ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Research on the Application of Deep Learning in College English Personalized Learning Path

Yuanyuan Xue[1,*]

[1] Jilin Animation Institute, Changchun, Jilin, 130012 China

SUMMARY: *With the rapid development of digital education, college English teaching has put forward higher requirements for the accurate design and dynamic adaptation of learning paths. Focusing on the path recommendation task driven by multi-source learning data, this paper constructs a technical framework covering data collection and preprocessing, learner portrait modeling, deep feature extraction, path generation, resource matching, dynamic adjustment and teaching feedback closed loop. Intelligent recommendation and continuous update of learning paths are realized. The experimental results show that the Precision, Recall and F1-score of the proposed method reach 0.846, 0.821 and 0.833, the Top-10 hit rate reaches 0.862, the path generation accuracy and knowledge coverage rate reach 0.852 and 0.874, respectively. After feedback-driven adjustment, the task accuracy is increased to 84.9%, and the path deviation index is reduced to 0.121. The research shows that this method can improve the accuracy, coherence and dynamic adaptation ability of college English learning path, and provide technical support for the intelligent transformation and accurate implementation of college English teaching.*

KEYWORDS: *Deep learning; College English; Learning path recommendation; Multi-source learning data*

1 Introduction

With the continuous entry of digital technology into the field of higher education, the organizational mode, resource form and learning evaluation mechanism of college English teaching have changed significantly. Compared with the traditional teaching mode based on classroom teaching and unified progress promotion, current college English learning emphasizes on difference identification, process tracking and ability orientation, and requires that teaching activities can respond to students' individual differences in language foundation, cognitive style, learning interest, task completion rhythm and target needs. We note that in real teaching, there are often some phenomena within the same classroom, such as a wide range of starting point levels, uneven learning investment, low efficiency of resource use, and convergence of learning paths. Although teachers can rely on experience for hierarchical guidance to a certain extent, in the face of large-scale, multi-dimensional and continuously changing learning data, it is difficult to rely on manual judgment alone to achieve accurate, dynamic and sustainable personalized support.

The development of deep learning technology provides new research ideas for the construction of college English personalized learning paths. Compared with traditional rule matching or static recommendation methods, deep learning can extract high-level features

*XueYuanYuan2026@163.com
https://doi.org/10.65102/is2026312 Published: 30 April, 2026

from multi-source data such as learning behavior records, evaluation results, resource access trajectories, assignment performance and interactive feedback, and mine the potential relationship between learners 'ability structure, knowledge mastery status and learning preferences. On this basis, a more adaptive resource recommendation and path generation scheme is formed. We can see that in the context of college English teaching, the ability dimensions of vocabulary, grammar, reading, writing, listening and speaking are interrelated, and the learning process is continuous, phasic and different. If learner portrait modeling and path intelligent adjustment can be realized with the help of deep learning, it will not only help to improve the efficiency of resource allocation, but also help to enhance the pertinence and sustainability of students 'learning.

However, from the existing research, college English personalized learning research focuses more on platform construction, hierarchical teaching or general recommendation, and the integrated research on multi-source learning data collaborative processing, learners' deep feature extraction, learning path dynamic generation and feedback closed-loop optimization is still insufficient. Based on this, we explore the application mechanism of deep learning in personalized learning path design around college English teaching scenarios, construct a technical framework covering data collection, feature extraction, path generation, resource matching and feedback regulation, and analyze its application effect combined with experimental results, expecting to provide reference for the intelligent transformation and precise implementation of college English teaching.

2 Related work

On the college English personalized learning path, the existing research has generally experienced a process from resource recommendation to path planning, and then to dynamic optimization. Huang and Zhu built an English learning material recommendation system based on knowledge graph, which represented learning resources, knowledge points and learning needs, indicating that semantic relationship modeling could improve the pertinence of English material recommendation, but its focus was still on the matching of resources, and the description of continuous learning sequence was relatively limited [1]. van der Velde et al. further started from the cold start problem in adaptive learning and proposed to use data-driven difficulty estimation to alleviate the recommendation bias caused by insufficient portraits at the early stage of learning, indicating that early ability judgment plays a fundamental role in subsequent personalized support [2]. Wang and Huang applied deep learning to the college English writing evaluation system to provide feedback for teaching through automatic analysis of writing performance, expanding the application boundary of deep models in English learning diagnosis, but the research focus was still on single ability evaluation rather than the generation of complete learning paths [3].

In terms of learning path planning, Jiang et al. proposed a data-driven personalized path planning method based on the cognitive diagnosis results in MOOCs, which combined the learning state, the difficulty of knowledge points and the mastery degree to generate a more appropriate learning sequence, and promoted the path recommendation from static experience judgment to state perception driven [4]. Li et al. integrated deep learning and big data technologies to carry out personalized recommendation in MOOC scenarios, emphasizing the important value of large-scale learning behavior data for user interest recognition and course recommendation accuracy [5]. Zhang et al. proposed a fine-grained and multi-context-aware knowledge graph learning path recommendation model, which can capture the dynamic characteristics and potential knowledge level of learners from online learning communities, and improve the context adaptation of path recommendation [6]. Another study combined knowledge graph with graph convolutional network to deal with the problem of learning resource association and user preference propagation, indicating that graph neural network method has strong potential in learning path modeling [7].

Recent studies pay more attention to the cognitive interpretation and dynamic decision-making ability of path recommendation. Mu and Yuan introduced cognitive map to present learners' knowledge structure and cognitive state, so that path recommendation emphasized knowledge mastery logic and self-reflection support [8]. Tzeng et al. built a MOOC learning material recommendation system based on LSTM, used video viewing behavior to identify similar learning patterns, and recommended learning paths based on them, showing the advantages of sequence model in describing the evolution of learning process [9]. Ma et al. further adopted multi-behavior user modeling and cascaded deep Q-network to introduce reinforcement learning into the path recommendation process, so that the system could simultaneously take into account short-term behavior changes and long-term learning objectives [10]. In general, the existing results have provided various ideas for the research of personalized learning path, such as knowledge graph, deep sequence modeling and reinforcement learning. However, there is still room for further expansion of multi-source data collaborative processing for college English scenarios, deep feature extraction of learner profiles, and the integrated design of "path generation-feedback adjustment", which also constitutes the entry point of this paper.

3 The application design of deep learning in College English personalized learning Path

3.1 Learner characteristic data composition and personalized learning problem description

The process of college English learning generates various kinds of traceable data, including stage test scores, writing performance, resource access records, learning duration, task completion and interactive feedback. These data are scattered in different teaching links, which not only contain students' current language ability information, but also reflect their learning engagement, resource use preference and stage goal differences.

From the perspective of teaching application, learner characteristics are dynamic sets composed of multiple elements. We summarize them into four aspects: first, basic ability characteristics, which mainly correspond to language ability indicators such as vocabulary mastery, reading comprehension, writing expression and listening and speaking performance. The second is the behavior participation characteristics, which mainly reflects the process information such as learning duration, resource access frequency, task completion rate and stage activity. The third is the resource preference feature, which is used to characterize students 'different acceptance tendencies for video explanations, reading materials, exercises or interactive tasks. The fourth is the characteristics of learning objectives, which mainly reflects the differentiated demands of students in terms of exam attainment, ability improvement and special breakthrough. The above characteristics jointly determine the real learning state of learners in the teaching platform, and also make the personalized path design have the basis of data description.

For ease of calculation and modeling, let the original features of the u-th learner be represented as follows:

$$x_u=[x_{u1},x_{u2},\ldots,x_{um}]^{\top} \tag{1}$$

Here, m represents the total number of feature dimensions and x_{um} represents the observed value of the learner on the JTH dimension. The vector can uniformly describe the comprehensive state of learners at a certain stage. However, due to the obvious differences in the numerical range and dimension of different features, direct input into the model will affect the subsequent feature extraction effect, so it needs to be standardized first. We use the range normalization method to map the features of each dimension to a unified interval, and the calculation formula is as follows:

$$\bar{x}_{uj}=\frac{x_{uj}-x_j^{min}}{x_j^{max}-x_j^{min}+\varepsilon} \tag{2}$$

Here, x_j^{min} and x_j^{max} denote the minimum and maximum value of the JTH dimension feature in the sample, respectively, and ε is a minimal constant. After this process, data from different sources are comparable, and more stable input conditions are provided for subsequent deep representation learning.

The key task of personalized learning path design is to select a set of learning content with reasonable order, gradual difficulty and resource adaptation according to the difference between the learner's current state and the goal state. Let the learner's current ability be represented as a_u and the target ability be represented as q_u, then its ability gap can be defined as follows:

$$g_u=\max(0,q_u-a_u) \tag{3}$$

This vector is used to describe the learner's degree to be improved in each ability dimension. If a dimension has met the expected requirements, the gap is recorded as 0; If the target level is not reached, the larger the difference is, the higher the priority of this capability dimension in path generation.

Based on this, let the candidate path of the learner be $P_u=\{r_1,r_2,\ldots,r_T\}$, where rt denotes the t-th learning resource or learning task. Path generation needs to simultaneously consider the matching degree between resources and capacity gaps, the connection relationship between adjacent learning contents, and the overall learning cost. Therefore, the problem can be expressed as follows.

$$P_u^*=\arg\max_{P_u}\left(\alpha\sum_{t=1}^{T}\mathrm{Rel}(r_t,g_u)+\beta\sum_{t=2}^{T}\mathrm{Con}(r_{t-1},r_t)-\gamma\sum_{t=1}^{T}c(r_t)\right) \tag{4}$$

Among them, $\mathrm{Rel}(r_t,g_u)$ represents the matching degree between resources and learners' ability gap, $\mathrm{Con}(r_{t-1},r_t)$ represents the knowledge cohesion strength between adjacent resources, $c(r_t)$ represents the time or cognitive cost brought by learning resources, α, β, γ are weight coefficients. Taken together, the problem of college English personalized learning can be viewed as a sequential decision-making process with multi-source learning features as input, ability gap identification as intermediary, and path optimization as output.

3.2 College English multi-source learning data collection and preprocessing mechanism

The generation of college English personalized learning path is inseparable from the continuous, complete and unified structure of data support. The data sources of this study mainly include four categories: first, platform log data, such as login frequency, learning duration, resource click, task submission time, etc. The second is outcome data, such as stage test scores, vocabulary breakthrough results, reading exercise scores and writing evaluation results. The third is the process text data, such as homework notes, learning reflection, question and answer interactive content. Fourth, questionnaire and feedback data are used to supplement information such as learning objectives, resource preferences and self-perceived difficulties. Data records from different sources have different granularity, different time scales, and different field formats. Therefore, it is necessary to complete unified access, cleaning transformation, and feature alignment before entering the model.

In the data collection stage, learner number, course number and timestamp are used as three types of core indexes to associate heterogeneous data. For behavior stream data, we use time window aggregation to transform discrete records into stage features. Suppose that the observed value of the KTH behavior of learner u in time window τ is yu,k,t, then the behavior statistics of this stage can be expressed as follows:

$$b_{u,k}^{(\tau)} = \frac{1}{|\Omega_\tau} \sum_{t \in \Omega_\tau} y_{u,k,t} \tag{5}$$

Here, Ω_τ denotes the set of records within the time window τ. Figure 1 shows the multi-source data processing flow. The system first completed the original data access and identity mapping, and then carried out outlier elimination, missing completion, time alignment, text coding and feature fusion, and finally formed a unified learner input matrix.

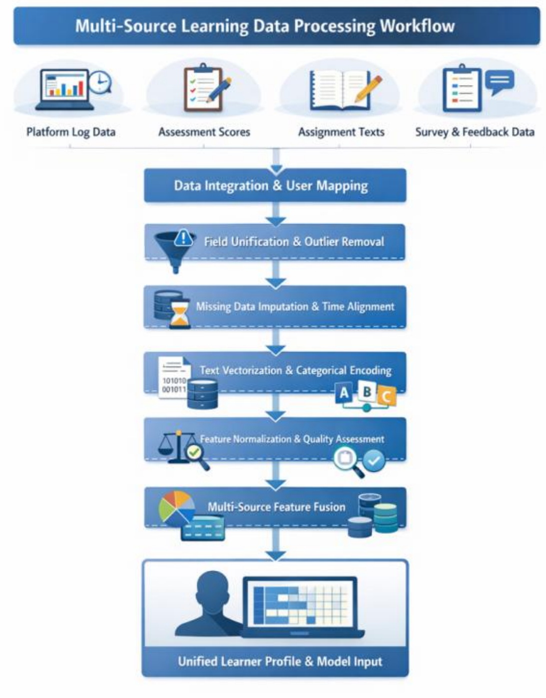

Figure 1: Flowchart of data processing for College English multi-source learning

The preprocessing stage focuses on solving the problems of missing, noise and scale inconsistency. Considering that learning platform data is susceptible to sporadic behavioral fluctuations, this paper uses a robust standardization method based on median and median absolute difference to transform continuous variables:

$$\tilde{b}_{u,k}^{(\tau)}=\frac{b_{u,k}^{(\tau)}-Med(b_k)}{MAD(b_k)+\delta} \tag{6}$$

where $Med(b_k)$ is the sample median of the KTH class feature, $MAD(b_k)$ is the absolute median difference, and δ is the smoothing term. This method is more stable for skewed data and outliers, and is more suitable for behavioral data with large fluctuations in educational scenarios.

For missing records, we combined learner historical performance with neighboring group characteristics for imputation. If the estimated value after a certain feature is missing is $\bar{b}_{u,k}^{(\tau)}$, then:

$$\bar{b}_{u,k}^{(\tau)}=\lambda b_{u,k}^{his}+(1-\lambda)\frac{1}{|N_u|}\sum_{v\in N_u} b_{v,k}^{(\tau)} \tag{7}$$

where $b_{u,k}^{his}$ represent the individual historical mean of learners, N_u represents the set of neighborhood samples with similar ability level to learner u, and λ is the weight coefficient.

After encoding textual and categorical data, they need to enter a unified representation space together with numerical features. In order to improve the fusion quality of data from different sources, this paper introduces a weighted integration strategy based on data quality. Let the representation of the MTH data source after preprocessing be su(m), then the learner's final input vector is written as follows:

$$d_u=\sum_{m=1}^{M}\omega_m\,\psi_m(s_u^{(m)}),\quad \omega_m=\frac{q_m}{\sum_{j=1}^{M} q_j} \tag{8}$$

Here, $\psi_m(\cdot)$ represents the mapping function of different data sources, q_m represents the integrity and credibility score of the data source, and ω_m is the normalized fusion weight. Through this method, the platform behavior, test results, text feedback and target information are unified into a computable multi-source feature representation, which provides a stable input for subsequent learner profile modeling and path generation.

3.3 Deep feature Extraction Method for Learner Portrait Modeling

After completing the multi-source data access and preprocessing, the research focus turned to the modeling of learner portraits. For college English learning, using test scores or platform activity alone can only reflect part of the learning state, and it is difficult to accurately describe the comprehensive differences in students 'language foundation, learning preference, behavior rhythm and ability improvement direction. To this end, we define the learner profile as a low-dimensional semantic representation composed of static attribute features and dynamic behavior features, which is used to support subsequent learning path generation and resource matching.

In the modeling process, the static features mainly include vocabulary level, reading achievement, writing performance, resource preference and stage goal, which are relatively stable and suitable for characterizing the long-term attributes of learners. Let the MTH static input be $d_u^{(m)}$, and its corresponding deep representation is obtained after linear mapping and nonlinear activation:

$$h_u^{(m)}=\phi\left(W^{(m)}d_u^{(m)}+b^{(m)}\right) \tag{9}$$

Here, $W^{(m)}$ and $b^{(m)}$ denote the transformation parameters of the m-th feature, respectively, and $\phi(\cdot)$ denotes the activation function. After this step, the original features of different dimensions and types are projected into a unified representation space to facilitate subsequent fusion.

Compared with static features, the learned behavior is more temporal. Students' resource click, task completion, repeated practice and feedback correction in a period of time often imply a more obvious track of ability change. Therefore, we denote the sequence of actions as $\{e_{u,t}\}_{t=1}^{T}$ and extract the phase state features in a recursive manner:

$$s_{u,t}=\tanh\left(Ae_{u,t}+Bs_{u,t-1}+c\right) \tag{10}$$

Here, $s_{u,t}$ denotes the hidden state of the learner at time t, and A, B, and c are trainable parameters. This formula can compress the scattered learning behavior into a continuous state representation, so that the model can observe the overall trend in the process of behavior evolution.

Considering that the importance of different time slices is not the same, this paper further introduces temporal attention weights to highlight key learning segments. Let the importance weight of the behavioral state $s_{u,t}$ be $\alpha_{u,t}$, then we have:

$$\alpha_{u,t}=\frac{\exp\left(q^{\top}s_{u,t}\right)}{\sum_{\tau=1}^{T}\exp\left(q^{\top}s_{u,\tau}\right)},r_u=\sum_{t=1}^{T}\alpha_{u,t}\,s_{u,t} \tag{11}$$

where q is the learnable query vector and ru represents the dynamic representation extracted based on the behavior sequence. Through this processing, the key segments such as task focused completion period, frequent review period and performance fluctuation period can obtain higher weights in the portrait modeling, thereby enhancing the discrimination of feature expression.

After the static attributes and dynamic behaviors are encoded separately, cross-source fusion needs to be completed to form a unified learner profile. In this paper, cross-modal weight allocation mechanism is used to combine features from different sources, and the final learner profile vector is expressed as follows:

$$p_u=\mathrm{LayerNorm}\left(r_u+\sum_{m=1}^{M}\beta_u^{(m)}\,h_u^{(m)}\right) \tag{12}$$

where p_u is the learner profile vector, $\beta_u^{(m)}$ represents the contribution coefficient of the m-th static feature on the current learner. It is calculated as follows:

$$\beta_u^{(m)}=\frac{\exp\left(w^{\top}h_u^{(m)}\right)}{\sum_{j=1}^{M}\exp\left(w^{\top}h_u^{(j)}\right)} \tag{13}$$

Here, w is the weight evaluation vector. According to the actual situation of different learners, the system can adaptively adjust the influence degree of information such as vocabulary score, behavior activity, resource preference and goal demand. After deep feature extraction, the learner portrait is no longer a simple result of index splicing, but a comprehensive representation that retains long-term attributes, stage states and key behavioral signals at the same time. The portrait vector will be used as the core input of the subsequent personalized learning path generation model, making the path design closer to the real needs of learners, and also providing a more stable basis for resource matching and dynamic adjustment. The learner portrait modeling and deep feature extraction are shown in Figure 2.

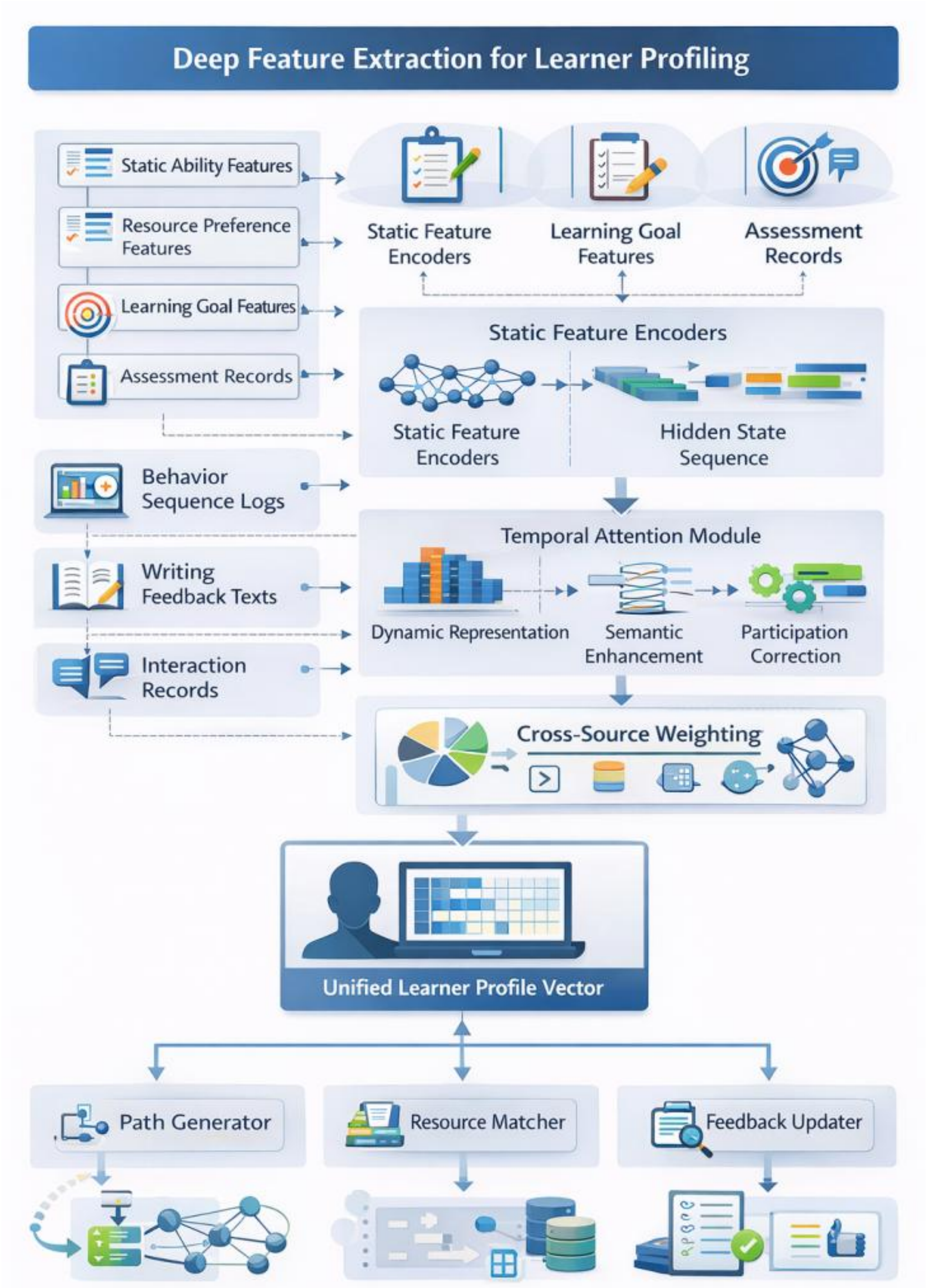

Figure 2: Schematic diagram of learner portrait modeling and deep feature extraction

3.4 College English Personalized Learning Path Generation Model based on Deep learning

After completing the learner profile modeling, the system needs to further answer the core question of "what to recommend to whom and in what order". For example, vocabulary accumulation will affect reading comprehension, and reading input will have a continuous effect on writing output. Therefore, the generation of learning paths should take into account knowledge dependence, difficulty progression and consistency of stage goals.

As shown in Figure 3, the personalized learning path generation model we constructed consists of the learner profile input layer, the candidate resource representation layer, the path matching network, the sequence decision-making layer and the dynamic update interface. The model first screened the candidate set from the resource library according to the learner portrait, and then completed the path sorting by combining the resource knowledge association, historical learning progress and stage goal, and finally output the learning sequence suitable for the current learning state. It should be noted that although the model in this paper absorbs the resource correlation information provided by the knowledge graph, it is not equivalent to the "knowledge graph based method" in the comparative experiment. The latter mainly uses the semantic relationship between resources and knowledge points for static path ranking. On this basis, the proposed method further integrates the deep representation of learner portrait, resource difficulty constraint, path sequence optimization, and feedback-driven dynamic update mechanism, so it emphasizes the ability of continuous path generation under learner state awareness.

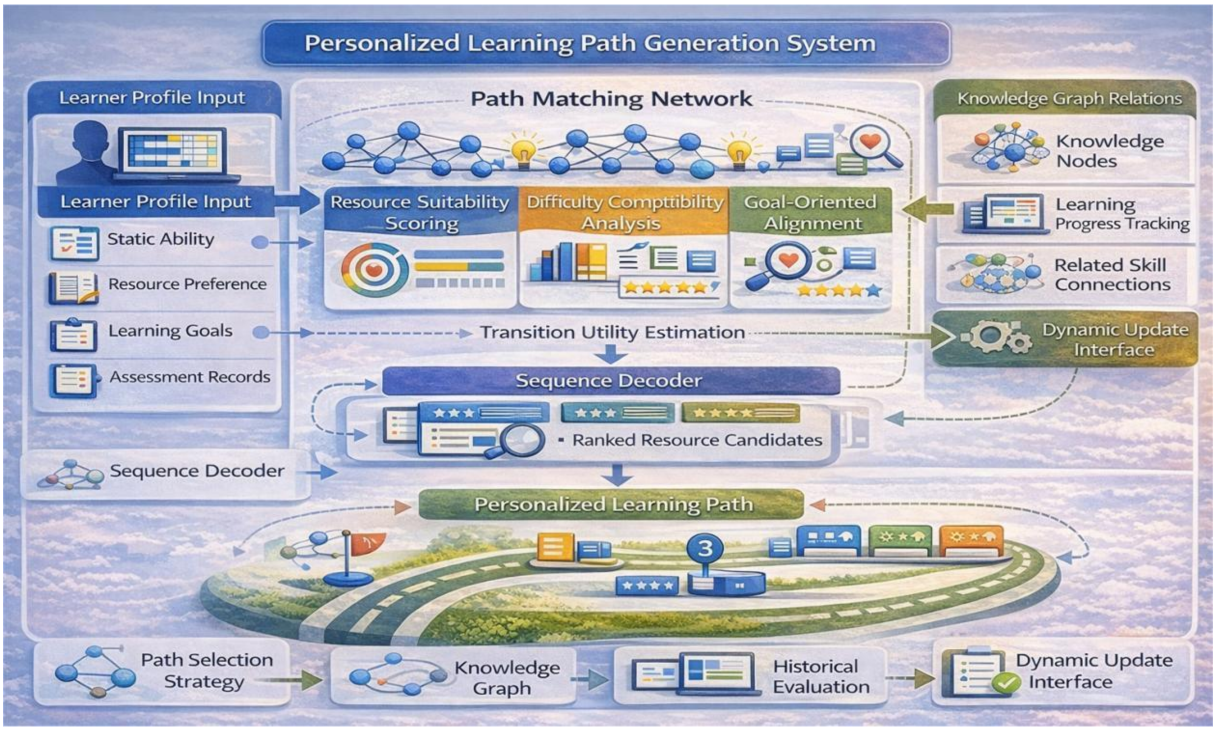

Figure 3: Overall architecture diagram of personalized learning path generation

Let the learner profile vector be pu, the embedding representation of candidate resource ri be vi, and the difficulty of resource be di, then the adaptation score between learner and resource is defined as follows:

$$s_{u,i}=\lambda_1 \frac{p_u^T v_i}{\|p_u\|\|v_i\|}+\lambda_2 \exp\left(-|g_u-d_i|\right)+\lambda_3\chi_i \tag{14}$$

Here, g_u represents the target difficulty level of the learner at the current stage, χ_i represents the consistency coefficient between the resource and the target ability dimension, and λ_1,λ_2, and λ_3 are the balance parameters. This formula considers the semantic similarity of the portrait, the degree of difficulty fitting and the degree of target matching at the same time, avoiding excessive dependence on a certain type of information for path generation.

After the score of a single resource is determined, it is also necessary to judge whether a smooth connection can be formed between adjacent resources. To this end, we introduce the path transfer payoff function. Given that the knowledge association strength between resources r_i and r_j is aij, the difficulty difference between them is $|d_i-d_j|$, and the learner's individual preference weight for r_j is $\pi_{u,j}$, then the transfer benefit is written as follows:

$$\delta_u(i,j)=\mu_1 a_{ij}+\mu_2 \frac{1}{1+|d_i-d_j|}+\mu_3\pi_{u,j} \tag{15}$$

Here,μ_1,μ_2,μ_3 are the weight parameters. This index reflects the degree of coherence between the two learning nodes before and after in the path.

In the path decoding stage, this paper uses a stepwise cumulative sequence optimization to generate a personalized learning path of length T. Let the optimal cumulative utility at step t ending with resource rj be $F_t(j)$, then we have:

$$F_t(j)=\max_{i\in C_{t-1}}\left[F_{t-1}(i)+\delta_u(i,j)\right]+s_{u,j} \tag{16}$$

Here, C_{t-1} represents the set of candidate resources at step t−1. Through this recursive process, the model can comprehensively evaluate the effect of resource matching and sequence cohesion in a global perspective, and finally backtrack the obtained sequence $P_u^*=\{r_1,r_2,\ldots,r_T\}$ is the recommended path of the learner at the current stage.

In order to improve the sorting discrimination ability of the model training, we denote the positive sample resources in the real learning sequence as ri+, and the non-optimal resources obtained by random sampling as ri−, and construct the margin constraint loss function:

$$L=\sum_u\sum_i \max\left(0,\zeta-s_{u,i}^{+}+s_{u,i}^{-}\right)+\eta\|\Theta\|_2^2 \tag{17}$$

Here, L is the safety interval, Θ represents the set of model parameters, and η is the coefficient of the regularization term. The loss can promote the score of positive sample resources to be continuously higher than that of negative sample resources, making the model easier to learn a stable path ordering law.

The value of this model lies in putting the learner profile, resource characteristics, knowledge relationship and path sequence into the same framework. For college English teaching, this way of path generation is more in line with the actual needs of the gradual development of ability, and also provides a clear interface for the subsequent optimization and dynamic adjustment of resource matching.

3.5 Learning resource matching and path dynamic adjustment mechanism design

Learning resource matching is expanded based on the resource description vector. For any resource ri, its attributes not only include the coverage of knowledge points, but also include the difficulty level, resource form, expected learning time and target ability dimension. Let the support strength of resources on the CTH ability dimension be $q_{i,c}$, and learners' current demand for improvement on this dimension be $g_{u,c}^{(t)}$, then the coverage degree of resources on demand can be expressed as follows:

$$\Gamma_{u,i}^{(t)}=\sum_{c=1}^{C} \omega_c \min\left(g_{u,c}^{(t)},q_{i,c}\right) \tag{18}$$

Here, C is the number of capability dimensions and ω_c is the dimension weight. This formula can directly reflect the degree of fit between resources and learning needs.

Capacity coverage alone is not enough to ensure that the resources are truly applicable, and the resource difficulty and learning load also affect the path execution effect. Set the resource difficulty as l_i, the learner's current acceptable target difficulty center as $\bar{l}_u^{(t)}$, the resource learning load as h_i, and the learner stage load threshold as $H_u^{(t)}$, then the comprehensive matching score of resources is defined as follows:

$$\Psi_{u,i}^{(t)}=\Gamma_{u,i}^{(t)}\cdot \exp\left[-\tau\left|l_i-\bar{l}_u^{(t)}\right|\right]\cdot\left(1-\frac{\max\left(0,h_i-H_u^{(t)}\right)}{H_u^{(t)}+\epsilon}\right) \tag{19}$$

Here, τ is the difficulty deviation penalty coefficient and ϵ is the smoothing term. The system will not simply give priority to the resources with the "most coverage capacity", but control the difficulty jump and overload problems at the same time, so that the resource allocation is closer to the real learning rhythm.

After the path enters the execution phase, the learner feedback needs to flow back into the system continuously. In this paper, the stage feedback is divided into three categories: accuracy rate, completion efficiency and interaction participation, and the learner's mastery status is updated accordingly. Let the mastery estimate of the CTH competence dimension after the TTH learning round be $m_{u,c}^{(t)}$, then the update rule is written as follows:

$$m_{u,c}^{(t+1)}=(1-\rho)m_{u,c}^{(t)}+\rho\Phi_{u,c}^{(t)} \tag{20}$$

where ρ is the update rate and $\Phi_{u,c}^{(t)}$ represents the capability performance value after this round of feedback mapping. This formula preserves the historical state and absorbs the latest learning performance, so that the portrait and path can gradually evolve with the learning process.

On this basis, the system makes shunting adjustments to the path according to the updated mastery level. Let the overall learning state of the current node be $z_u^{(t)}$, then the path adjustment strategy is defined as follows:

$$A_u^{(t)}=\begin{cases}\text{Advance}, & z_u^{(t)}\geq\theta_2,\\ \text{Reinforce}, & \theta_1\leq z_u^{(t)}<\theta_2,\\ \text{Remediate}, & z_u^{(t)}<\theta_1,\end{cases} \tag{21}$$

Here, θ_1 and θ_2 are the remedy threshold and the progression threshold, respectively. When the learning state reached a high level, the system allowed the path to advance to higher difficulty nodes. In the middle interval, the main path is maintained and the consolidation resources are added. If the degree of mastery is obviously insufficient, it will fall back to the previous knowledge point or insert a compensatory task.

In summary, the mechanism in this section completes two key tasks. First, the learning resources are transformed from "recommended objects" to "computable matching units", which makes the resource selection more targeted. Secondly, the learning feedback is embedded in the path updating process, so that the personalized learning path has the ability of continuous modification.

3.6 Application process of personalized learning path and construction of teaching feedback mechanism

After the generation of personalized learning path, it also needs to enter the stage of executable, traceable and reflowable teaching application. College English teaching has the characteristics of clear task cycle and parallel advancement of classroom activities and platform learning. Therefore, path application should be further implemented in resource allocation, learning promotion, process monitoring, feedback aggregation and teacher intervention. We divide the application process into five parts: initial diagnosis, path release, task execution, feedback collection and path callback, and construct the corresponding teaching feedback mechanism, so that the recommendation results can really enter the teaching implementation process.

In the path release phase, the system releases subsequent tasks in batches according to the current node status. Let the task release priority value of learner u on the KTH path node be $o_{u,k}$, the node readiness be $\eta_{u,k}$, the completion quality of the previous node be $c_{u,k-1}$, and the urgency of the learning goal be $r_{u,k}$, then the release weight of the current node can be expressed as follows:

$$o_{u,k}=\frac{\exp\left(\nu_1\eta_{u,k}+\nu_2 c_{u,k-1}+\nu_3 r_{u,k}\right)}{\sum_{j\in P_u}\exp\left(\nu_1\eta_{u,j}+\nu_2 c_{u,j-1}+\nu_3 r_{u,j}\right)} \tag{22}$$

Here, Pu represents the set of nodes currently accessible to the learner, and ν1,ν2,ν3 are the regulation parameters. This formula reflects the stage control idea in path execution: the nodes with higher readiness, more stable completion of preorder tasks, and more urgent goals will enter the actual teaching process earlier.

After the learning task is performed, the system needs to extract feedback information from the platform records and teaching interactions. Considering that college English learning involves multiple dimensions such as accuracy, completion efficiency, classroom participation and teacher evaluation, this paper uses the fusion scoring method to form the node feedback value. Let the feedback strength of the KND node be $f_{u,k}$, where $a_{u,k}$ represent the task accuracy rate, $t_{u,k}$ represent the standardized time efficiency, $e_{u,k}$ represent the interactive participation level, $q_{u,k}$ represent the teacher process evaluation, then there is

$$f_{u,k}=\omega_1 a_{u,k}+\omega_2 \exp(-t_{u,k})+\omega_3 \ln(1+e_{u,k})+\omega_4 q_{u,k} \tag{23}$$

Here,$\omega_1,\omega_2,\omega_3,\omega_4$ are the feedback fusion weights. After the feedback enters the system, it is necessary to determine whether the current path should be maintained, locally corrected or globally reconstructed. To this end, the path deviation index is introduced in this paper. Let the expected behavior of the system for the KTH node be $\hat{f}_{u,k}$, then the degree of path deviation is defined as follows:

$$d_{u,k}=\frac{|f_{u,k}-\hat{f}_{u,k}|}{1+\hat{f}_{u,k}} \tag{24}$$

The larger this index is, the more obvious is the gap between the learner's state in actual execution and the model preset, and the weaker is the adaptation of the original path to the current learning situation. When $d_{u,k}$ is in the lower range, the system maintains the original main path. When it reaches the medium range, insert supplementary exercises, micro videos, or explanation resources. If the deviation continues to expand, the teacher intervention is triggered and the subsequent node order is recalculated.

The key of the teaching feedback mechanism is to let the teacher and the system participate in path adjustment together. The platform generates learning reports according to node feedback, and teachers combine classroom observation results to determine whether students have problems in knowledge understanding, task execution or learning engagement, and return correction labels to the system.

4 Results and discussion

4.1 Experimental Environment and Data set Construction

In order to verify the applicability of the proposed model in the generation of college English personalized learning paths, the experimental data are collected from the teaching records of a whole semester of a college English smart teaching platform. The data covered many links such as vocabulary learning, reading training, writing tasks, stage evaluation, classroom interaction and teacher feedback, which could completely reflect learners 'ability status and behavior changes in the platform environment. After anonymizing the original data, the learning behavior, academic performance, resource access and feedback records directly related to path modeling are retained, and the field cleaning, time alignment and outlier removal are completed according to the above Settings. To ensure the comparability of the experimental results, we divided the path sequence samples into training set, validation set and test set according to the ratio of 7:1.5:1.5, of which 2397 were in the training set, 513 were in the validation set and 514 were in the test set. The running environment of the experimental platform was Python 3.10 and PyTorch 2.1, the Adam optimizer was used for model training, the initial learning rate was set to 0.001, the batch size was set to 32, and the maximum training round was 100. On the whole, the dataset contains both static ability information and dynamic feedback characteristics in the learning process, which can support the comprehensive verification of personalized learning path recommendation, path generation accuracy and dynamic adjustment effect. The specific composition and sample distribution of the dataset are shown in Table 1.

Table 1: Data set composition and sample distribution table

Data Module	Quantity	Proportion of Total Data Entries / %	Description
Learner Samples	428 students	—	Students who participated in a full semester of College English learning
Platform Behavior Logs	18,640 records	53.07	Includes login, click, dwell time, and submission records
Stage Assessment Records	5,120 records	14.58	Includes vocabulary, reading, and comprehensive assessment results
Assignments and Writing Texts	2,140 files	6.10	Includes essays, translation tasks, and stage assignment texts
Teacher Feedback Records	1,680 records	4.79	Includes annotations, evaluations, and classroom observation information
Learning Resource Items	1,260 items	3.59	Includes videos, courseware, exercises, and supplementary materials
Knowledge Point Nodes	186 nodes	0.53	Corresponding to vocabulary, grammar, reading, and writing knowledge points
Path Sequence Samples	3,424 records	9.75	Used for personalized learning path generation and evaluation
Training Set Samples	2,397 records	70.00	Used for model parameter learning
Validation Set Samples	513 records	14.98	Used for model tuning and early stopping control
Test Set Samples	514 records	15.02	Used for final performance evaluation

4.2 Analysis of personalized learning path recommendation effect

After completing the configuration of the experimental environment and the division of the data set, we further tested the actual performance of the model from two aspects of recommendation accuracy and hitting ability. In order to ensure that the comparison results are representative, the rule matching method, collaborative filtering method, knowledge graph-based method and the method of this paper are selected as the comparison objects, and tested under the same data set, training rounds and evaluation criteria. The evaluation metrics mainly include Precision, Recall, F1-score and Top-N recommendation hit rate, which reflect the accuracy, coverage ability and ranking effectiveness of path recommendation results respectively. As shown in Figure 4, different methods show obvious differences in the recommendation effect.

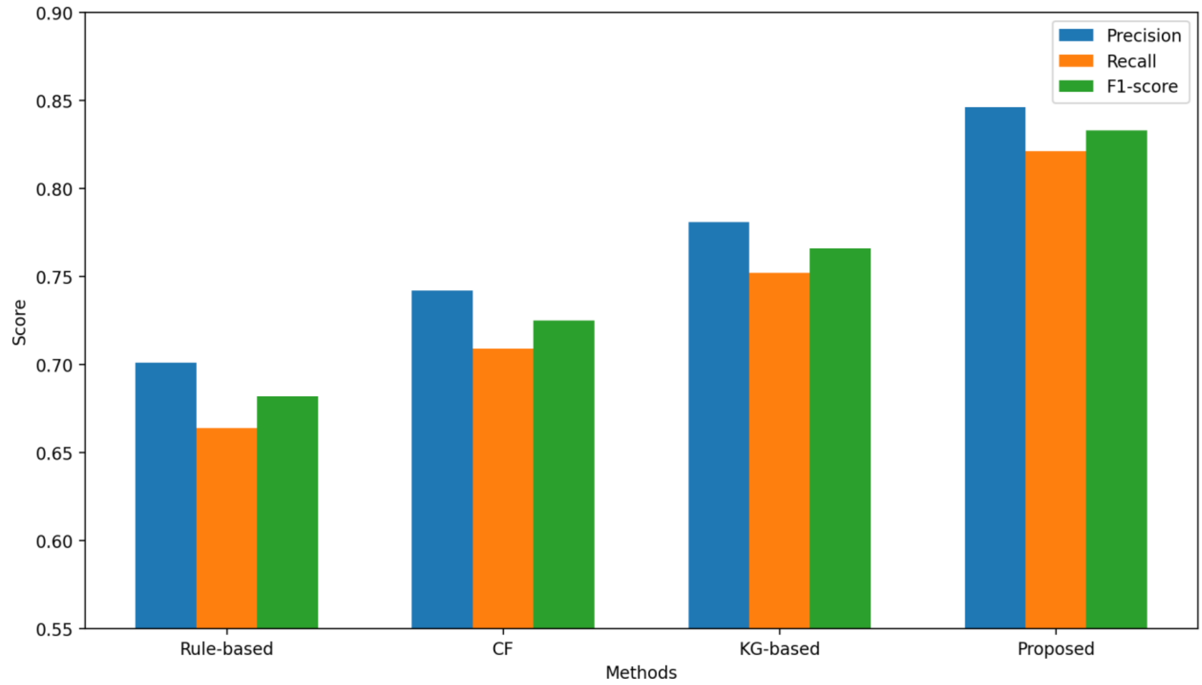

Figure 4: Bar chart of comparison of recommendation effect of different methods

Although the rule matching method can complete the preliminary recommendation according to the preset knowledge rules, it has a weak response to the individual differences of learners, and the Precision, Recall and F1-score are only 0.701, 0.664 and 0.682, respectively. The collaborative filtering method has a certain ability to utilize similar learners 'behavior patterns, and the three indicators are improved to 0.742, 0.709 and 0.725, but it is still susceptible to sparse data and cold start problems. The method based on knowledge graph performs better in resource relationship modeling, with Precision of 0.781, Recall of 0.752, and F1-score of 0.766, indicating that semantic association is helpful to improve the rationality of path recommendation. In contrast, the proposed method achieves the highest values on three indicators, with Precision reaching 0.846, Recall reaching 0.821, and F1-score improving to 0.833, indicating that the fusion of deep feature extraction and learner profiling can more accurately capture college English learning needs. And generate recommended paths that are more consistent with the capability gap and target preference.

Furthermore, in order to investigate the hit performance of different methods under different candidate sizes, this paper counts the recommendation hit rate under Top-3, Top-5 and Top-10 conditions, and the results are shown in Figure 5.

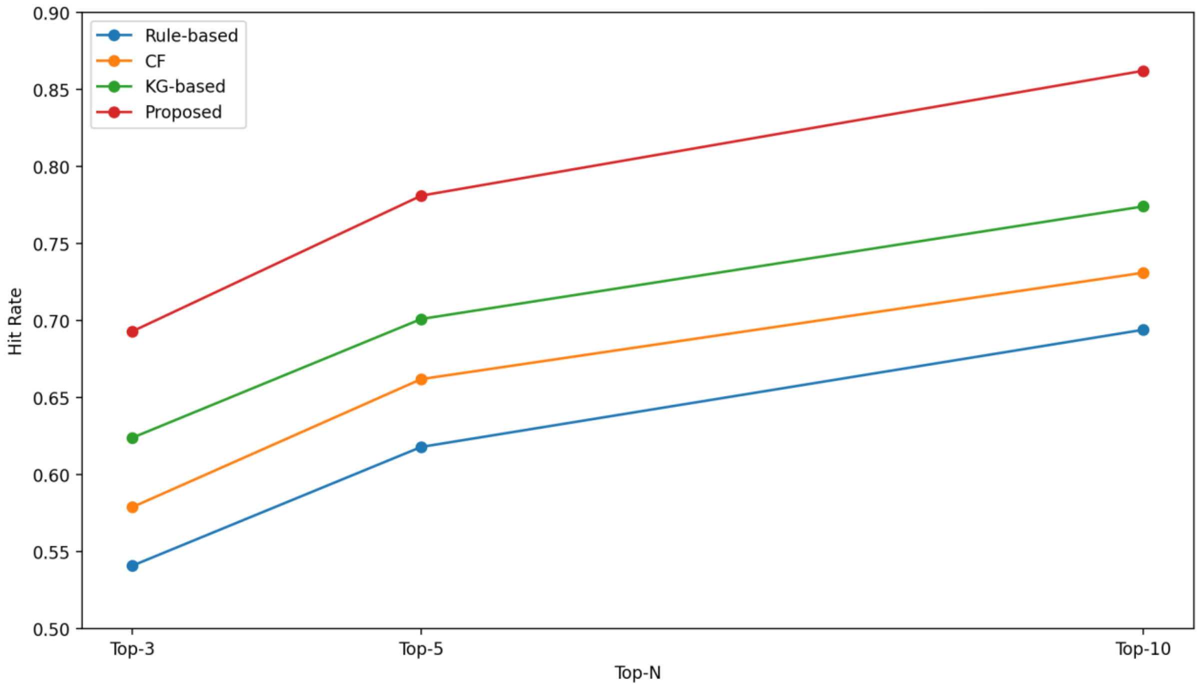

Figure 5: Line plots of Top-N recommendation hit rate for different methods

It can be seen that the hit rate of each method increases with the expansion of Top-N range, but there are differences in the growth rate and the final level. The rule matching method improved from 0.541 in Top-3 to 0.694 in Top-10, and the overall performance was weak. The collaborative filtering method achieves 0.579, 0.662 and 0.731, respectively, which indicates that certain improvement can be obtained under a large candidate range. The method based on knowledge graph reaches 0.701 in Top-5 and 0.774 in Top-10, which is more stable. The hit rate of the proposed method is 0.693 in Top-3, 0.781 in Top-5, and 0.862 in Top-10, which shows that the proposed method can not only maintain high accuracy in a smaller recommendation set, but also maintain good ranking quality in a wider recommendation range.

4.3 Analysis on the accuracy of College English personalized learning path generation

After completing the comparison of path recommendation effects, it is necessary to further judge whether the learning paths generated by the model truly conform to the organizational logic of college English learning tasks. Different from general resource recommendation, personalized learning path requires not only the recommended content to be selected correctly, but also the recommended order to be correctly ranked, that is, each node in the path should be consistent with the learners' current ability state, the target difficulty level and the coverage requirements of knowledge points. Based on this consideration, this paper compares different methods from two aspects of path generation accuracy and knowledge coverage, and analyzes the rationality of the model in path structure organization by combining with the difficulty matching of path nodes.

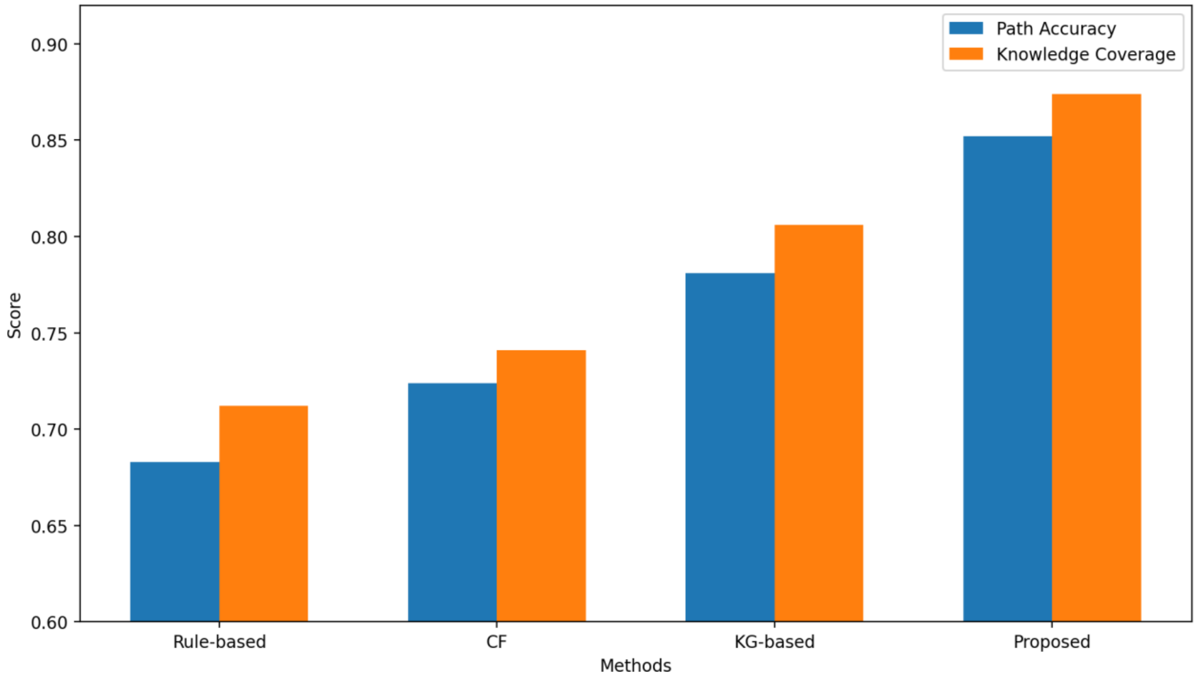

Figure 6: Comparison of path generation accuracy and knowledge coverage

As shown in Figure 6, the accuracy of path generation and the knowledge coverage of the rule matching method are 0.683 and 0.712 respectively, indicating that although this kind of method can give the learning order according to the preset rules, it still has obvious limitations in the face of complex learning requirements. The path generation accuracy of collaborative filtering method is improved to 0.724, and the knowledge coverage rate is 0.741, which is improved compared with the rule matching method, but its utilization of knowledge dependencies between learning resources is still insufficient. The method based on knowledge graph continues to improve in the two indicators, the path generation accuracy reaches 0.781, and the knowledge coverage rate is 0.806, indicating that semantic association modeling can enhance the integrity of path organization. In contrast, the method in this paper achieves the best results on both indicators, the path generation accuracy reaches 0.852, and the knowledge coverage rate reaches 0.874, indicating that the deep feature extraction and path sequence optimization mechanism can more effectively take into account learners' needs, resource matching and knowledge cohesion.

In addition to the overall accuracy and coverage rate, the incremental difficulty of the path nodes is also an important basis for measuring the quality of the path. To this end, this paper further compares the target difficulty curve, the path difficulty curve generated by the knowledge graph-based method, and the path difficulty curve generated by the proposed method, and the results are shown in Figure 7.

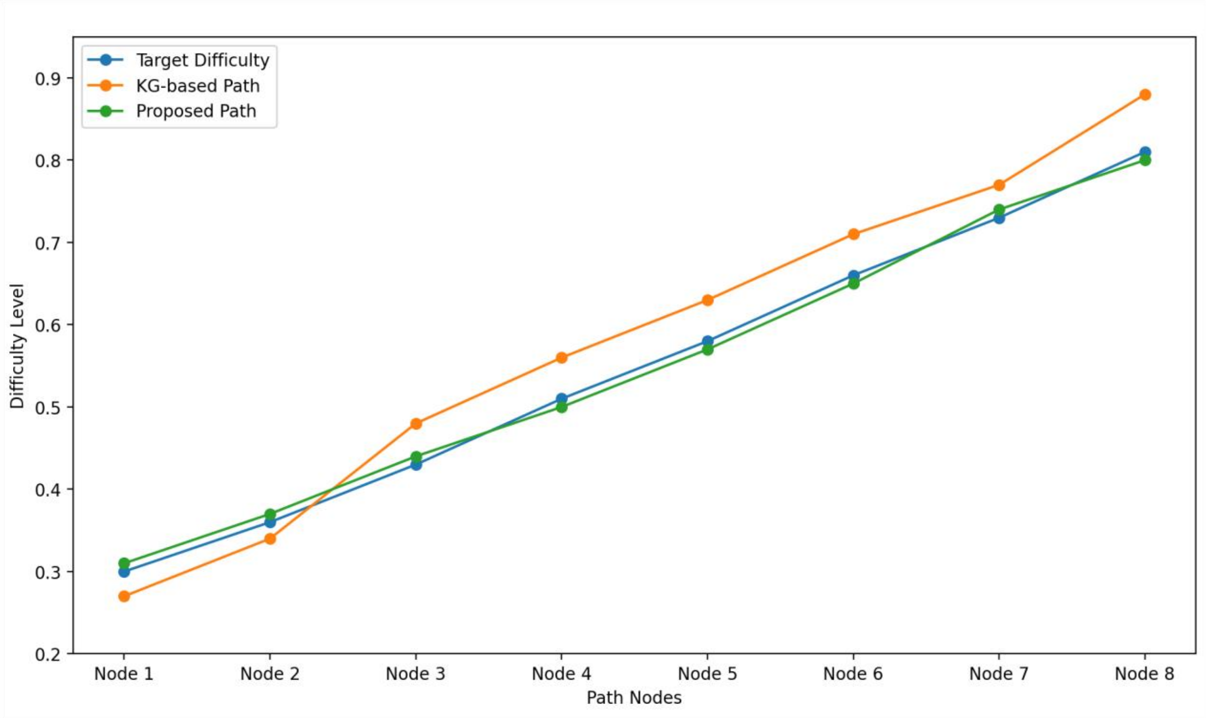

Figure 7: Difficulty matching curve of path nodes

As can be seen from Figure 7, the target difficulty gradually rises from 0.30 of Node 1 to 0.81 of Node 8, showing a relatively stable progressive trend. The method based on knowledge graph is close to the target value on the early Node, but gradually deviates from Node 4, and the difficulty of Node 8 reaches 0.88, which is 0.07 higher than the target value, indicating that there is a certain difficulty jump in the later path. The difficulty values of Node 4, Node 6 and Node 8 are 0.50, 0.65 and 0.80, respectively, and the deviations from the target value are all controlled at about 0.01. It shows that this method can better take into account the difficulty balance and learning adaptation in the process of path advancement.

In general, the proposed model is not only significantly better than other methods in path generation accuracy and knowledge coverage, but also shows stronger stability in node difficulty matching. Compared with the rule matching method, the path generation accuracy of the proposed method is increased by 0.169, and the knowledge coverage rate is increased by 0.162. Compared with the method based on knowledge graph, the two indicators are also increased by 0.071 and 0.068 respectively. Combined with Figure 7, it can be further explained that the learning path generated by the proposed method is more in line with the organizational requirements of college English learning tasks from shallow to deep and progressive while maintaining the integrity of knowledge coverage.

4.4 Effect analysis of path dynamic adjustment driven by learning feedback

This paper further investigates the effect of path dynamic adjustment mechanism after learning feedback reflux, and analyzes it from two aspects: the improvement of learning performance and the change of path deviation. The learning effects before and after path dynamic adjustment are shown in Table 2. Without dynamic adjustment, the average task accuracy rate of learners was 76.8%, the task completion rate was 73.5%, the average learning time was 34.6 min, the path deviation index was 0.214, and the satisfaction score was 3.72. After introducing the dynamic adjustment mechanism driven by learning feedback, the system can timely correct subsequent learning nodes according to stage performance, task completion efficiency and teacher evaluation, so that the average task accuracy rate is increased to 84.9%, task completion rate is increased to 82.7%, and the satisfaction score is increased to 4.31. At the same time, the average learning time decreases to 29.8 min, and the path deviation index decreases to 0.121. This result shows that dynamic adjustment does not simply rely on increasing learning tasks to improve performance, but improves learning efficiency and resource adaptation under a more reasonable path arrangement.

Table 2: Comparison table of learning effects before and after path dynamic adjustment

Metric	Before Adjustment	After Adjustment	Change Magnitude
Task Accuracy / %	76.8	84.9	+8.1
Task Completion Rate / %	73.5	82.7	+9.2
Average Learning Time / min	34.6	29.8	-4.8
Path Deviation Index	0.214	0.121	-0.093
Satisfaction Score	3.72	4.31	+0.59

In order to observe the path correction effect in the dynamic adjustment process more intuitively, this paper further statistics the change trend of the path deviation index under different learning rounds, as shown in Figure 8.

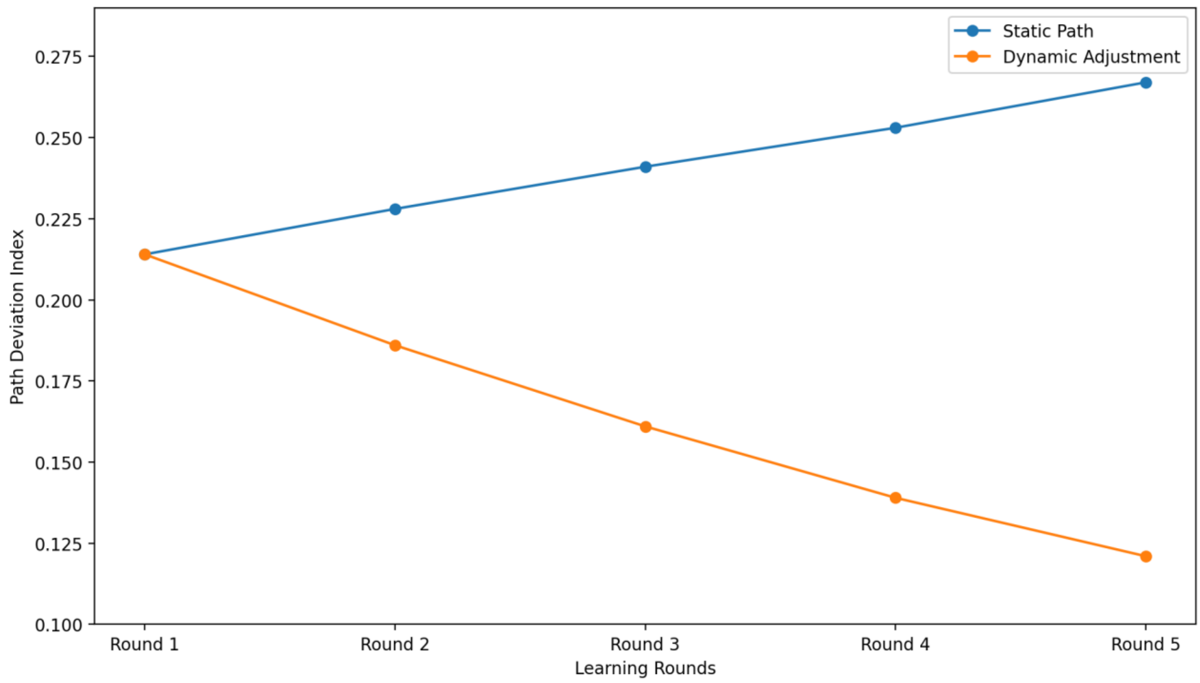

Figure 8: Curve of path deviation index change

It can be seen that the deviation index under the static path condition continues to rise from 0.214 in Round 1 to 0.267 in Round 5, indicating that the difference between the original fixed path and the actual state of learners is expanding with the progress of learning. In contrast, the dynamically adjusted path drops to 0.186 in Round 2, further drops to 0.161 in Round 3, and stabilizes at 0.121 in Round 5. In other words, the feed-back driven mechanism can continuously compress the path error in the learning process, so that the path arrangement gradually converges to the real needs of learners, instead of constantly deviating with the increase of rounds.

From the overall results, the dynamic adjustment mechanism driven by learning feedback has a significant role in promoting college English personalized learning. On the one hand, it improved the performance of learning results, and the accuracy and completion rates were increased by 8.1 and 9.2 percentage points respectively. On the other hand, it reduced the invalid learning consumption, shortened the average learning time by 4.8 min, and decreased the path deviation index by 0.093. Combined with Figure 8, it can be shown that the proposed method can not only generate a more accurate initial learning path, but also continuously correct subsequent nodes according to feedback in the real teaching process, so as to improve the continuous adaptation ability and application stability of the path.

5 Conclusion

Focusing on the personalized learning path design of college English, this paper constructs a complete method chain consisting of multi-source data collection, robust preprocessing, learner portrait modeling, deep feature extraction, path sequence generation, resource matching and feedback regulation. The results show that the unified modeling of multi-source learning data can better support learners 'state recognition and ability gap characterization. On this basis, the path generation mechanism driven by deep learning can simultaneously improve the recommendation accuracy, knowledge coverage level and difficulty matching stability. In the experiment, the F1-score of the proposed method reaches 0.833, the path generation accuracy reaches 0.852, and the knowledge coverage reaches 0.874. In the dynamic adjustment stage, the task completion rate is increased to 82.7%, the average learning time is shortened to 29.8 min, and the path deviation index is reduced from 0.214 to 0.121. On the whole, the proposed method takes into account the recommendation accuracy, path coherence and application adaptation ability. In the future, it can combine long-term classroom tracking data and finer-grained ability diagnosis mechanism to further improve the generalization and teaching application value of the model.

References

[1] Huang Y, Zhu J. A Personalized English Learning Material Recommendation System Based on Knowledge Graph[J]. International Journal of Emerging Technologies in Learning, 2021, 16(11): 160-173. DOI:10.3991/ijet.v16i11.23317.

[2] van der Velde M, Sense F, Borst J, van Rijn H. Alleviating the Cold Start Problem in Adaptive Learning using Data-Driven Difficulty Estimates[J]. Computational Brain & Behavior, 2021, 4(2): 231-249. DOI:10.1007/s42113-021-00101-6.

[3] Wang P, Huang X. An Online English Writing Evaluation System Using Deep Learning Algorithm[J]. Mobile Information Systems, 2022, 2022: 1-8. DOI:10.1155/2022/7605989.

[4] Jiang B, Li X, Yang S, Kong Y, Cheng W, Hao C, Lin Q. Data-Driven Personalized Learning Path Planning Based on Cognitive Diagnostic Assessments in MOOCs[J]. Applied Sciences, 2022, 12(8): 3982. DOI:10.3390/app12083982.

[5] Li B, Li G, Xu J, Li X, Liu X, Wang M, Lv J. A personalized recommendation framework based on MOOC system integrating deep learning and big data[J].

Computers & Electrical Engineering, 2023, 106: 108571. DOI:10.1016/j.compeleceng.2022.108571.

[6] Zhang S, Hui N, Zhai P, Xu J, Cao L, Wang Q. A fine-grained and multi-context-aware learning path recommendation model over knowledge graphs for online learning communities[J]. Information Processing & Management, 2023, 60(5): 103464. DOI:10.1016/j.ipm.2023.103464.

[7] Zhang X, Liu S, Wang H. Personalized Learning Path Recommendation for E-Learning Based on Knowledge Graph and Graph Convolutional Network[J]. International Journal of Software Engineering and Knowledge Engineering, 2023, 33(1): 109-131. DOI: 10.1142/S0218194022500681.

[8] Mu M, Yuan M. Research on a personalized learning path recommendation system based on cognitive graph[J]. Interactive Learning Environments, 2024, 32(8): 4237-4255. DOI:10.1080/10494820.2023.2195446.

[9] Tzeng J-W, Huang N-F, Chen Y-H, Huang T-W, Su Y-S. Personal learning material recommendation system for MOOCs based on the LSTM neural network[J]. Educational Technology & Society, 2024, 27(2): 25-42. DOI:10.30191/ETS.202404_27(2).SP03.

[10] Ma D, Zhu H, Liao S, Chen Y, Liu J, Tian F, Chen P. Learning path recommendation with multi-behavior user modeling and cascading deep Q networks[J]. Knowledge-Based Systems, 2024, 294: 111743. DOI:10.1016/j.knosys.2024.111743.

[11] Zhou L, Wang C. Research on Recommendation of Personalized Exercises in English Learning Based on Data Mining[J]. Scientific Programming, 2021, 2021: 1-9. DOI:10.1155/2021/5042286.

[12] Zhang Y. A Personalized Recommendation System for English Teaching Resources Based on Learning Behavior Detection[J]. Mobile Information Systems, 2022, 2022: 1-8. DOI:10.1155/2022/4531867.

[13] da Silva F L, Slodkowski B K, da Silva K K A, Cazella S C. A systematic literature review on educational recommender systems for teaching and learning: research trends, limitations and opportunities[J]. Education and Information Technologies, 2023, 28(3): 3289-3328. DOI:10.1007/s10639-022-11341-9.

[14] Liu T, Wu Q, Chang L, Gu T. A review of deep learning-based recommender system in e-learning environments[J]. Artificial Intelligence Review, 2022, 55(8): 5953-5980. DOI:10.1007/s10462-022-10135-2.

[15] Wang J, Xie H, Wang F L, Lee L-K, Au O T S. Top-N personalized recommendation with graph neural networks in MOOCs[J]. Computers and Education: Artificial Intelligence, 2021, 2: 100010. DOI:10.1016/j.caeai.2021.100010.

[16] Lin Y, Feng S, Lin F, Zeng W, Liu Y, Wu P. Adaptive course recommendation in MOOCs[J]. Knowledge-Based Systems, 2021, 224: 107085. DOI:10.1016/j.knosys.2021.107085.

[17] Lin Y, Lin F, Yang L, Zeng W, Liu Y, Wu P. Context-aware reinforcement learning for course recommendation[J]. Applied Soft Computing, 2022, 125: 109189. DOI:10.1016/j.asoc.2022.109189.

[18] Liu Y, Dong Y, Yin C, Chen C, Jia R. A Personalized Course Recommendation Model Integrating Multi-granularity Sessions and Multi-type Interests[J]. Education and Information Technologies, 2024, 29(5): 5879-5901. DOI:10.1007/s10639-023-12028-5.

[19] Tzeng J-W, Huang N-F, Chuang A-C, Huang T-W, Chang H-Y. Massive open online course recommendation system based on a reinforcement learning algorithm[J]. Neural Computing and Applications, 2025, 37(18): 11607-11618. DOI:10.1007/s00521-023-08686-8.

[20] Troussas C, Krouska A. Path-Based Recommender System for Learning Activities Using Knowledge Graphs[J]. Information, 2023, 14(1): 9. DOI:10.3390/info14010009.

[21] Niu Y, Lin R, Xue H. Research on Learning Resource Recommendation Based on Knowledge Graph and Collaborative Filtering[J]. Applied Sciences, 2023, 13(19): 10933. DOI:10.3390/app131910933.

[22] Pelánek R, Effenberger T, Jarušek P. Personalized recommendations for learning activities in online environments: a modular rule-based approach[J]. User Modeling and User-Adapted Interaction, 2024, 34: 1399-1430. DOI:10.1007/s11257-024-09396-z.

[23] Ahmadian Yazdi H, Seyyed Mahdavi S J, Ahmadian Yazdi H. Dynamic educational recommender system based on Improved LSTM neural network[J]. Scientific Reports, 2024, 14(1): 4381. DOI:10.1038/s41598-024-54729-y.

[24] Zhang F, Feng X, Wang Y. Personalized process-type learning path recommendation based on process mining and deep knowledge tracing[J]. Knowledge-Based Systems, 2024, 303: 112431. DOI:10.1016/j.knosys.2024.112431.

[25] Zhou L-Y, Wang Y-Y. Simulation of personalized english learning path recommendation system based on knowledge graph and deep reinforcement learning[J]. Scientific Reports, 2025, 15: 34554. DOI:10.1038/s41598-025-17918-x.