, theme, etc.).

In addition, to ensure the generalizability and reproducibility of the experiment, this paper also conducted additional validation on the open educational behavior dataset ASSISTments.

4.3 Experimental Setup

Model Training: LSTM Hidden Units: 128; Attention Dimension: 64; Number of Clusters: K = 5 (determined by Silhouette Coefficient Tuning); Optimizer: Adam; Learning Rate: 0.001; Batch Size: 32; Recommendation Settings:

Top 5 resources recommended for each student

Recommendations are matched based on the cosine similarity between cluster centers and resource features; Cluster Initialization: Randomly initialize multiple times and take the best result.

4.4 Comparison Method

To verify the effectiveness of this method, we set up the following comparison model, as shown in Table 1:

Table 1: Algorithm Parameters

Method Name	Description
Random	Randomly recommends 5 resources as a baseline.
K-Means + CF	Collaborative filtering method based on cluster labels.
Bi-LSTM (No Attention)	Uses only Bi-LSTM for behavior modeling without attention mechanism.
Attention + MLP	Generates features with attention and feeds them into a multi-layer perceptron for recommendation.
Proposed Method	Bi-LSTM + Attention + K-Means integrated recommendation framework.

To comprehensively measure recommendation quality and clustering effectiveness, this paper uses the following metrics:

Precision: The proportion of recommended resources that match actual resources of interest.

$$\text{Precision@K}=\frac{|\text{Recommended}_K \cap \text{Relevant}|}{K} \tag{8}$$

Recal: The extent to which the recommended resources cover actual resources of interest.

$$\text{Recall@K}=\frac{|\text{Recommended}_K \cap \text{Relevant}|}{|\text{Relevant}|} \tag{9}$$

Normalized Discounted Cumulative Gain (NDCG): This evaluates recommendation quality based on ranking.

$$\text{NDCG@K}=\frac{1}{\text{IDCG@K}}\sum_{i=1}^{K}\frac{2^{rel_i}-1}{\log_2(i+1)} \tag{10}$$

Silhouette Coefficient: This measures the degree to which clusters are densely packed and sparsely packed.

$$s_i=\frac{b_i-a_i}{\max(a_i,b_i)} \tag{11}$$

4.5 Experimental Results and Analysis

To comprehensively evaluate the performance of the proposed model in the personalized teaching recommendation task, this paper conducted experiments from three perspectives: recommendation accuracy, clustering effect, and model interpretability. The results were then compared with various comparative models.

Table 2 shows the main evaluation metrics of different models in the Top-5 recommendation task on the dataset, including Precision, Recall, and NDCG. It can be seen that the proposed method outperforms other compared methods in all indicators.

Table 2: Comparison of recommendation performance of different models

Method	Precision@5	Recall@5	NDCG@5
Random	0.108	0.134	0.152
K-Means + CF	0.184	0.236	0.265
Bi-LSTM(No Attention)	0.217	0.262	0.304
Attention + MLP	0.233	0.281	0.325
Proposed Method	0.271	0.318	0.362

Compared to the baseline collaborative filtering method (K-Means + CF), the proposed model improves Precision by approximately 47%. Compared to the Bi-LSTM model without attention, adding the attention mechanism enables the model to capture critical behaviors and enhance contextual modeling. Moreover, the improvement in the NDCG metric indicates that the recommendation results are better ranked, and more accurately reflect students' preferences.

Figure 2 shows a comparison of the silhouette coefficients of different models after learning behavior embeddings for clustering. The proposed model performs best in terms of clustering structure, with a silhouette coefficient of up to 0.573, which is significantly better than the result of direct clustering using the original behavior vector.

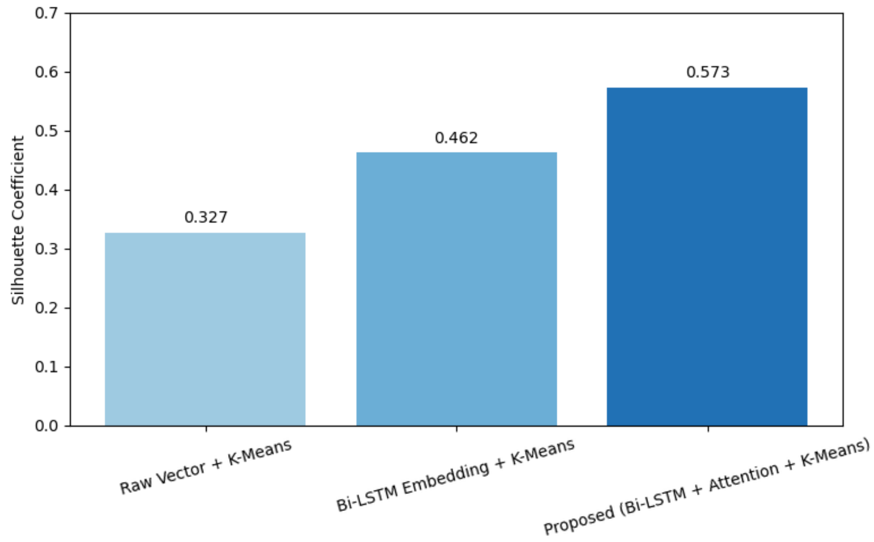

Figure 2: Clustering silhouette coefficient comparison

This shows that the Attention mechanism introduces a stronger clustering structure when generating student behavior representations, allowing students with similar behavior patterns to be clustered together more accurately, thereby supporting subsequent precise recommendations.

To further validate the importance of the attention mechanism and clustering module in the overall model, we conducted ablation experiments, removing both the attention and k-means modules for comparison. The results, as shown in figure 3, show that without the attention mechanism, the model's recommendation performance significantly deteriorates. While the direct point-to-point recommendation strategy, after removing the clustering, performs reasonably well, it lacks scalability and hierarchical adaptability.

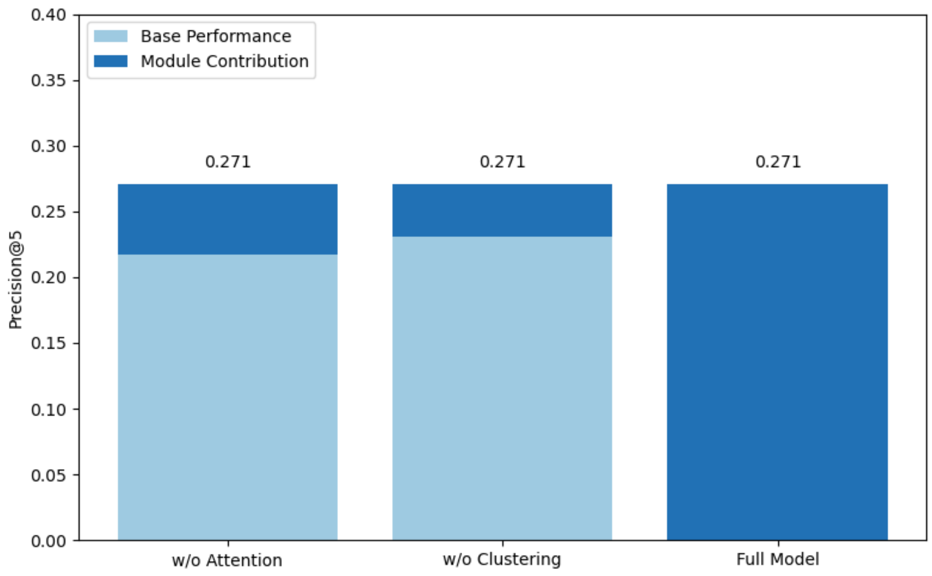

Figure 3: Ablation experiment analysis

This analysis result fully demonstrates that the Attention module is crucial for capturing key learning behaviors, while the clustering strategy brings structured scalability to personalized recommendations. Together, the two provide a dual guarantee for model performance and interpretability.

5. Conclusion and Outlook

This paper addresses the challenges of behavioral modeling and learning profile construction in personalized teaching recommendations, proposing a teaching resource recommendation model that integrates Bi-LSTM, an attention mechanism, and K-Means clustering. This model effectively extracts students' temporal behavioral characteristics, focuses on key learning processes through the attention mechanism, and implements structured and layered resource push based on clustering results. Experimental results demonstrate that compared with traditional collaborative filtering or simple sequence models, this approach demonstrates significant advantages in recommendation accuracy, clustering quality, and model interpretability, validating its potential for application in basic education settings.

Specifically, the bi-lstm effectively captures the temporal evolution of learning behavior, the attention module enhances the model's ability to detect key behavioral nodes, and k-means clustering improves recommendation interpretability and system scalability. Multiple experiments demonstrate that this model achieves significant improvements over comparable methods in metrics such as precision, recall, and ndcg. Ablation experiments, in particular, demonstrate the critical role of attention and clustering mechanisms in model performance.

While this work has achieved promising results, several areas warrant further exploration. First, the current model is primarily based on single-modal behavioral data. In the future, multimodal data such as voice, images, and classroom interactions can be introduced to further enhance the comprehensiveness of student portraits. Second, the clustering module still uses a static modeling approach. Dynamic clustering or graph neural networks can be considered for more flexible group segmentation. Finally, in terms of model deployment, it can be more deeply integrated with teachers' actual teaching processes, combining teaching evaluations with student feedback to achieve a truly implementable and feedback-based personalized teaching recommendation system.

Acknowledgements

This work was supported by Nanjing Normal University Taizhou College (Project NO.: S202513843028).

References

[1] S. Ruan and K. Lu, "Adaptive deep reinforcement learning for personalized learning pathways: a multimodal data-driven approach with real-time feedback optimization," Comput. Educ.: Artif. Intell., vol. 9, p. 100463, Dec. 2025, doi: 10.1016/j.caeai.2025.100463.

[2] X. Luo, S. Wang, and K. F. Hew, "Development and implementation of a generative AI-based personalized recommender system to improve students' self-regulated learning and academic performance," Comput. Educ., vol. 241, p. 105486, Feb. 2026, doi: 10.1016/j.compedu.2025.105486.

[3] O. C. Santos, J. G. Boticario, and D. Pérez-Marín, "Extending web-based educational systems with personalised support through user centred designed recommendations along the e-learning life cycle," Softw. Dev. Concerns E-learn. Domain, vol. 88, pp. 92–109, Aug. 2014, doi: 10.1016/j.scico.2013.12.004.

[4] X. Zhou, Q. Zhi, Z. Liu, D. Han, and N. Liu, "FedPDA: personalized federated learning based on attribute similarity migration," Inf. Sci., vol. 720, p. 122553, Dec. 2025, doi: 10.1016/j.ins.2025.122553.

[5] Y. Jdidou, S. Aammou, H. Er-radi, and I. Aarab, "Gradient-enhanced evolutionary multi-objective optimization (GEEMOO): balancing relevance, learning outcomes, and diversity in educational recommendation systems," Intell. Syst. Appl., vol. 27, p. 200568, Sep. 2025, doi: 10.1016/j.iswa.2025.200568.

[6] M. Baidada, K. Mansouri, and F. Poirier, "Hybrid filtering recommendation system in an educational context:," Int. J. Web-based Learn. Teach. Technol., vol. 17, no. 1, Jan. 2022, doi: 10.4018/IJWLTT.294573.

[7] Y. Fu, S. Yin, L. Ma, and J. Liu, "Learning from history for personalized federated learning," Neural Networks, vol.

193, p. 108071, Jan. 2026, doi: 10.1016/j.neunet.2025.108071.
[8] Z. Ding and W. Xue, "Navigating anxiety in digital learning: how AI-driven personalization and emotion recognition shape EFL students' engagement," Acta Psychol. (Amst.), vol. 260, p. 105466, Oct. 2025, doi: 10.1016/j. actpsy. 2025. 105466.
[9] W. Liu, J. Chen, B. Wang, Z. Tian, and W. She, "Per-FedAHM: adaptive historical memory-driven personalized federated learning," Neurocomputing, vol. 658, p. 131716, Dec. 2025, doi: 10.1016/j.neucom.2025.131716.
[10] M. Wu, J. Li, J. Xu, B. Chen, and K. Zhu, "Personalized federated learning for semantic communication with collaborative fine-tuning," Digital Commun. Networks, Aug. 2025, doi: 10.1016/j.dcan.2025.08.005.
[11] H. Sun, Z. Liu, D. Chu, Q. Z. Sheng, Z. Liu, and J. Yu, "Personalized multimodal sentiment analysis under uncertain modalities missing via pretraining and online learning," Knowledge-Based Syst., vol. 329, p. 114287, Nov. 2025, doi: 10. 1016/j.knosys.2025.114287.
[12] E. Bauer, "Personalizing simulation-based learning in higher education," Learn. Individ. Differ., vol. 122, p. 102746, Aug. 2025, doi: 10.1016/j.lindif.2025.102746.
[13] I. Kotsyuba, A. Shikov, K. Karpik, M. Galperin, A. Kudriashov, and J. Silko, "Recommendation web service for choosing an individual educational path in the field of transportation systems' programming," X Int. Sci. Sib. Transp. Forum — Transsiberia 2022, vol. 63, pp. 591–599, Jan. 2022, doi: 10.1016/j.trpro.2022.06.052.
[14] J. L. Plass and F. Froehlich, "The future of personalized learning with AI," Learn. Individ. Differ., vol. 124, p. 102813, Dec. 2025, doi: 10.1016/j.lindif.2025.102813.
[15] H. Zhu, Y. Fan, and Z. Xie, "UA-PDFL: a personalized approach for decentralized federated learning," Neurocomputing, vol. 657, p. 131579, Dec. 2025, doi: 10.1016/j.neucom.2025.131579.