International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 5171-5181

https://doi.org/10.70517/ijhsa464358

Optimization of Multi-scenario Teaching Strategies for Higher Vocational Railway English Based on Reinforcement Learning

Yao Li[1,*]
[1] Hunan Technical College of Railway High-speed, Hengyang, Hunan, 421002, China
Corresponding authors: (e-mail: ly380512560@126.com).

Abstract With the rapid development of information technology, the field of education is gradually adopting intelligent technology to enhance the teaching effect. This study proposes a multi-scenario teaching strategy optimization path based on reinforcement learning for English in the higher vocational railroad industry, which recommends personalized learning paths for learners by combining the Deep Knowledge Tracking (DKT) model with the Reinforcement Learning (RL) method. In the experiments, the model is validated on different datasets, in which the AUC in the Skill-builder-data-2023-2024 dataset reaches 0.83837, and the accuracy is 0.75654. By comparing with the traditional models (e.g., BKT, DKT, KNN, etc.), the method of this paper demonstrates a significant advantage in the accuracy of the learning path recommendation. At the same time, it is also able to adjust the recommendation strategy according to the dynamic performance of the learner to further optimize the learning effect. The results show that the learning path recommendation based on reinforcement learning can significantly improve the degree of personalization and adaptability of higher vocational English teaching, and has high practical value.

Index Terms Reinforcement learning, deep knowledge tracking, learning path, personalized recommendation, AUC, accuracy rate

I. Introduction

With the rapid development of China's railroad business, the requirements for English speaking ability in the railroad industry are getting higher and higher [1]. However, the current level of English application ability of railroad employees is low, which is difficult to meet the needs of cross-linguistic communication on the railroad. In order to improve the English speaking level of railroad staff and adapt to the internationalized working environment, it is imperative to improve the quality of English teaching in the higher vocational railroad industry [2]-[4].

At present, most of the higher vocational colleges and universities have offered professional English courses, but there are students' basic knowledge of English is not solid, vocational colleges and universities have weak English teachers, especially the teaching strategy is relatively backward [5], [6]. Therefore, in order to improve the cultivation efficiency of railroad professional English talents, the optimization of teaching strategy of reinforcement learning under the background of artificial intelligence has received wide attention [7], [8]. Reinforcement learning is an important branch of machine learning, which aims to make the intelligent body learn the optimal decision-making strategy through interaction with the environment [9], [10]. In this process, “strategy search” and “optimization” are two key techniques, which have a direct impact on the learning efficiency and final performance of the intelligences [11], [12]. In the English teaching of railroad majors, reinforcement learning can provide new ideas for teaching strategies, and railroad majors, as an emerging course, have relatively small class size, which is conducive to the flexible use of the above teaching methods [13]-[15].

This study combines reinforcement learning with knowledge tracking technology to model students' learning state through deep learning models and optimize learning path recommendation using reinforcement learning. First, the deep knowledge tracking model is used to analyze students' historical learning records and assess their current knowledge mastery; then, the reinforcement learning model is used to dynamically recommend learning paths based on students' cognitive state and learning progress. This method not only improves the accuracy and efficiency of recommendation, but also continuously adjusts the learning strategy according to the students' learning progress, thus realizing personalized and intelligent multi-scenario teaching support.

II. Theoretical foundations

II. A. Overview of knowledge tracking

Knowledge Tracking Overview Knowledge Tracking (KT), a technique designed to track the state of a learner's knowledge level, which predicts the current learner's mastery of a knowledge point at a given point in time based on the learner's historical learning record. In the knowledge tracking task, the first task is to model the user's interaction with the exercises. Given a student's historical learning record $E_t = (e_1, e_2, \ldots, e_t)$, the knowledge-tracking task is to predict the student's performance on the next time window e_{t+1}, which can be interpreted as the degree of mastery of the next learning content. Where $e_i = (x_1, x_2, \ldots, x_n)$ denotes the student's learning record on the i th time window, n refers to the number of completed questions in the current time period, and x_j is specifically an ordered pair of the student's record of doing the questions, i.e. $x_j = (q_j, a_j)$, which is an ordered pair representing the j th question answered by the student under the current time window, where a_j denotes the result of that answer, and in general a_j is either 0 or 1, with 0 being an incorrect answer for that question, and 1 being a correct answer [16]. The process of knowledge tracking task for users is shown in Figure 1.

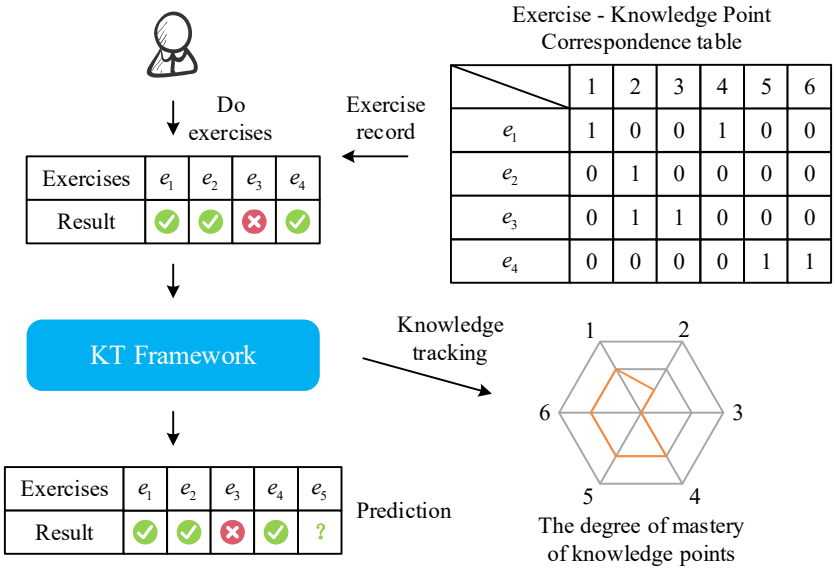


Figure 1: conducts knowledge tracking for users

The Deep Knowledge Tracking (DKT) model introduces Recurrent Neural Networks (RNN) to the knowledge tracking task for the first time, tracking the process of dynamic changes in the learner's knowledge proficiency over time and modeling the state of the learner's knowledge level. The deep knowledge tracking model is shown in Figure 2.

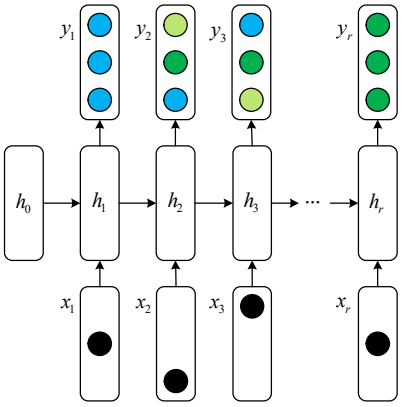


Figure 2: Deep Knowledge Tracing (DKT)

In the figure, x_i represents the input vector under the i th time window, h_i is the state of the hidden layer, and is the student's knowledge state matrix, and the final output of y_t is the degree of the student's knowledge of the point of x_t, which is also the probability of answering the right question, under the t th time window.The DKT model's The formula for calculating the DKT model is:

$$h_t = \tanh(W_{h_x} x_t + W_{hh} h_{t-i} + b_h) \tag{1}$$

$$y_t = \sigma(W_{yh} + b_y) \tag{2}$$

In Equation (1), tanh is the tanh function, W_{h_x} is the weight matrix, b_h is the bias of the hidden layer, and W_{hh} is the recursive weight matrix. In Equation (2), σ is the Sigmoid function, W_{yh} is the input weight matrix, and b_y is the bias of the output layer.

II. B.Overview of Enhanced Learning

II. B. 1) Basic Concepts of Reinforcement Learning Techniques

The model of reinforcement learning technique is shown in Figure 3. The task is to continuously explore in the constructed environment, the specific process is to observe the state in the environment, and then take the corresponding action, thereby obtaining the environment feedback to the intelligent body's current exploration of the action of the reward Finally, through the maximization of the reward to learn and update the strategy, and constantly interact with the environment, so as to find the optimal solution to reach the end state. Reinforcement learning simulates the environment explored by the intelligent body as the environment in which the current learning task takes place, which is the core element of reinforcement learning.State, i.e., state. The state of the intelligent body in the environment can be understood as a property of the intelligent body itself, which can be location information or other variables that characterize the environment. Action is the behavior produced by the intelligent body during the interactive exploration of the current environment, and the action space is represented as A, where the action under the time window t is $a_t \in \mathrm{A}$. Reward is a certain reward that the environment feeds back to the intelligent body based on the action taken by the intelligent body in its current state during the interactive exploration of the current environment.

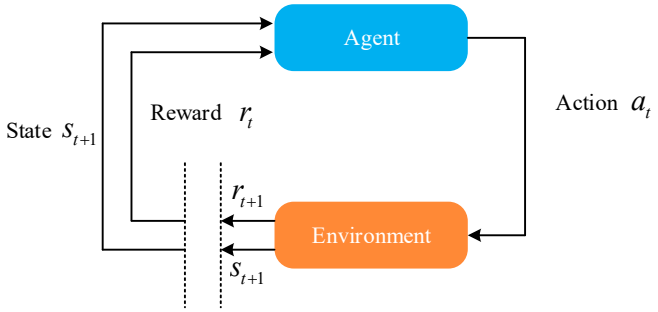


Figure 3: Reinforcement Learning Model

II. B. 2) Markov decision-making process

Reinforcement learning problems are often described as Markov Decision Processes (MDPs). A Markov decision process has dependencies between states and a temporal order and is a sequential decision process. The Markov decision process contains five elements which can be represented by a quintuple, $M = (S, A, P, R, \gamma)$. Where S is a finite state space, A is a finite action space, also called the set of actions, $P = \{P_s a(\cdot) \mid s \in S, a \in A\}$ is the probability of transferring to the next state, R is the reward function, and γ is the discount factor.

Where the transfer probability matrix P is specified as $P(s,a,s') = P_{(s,a)} r(s_{t+1} = s' \mid s_t = s, a_t = a)$, and the instantaneous reward r is expressed as $r = R(s,a,s') = E\{r_{t+1} \mid s_t = s, a_t = a, s_{t+1} = s'\}$.

The Markov decision process (MDP) is shown in Figure 4. s_0 is the initial state of the intelligent body, after taking the action a_0 in this state, it transfers to the state s_1 according to the probability p_0 and receives the immediate reward r_1 from the environment, and then repeats the above process continuously.

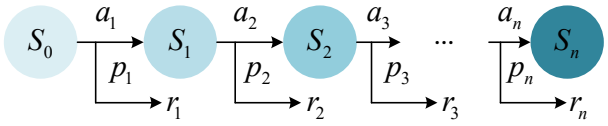


Figure 4: Markov Decision Process

In Markov Decision Process (MDP), which is based on the theory of stochastic processes, it is stipulated that the current state of the intelligent Agent is only related to the state of the previous moment and the action taken. Meanwhile, reinforcement learning methods based on this view the recommendation process as a Markov decision process [17]. As a result, the reinforcement learning based recommendation model can portray sequential features of user behavior and is able to capture the dynamic preferences of the user. The model also dramatically improves the diversity of recommendation results by allowing the intelligences to fully explore the state space and action space according to their exploration mechanisms. This type of model generally performs recommendation strategy updates by focusing on maximizing the cumulative gain of the recommendation system as an optimization goal, focusing on the long-term feedback of the dry users, thus focusing on the improvement of the long-term satisfaction of the users, and also improves the accuracy of the recommendations.

II. B. 3) Enhanced Learning Algorithm

(1) QLearning algorithm

The QLearning algorithm is an offline strategy algorithm and a value-based reinforcement learning algorithm. Among them, Q refers to the Q value, and the Q function in the algorithm, that is, Q (state, action), is expressed as the Q value obtained by executing the action A under the state S, also called $Q-value$. The QLearning algorithm aims to maximize Reward by selecting the optimal action A under the state S, i.e., the best action, in order to get the maximum return, with the ultimate goal of maximizing the Q value.

The QLearning algorithm always maintains a Q table ($Q-table$) to store the Q value ($Q-value$) corresponding to different actions taken in different states. In the interaction between the agent and the environment, the Q value is modified and the Q table is updated according to the instant reward r obtained and the algorithm rules. As the number of iterations increases, the agent improves its understanding of the environment, and the Q function gradually fits until it converges or reaches the set number of iteration ends.

The Value update formula of QLearning algorithm is shown in equation (3).

$$Q(s,a)arrowQ(s,a)+\alpha[R(s,a)+\gamma\max Q'(s',a')-Q(s,a)] \tag{3}$$

$$s\leftarrow s' \tag{4}$$

where α is the learning rate and denotes the update magnitude, $R(s,a)$ is the immediate reward, $\max Q'(s',a')$ is the value of Q obtained by selecting the action that maximizes the next state at the next state, and $\gamma\, maxQ'(s',a')$ denotes the future long-term reward. The $R(s,a)+\gamma\max Q'(s',a')$ is denoted holistically as the true Q -value of the action a_t to be taken in the current state s_t, consisting of the immediate reward R and the long-term reward. The $Q(s,a)$ in the middle bracket is the estimated Q value, and the difference between the actual Q value and the estimated Q value is denoted as $\Delta Q(s,a)$, and when this difference tends to 0, $Q(s,a)$ no longer changes and the whole tends to a convergent state.

(2) MCMC algorithm

Monte Carlo Method Control (MCMC), a learning method for estimating value functions, learns policy-based state-value functions, with the main idea of averaging sample gains and learning from actual or simulated interaction environments based on the states, actions, and rewards of a sequence of empirical samples. Monte Carlo methods do not require prior knowledge of the environment dynamics and still achieve optimal behavior.

III. Recommendation of Adaptive Learning Path for English in Higher Education Railway Industry

III. A. Problem definition

III. A. 1) Learning phase

Learning is a long process that can be broken down into more separate learning stages. Each stage of learning has separate learning objectives, which can be set by the teacher or the learner. Each learning stage consists of two main components: learning paths and quizzes. The learning validity of a stage E_P can be defined as:

$$E_p = \frac{E_e - E_s}{E_{sup} - E_s} \tag{5}$$

where E_s is the score of the quiz at the beginning of a given study, E_e is the score of the quiz at the end of the study, and E_{sup} is the total score of the quiz.

III. A. 2) Learning sequence diagrams

There are some domain-specific relationships in educational knowledge mapping, such as the learning sequence and similarity relationships between learning items, which can represent specific knowledge structures. Usually, when learners learn, they tend to follow the learning pattern of "first easy, then difficult". The learning sequence relationship graph is a subgraph of the Education Knowledge Graph, which represents the learning sequence relationship between learning items. The nodes in the graph represent learning programs, and the arrows from one to the other mean that the former is the prerequisite entity for the latter.

III. A. 3) Learning path recommendation problem

Learning is composed of multiple learning stages. In each phase, the learner will accomplish a specific learning objective $T = t_0, t_1, \ldots$, the learning objective can be to master the knowledge in one or more learning items. Learning items are nodes on the learning sequential relationship graph G, and edges represent learning sequential relationships, i.e., (i, j). The tuple (i, j) indicates that the learning item i is the prerequisite entity for j. For a learner who is about to start a new phase of learning, the historical learning records generated in previous learning phases are represented as $\mathrm{H} = \{h_0, h_1, ..., h_m\}$. Each record $h_t = \{k, score\}$ contains a learning item k and a corresponding learning situation $score$. Without loss of generality, it is assumed that $score$ is a discrete number 0 or 1, where 1 means that the corresponding item is mastered, while 0 means that it is not mastered. The goal of this paper is to recommend an optimal learning path $P = \{p_0, p_1, \ldots p_N\}$, containing N items, and by following the recommended path, the learner can achieve a greater improvement in knowledge mastery.

III. B. Cognitive Structure Enhanced Adaptive Learning Framework CSEAL

III. B. 1) Definition of MDP

The goal is to learn a policy that recommends targeted learning paths based on cognitive structures. In this paper, we model such sequential path generation as a decision problem and treat it as an MDP.The states, actions, and rewards of an MDP are defined as follows.

State: at each step of recommendation, the probability of selecting a certain learning item is closely related to the learning goal and the previous learning record. The former represents the focus and direction of the learner's learning, while the latter implies the learner's learning ability and cognitive state. The state at step i is defined as the combination of the learning objective τ and the current cognitive state s_i, denoted as $state_i$. The learning objective is represented using a unique heat code, denoted as $\mathrm{T} = \{0,1\}^M$:

$$T^j = \begin{cases} 1 & \text{If } j \text{ is in the learning goals} \\ 0 & \text{other} \end{cases} \tag{6}$$

where M is the total number of nodes in the learning sequential relationship graph G. As mentioned earlier, the learner's cognitive state is a hidden variable that cannot be directly observed, and L_{i-1} consists of two parts, one part is the historical learning record H before the current learning front-end, and the learning record generated in the previous $i-1$ steps within the current learning phase $F_{0,\ldots,i-1}$.

Action: action a_i means recommending learning items p_i at step i. If the probability of recommending each item is taken as an output, CSEAL can be viewed as a stochastic strategy which is implemented by sampling from the distribution $pi(a \mid state_i; \theta) = P(a \mid \mathrm{H}, \mathrm{F}_{0,\ldots,i-1}, \mathrm{T}; \theta)$ in sampling to determine the next action, where θ is a model parameter.

Reward: after taking an action, the model receives a reward signal r. During the learning process, the reward r_i for step i is kept 0. The goal of this paper is to maximize the sum of the discounted rewards for each step i. The reward is maximized:

$$R_i = \sum_{i}^{N} \gamma^i r_i \tag{7}$$

where γ is the discount factor, which is usually set to 0.99. Overall, the CSEAL model proposed in this paper consists of three sub-modules, i.e., Knowledge Tracking (KT), Cognitive Navigator (CN), and Actor-Reviewer-based Recommender (ACR). KT is responsible for mining the implicit cognitive state S based on the previous learning record at each step, and CN selects a set of candidate actions that matches the learning logic based on a learning sequential relational graph G to select the set of candidate actions that match the learning logics. The learning goal and the cognitive states mined by KT are input into ACR as states to decide the next optimal learning item by maximizing the overall gain over the entire learning path. At the end of the learning phase, the rewards obtained from the final full path are passed to CSEAL and used for reinforcement learning to adjust the model parameters.

III. B. 2) Knowledge tracking

The learner's current cognitive state has a very strong influence on his/her performance on the recommended learning program, and thus needs to be focused on when determining the desired path recommendation strategy. At the same time, as a key component of state in MDP, it needs to be accurately and quantitatively modeled [18]. In this paper, we use knowledge tracking techniques to model the implied cognitive states from previous learning records S_i, $\mathrm{L}_{i-1} = \mathrm{H} \oplus \mathrm{F}_{0,\ldots,i-1}$. In this paper, we extend the DKT model structure with an embedding layer that reduces the feature space. It is assumed that $score_t$ in the prior learning record $\mathrm{L}_t = \{p_t, score_t\}$ is either 0 or 1. L_t is characterized by a unique heat code as described in DKT:

First, the learning record L_i characterized by the unique heat code is converted into a low-dimensional dense vector x_i using the embedding operation.

The above process can be described as:

$$x_t = \mathrm{L}_t W_u \tag{8}$$

$W_u \in \mathbb{R}^{2 \cdot M \times d}$ denotes the parameters of the embedding layer, $x_t \in \mathbb{R}^d$, where d is the output dimension.

After obtaining a representation of the learning record, KT can track the dynamic cognitive state based on the learning record. The LSTM structure described in the DKT is used to represent the input record representation vector sequence $x_1, x_2, ..., x_N$ are mapped into intermediate vector sequences $o_1, o_2, ..., o_N$. The update formula for the LSTM hidden state h_r at step t is as follows:

$$\begin{aligned} i_t &= \sigma(W_{xi} x_t + W_{hi} h_{t-1} + b_t) \\ f_t &= \sigma(W_{xf} x_t + W_{hf} h_{t-1} + b_t) \\ o_t &= \sigma(W_{xo} x_t + W_{ho} h_{t-1} + b_0) \\ c_t &= f_t c_{t-1} + i_t \tanh(W_{xc} x_t + W_{hc} h_{t-1} + b_c) \\ h_t &= o_t \tanh(c_t) \end{aligned} \tag{9}$$

Over the fully connected layer from the intermediate vectors $o_1, o_2, ..., o_N$ from the cognitive state vector $S_t \in \mathbb{R}^M$:

$$S_t = \sigma(W_{oF} o_t + b_F) \tag{10}$$

where $i_\cdot, f_\cdot, c_\cdot, o_\cdot$ are the input gates, forgetting gates, memory cells and output gates of the LSTM, respectively. $w_\cdot$ and $b_\cdot$ are the learnable weight matrix and bias.

III. B. 3) Cognitive navigation

Humans need to follow general cognitive laws when learning, and such cognitive laws can be expressed as the structure of potential semantic networks between learning items, i.e., knowledge structures. In this paper, we focus on quickly filtering out potential learning items rather than directly selecting a learning path.

III. B. 4) Actor-critic based recommenders

In this paper, inspired by the phenomenon that teachers in practice dynamically adjust their teaching strategies based on student performance feedback, an actor-critic algorithm is used to adaptively learn the optimal recommendation strategies. A strategy network is used as an actor to generate actions from the filtered D sampled in the distribution $\pi(a \mid state_i; \theta)$, and a value network is used as a critic to evaluate the state. The value network $v(\cdot; \theta_v)$ is used to estimate the expected return for each state. It is a feed-forward network with inputs $state_i = \mathrm{S}_i \oplus T$, where S_i is the current cognitive state and T is the learning goal. The predicted gain V_i for the i th step can be computed in the following way:

$$v_i = \mathrm{V}(state_i;\theta_\mathrm{V}) = \mathrm{V}(\mathrm{S}_i \oplus \mathrm{T};\theta_\mathrm{V}) \tag{11}$$

With the stochastic policy and value network, the actor-critic algorithm is applied to the path generation problem, using the policy gradient at step i to train the policy network:

$$\nabla_\theta = \log\pi(a \mid state_i;\theta)(R_i - v_i) \tag{12}$$

And train the value network by minimizing the distance between the return prediction and the actual return:

$$\mathrm{L}oss_{value} = \| v_i - R_i \|_2^2 \tag{13}$$

Since too fast convergence of the value network may lead to slow convergence or even non-convergence of the strategy network, this paper proposes a strategy-enhanced loss term to solve this problem. The final loss function is:

$$\begin{aligned}\mathrm{L}oss = &\| \mathrm{V}(state_i;\theta_\mathrm{V}) - R_i \|_2^2 + \alpha \cdot -log\pi(a \mid state_i;\theta)(R_i - v_i) \\ &+ \beta \cdot -log\pi(a \mid state_i;\theta)R_i\end{aligned} \tag{14}$$

where α and β are hyperparameters.

IV. Model testing and application

IV. A. Model Performance Testing

IV. A. 1) Data set preparation

In this section, the performance test of the model proposed in this paper is carried out. The model is tested mainly for its generalization ability. The generalization ability is the prediction ability shown for unknown data, and its quantitative metrics are accuracy, precision, recall, subject characteristics (ROC) curve and area under the ROC curve - AUC, and confusion matrix F1 value. The test metrics Accuracy, AUC, etc. are described for this paper. The dataset should satisfy the principle of reasonable distribution of test and training data magnitude and the principle of independent and same distribution of data, i.e., the training set and the test set cannot cross and are in the same distribution state. Since the system is an online platform, learners can not only get the recommendation service, but also the learning process will be recorded, and these data will be filtered and processed to generate new experimental data.

IV. A. 2) Analysis of results

In this paper, the model incorporates the learning process data into the knowledge state measurement with the aim of improving the accuracy of the knowledge tracking model. In addition, in order to solve the problems of input information reconstruction and fluctuation of prediction results, the model introduces regularization coefficients. The AUC performance of different knowledge tracking algorithms is shown in Table 1. In the skill-builder-data-2023-2024 dataset, the AUC of BKT and DKT reaches 0.58803, 0.82309, and the AUC of this paper's model reaches 0.83837. In the dataset ADM-2018, the AUC of BKT and DKT reaches 0.61808, 0.68073, and this paper's model reaches 0.7691.

Table 1: The AUC performance of different knowledge tracking algorithms

	BKT	DKT	Ours
Skill-builder-data-2023-2024	0.58803	0.82309	0.83837
ADM-2018	0.61808	0.68073	0.7691

The model Acc performance is shown in Table 2. In the skill-builder-data-2023-2024 dataset, the Accuracy of BKT and DKT reaches 0.67233, 0.71058, and the Accuracy of this paper's model reaches 0.75654. In the dataset ADM-2018, the BKT, DKT's Accuracy are 0.59559, 0.6223 respectively, and the model in this paper reaches 0.69723.From the figure, it can be seen that the AUC and Accuracy of DKT are improved compared with BKT. Among them, the AUC performance improvement is most obvious in the skill-builder-data-2023-2024 dataset. This indicates that DKT utilizes the nonlinear fitting property of neural networks and the forgetting property of LSTM to improve the accuracy of predicting learners' knowledge states. In order to improve the algorithm computational performance and explore the potential connection between each learning feature, the learning information of this paper's method is not only not lost, but the AUC and Accuracy are also improved.

Table 2: Model ACC performance

	BKT	DKT	Ours
Skill-builder-data-2023-2024	0.67233	0.71058	0.75654
ADM-2018	0.59559	0.6223	0.69723

IV. B. Comparison Test Experiments

In this section, the effectiveness of the proposed type is compared with the classical models on several current recommendation problems. The comparison experiment is divided into two parts, one is to compare the effectiveness of this paper's model with the comparison model by fixing the learning path length. The second is to compare the effectiveness of each model in different path lengths by using the learning path length as a variable.

The models for comparison are KNN, GRU4Rec, NARM and DQN. the following is a brief description of the comparison models:

KNN: KNN finds the closest predefined number of new users by comparing the cosine distance of the learning paths and recommends what to learn next for the new users.

GRU4Rec: GRU4Rec is the classic session-based recommendation model. The input to the model is a sequence of sessions and the output is a probability distribution of the next occurrence of the learned item.

NARM: NARM is a state-of-the-art personalized trajectory-based recommendation method for RNN models. It uses an attentional mechanism to determine the relevance of past learning trajectories to the next learning element.

DQN: Several studies have proposed the use of reinforcement learning to solve the learning path recommendation problem.

A comparison of the average validity of the models on the two datasets is shown in Table 3. The results show that the model proposed in this paper outperforms the other comparison models. Comparing the results on the two datasets, KNN, GRU4Rec and NARM perform well on the Junyi dataset and slightly worse on the ASSISTments dataset. It can be seen that the traditional methods have high applicability on the data with one-to-one knowledge points of the exercises and poor applicability on the data with many-to-many knowledge points of the exercises. And the traditional DQN model is almost impossible to make effective recommendation on two datasets. The method proposed in this paper outperforms the other four comparative models on both datasets, which shows that splitting the learning path recommendation into the knowledge point level and the exercise level as well as introducing a course guide to the reinforcement learning intelligences can indeed improve the applicability of reinforcement learning algorithms in the learning path recommendation scenario.

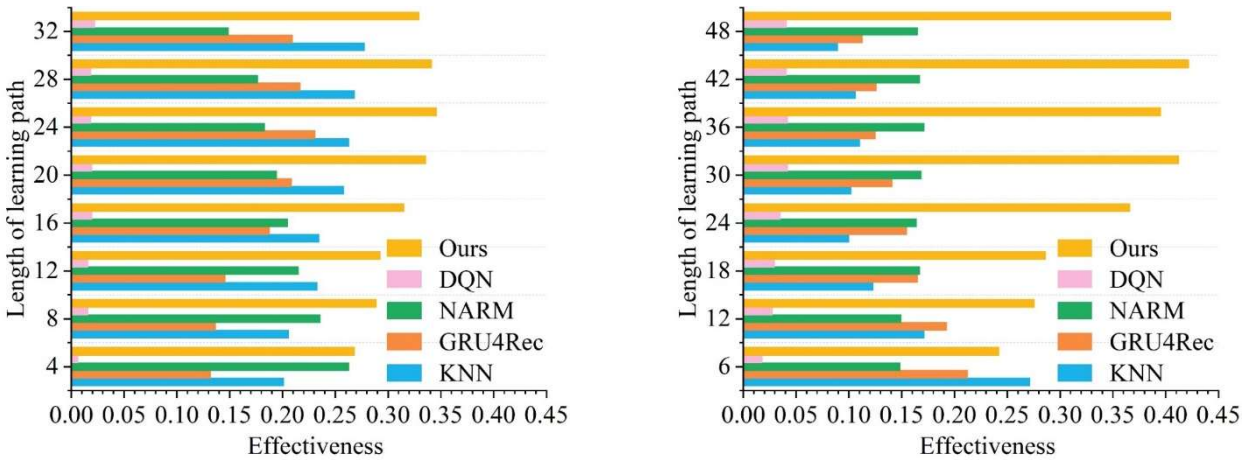
Table 3: Contrast model is compared to the average effectiveness of two data sets

Model name	Junyi	ASSISTments
KNN	0.25387	0.0878
GRU4Rec	0.19963	0.12402
NARM	0.19294	0.15586
DQN	0.00491	0.02187
Ours	0.33675	0.40879

Comparison of the efficiency of each model on different learning path lengths is shown in Fig. 5 (Figs. a and b show the experimental results under Junyi and Assistment datasets, respectively). The median number of exercises within a single answer log in the statistical information of the two datasets is 20 and 33, respectively, and it is hypothesized that the most suitable learning path length should be around 20 and 33. The experimental results show that the average efficiency of the model proposed in this paper on the two datasets achieves better results at path lengths of 22 and 31, respectively, and outperforms other comparative models in longer learning path scenarios.

IV. C. Example of Adaptive Learning Path Recommendation

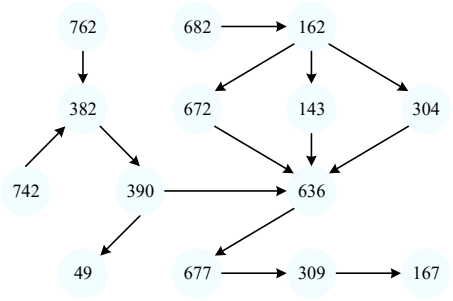
This subsection gives an example of multi-scenario adaptive learning path recommendation for English in the senior railroad industry. An example of different adaptive learning path recommendations regarding the same learning objective is shown in Figure 6. Figure (a) shows a part of the prerequisite graph after preprocessing the knowledge graph contained in the Junyi dataset, in this example, the learner's learning goal is a learning item containing knowledge point 636, and its history of interactions is grouped as {(636,0),(636,0),(636,0),(636,0),(636,0)}, which suggests that this learner may have dilemma in the learning process and is completely unable to master the knowledge point.636 Figure (b) shows the learning paths recommended for them using the two recommendation models GRU4Rec and the method in this paper, respectively.



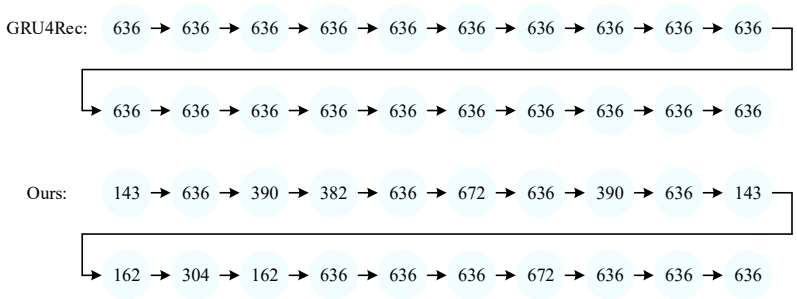
(a) Junyi (b) Assistment

Figure 5: Comparison of model efficiency on different learning path length

According to the experimental results, the multi-scenario learning path recommended by GRU4Rec for his English in the higher vocational railroad industry contains only the learning items corresponding to the target knowledge point 636. This learner had not mastered Knowledge Point 636 originally and would not be able to complete any step in the path along this learning path. The path recommended by this paper's method encourages the learner to review the knowledge points that have prerequisites for the target knowledge point. Based on the experimental results, it can be seen that this paper's method first recommends the learning item containing knowledge point 143 for this learner, and then chooses to suggest that he or she learns the target knowledge point or the knowledge point associated with the target knowledge point according to the learner's answers, for example, in the third time step it recommends the learning item containing knowledge point For example, in the third time step, the learning item containing knowledge point 390 is recommended for the learner, in the fourth time step, the learning item containing knowledge point 382 is recommended for the learner, etc., and in the later time steps of the path recommendation, the learning item containing knowledge point 636 is recommended, which shows that the method in this paper can recommend a more efficient and cognitively logical learning path for the learner.



(a) Partial prerequisite diagram



(b) Use the learning paths recommended by GRU4Rec and ours

Figure 6: Examples of different adaptive learning path recommendations

IV. D. Case Studies

In this study, we also randomly selected the adaptive learning path of a learner in the experimental class of a school for teaching English in multiple scenarios in the higher vocational railroad industry to conduct a case study, and the case study of the adaptive learning path is shown in Figure 7. In the figure, the horizontal axis (x-axis) represents the path recommended by the exercise, and the vertical axis (y-axis) represents the accuracy rate, the point colored green represents the average accuracy rate of all learners who answered the exercise in the system, and the point colored blue represents the student's answer in each exercise, if the answer is correct, y=1. If the answer is incorrect, y=0. From the figure, we can see:

(1) from the green point of the trend of change, the system recommended for the learner's accuracy of the exercises roughly show a wave, when the student answered incorrectly on an exercise, the system recommended for the learner a higher accuracy of the exercise, when the learner answered correctly on an exercise, the system recommended for the learner's accuracy of the exercise is getting lower and lower, this visualization shows that the system can be based on the learner's answer, for the This visualization shows that the system can recommend appropriate exercises for learners based on their answers, i.e., exercises with appropriate difficulty.

(2) When the average accuracy of an exercise is high, but the learner answers it incorrectly, the system will recommend an exercise with comparable accuracy again. If the answer is still wrong the second time, it means that the exercise is the student's weak knowledge concept. If both answers are correct, the previous error may be due to a mistake or other reasons.

(3) When the average accuracy of an exercise is low and the learner answers it incorrectly, the question may be a difficult one for the learner. In conclusion, from the learners' case study of adaptive learning paths for multi-scenario teaching of English in the higher vocational railroad industry, the systematic recommendation of exercises based on the method of this paper is able to locate weak knowledge point concepts, diagnose learning difficulties, and is able to recommend exercises of appropriate difficulty for the learners.

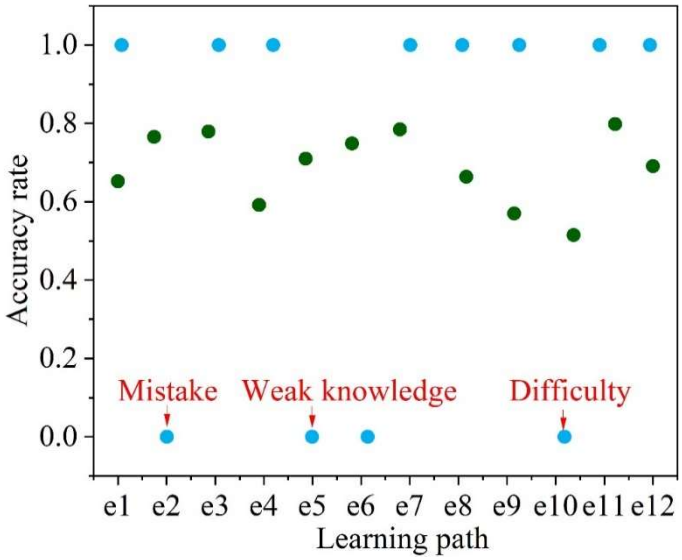


Figure 7: Adaptive learning path case analysis

V. Conclusion

Experimental results on different datasets show that the learning path recommendation system based on reinforcement learning outperforms traditional recommendation models in terms of performance. On the Skill-builder-data-2023-2024 dataset, the AUC value of this paper's model reaches 0.83837 with an accuracy of 0.75654, which is significantly higher than the BKT and DKT models. On the ADM-2018 dataset, the AUC of this paper's model is 0.7691 and the accuracy rate is 0.69723. By comparing with traditional methods such as KNN, GRU4Rec, NARM and DQN, the model proposed in this paper shows a large advantage in both path recommendation efficiency and accuracy.

In addition, the model in this paper is able to continuously adjust the recommendation strategy according to the dynamic feedback of learners, which greatly improves the degree of personalization and adaptability of learning. By accurately analyzing students' historical learning data, the system is able to recommend the most appropriate learning content for students and avoid over- or under-learning in the process. Therefore, the method in this paper not only provides a new learning path optimization idea in theory, but also provides an effective teaching strategy

optimization tool for multi-scenario teaching of English in the railroad industry in higher vocational colleges and universities in practice.

References

[1] Kobeleva, E., Stuchinskaya, E., & Dushinina, E. (2023). Development of professional foreign language competence within transport industry requirements. In E3S Web of Conferences (Vol. 402, p. 08031). EDP Sciences.

[2] Puspitasari, A., Fikria, A., Marsusiadi, E. N. S., & Widiastuti, S. (2024). English Module for Railway Engineering: a Need Analysis. JP (Jurnal Pendidikan): Teori dan Praktik, 9(2), 183-194.

[3] Tulkinovna, D. G., & Gulomovna, M. S. (2025). Needs Analysis for ESP: Enhancing English Proficiency Among Students at Tashkent Institute of Railway Engineers. American Journal of Philological Sciences, 5(04), 292-307.

[4] Candra, B., & Khoiriyah, K. (2024). Cultivating English proficiency for enhanced customer service in Indonesian railways: A study on vocational training needs. English Review: Journal of English Education, 12(1), 397-408.

[5] Tividad, M. J. (2024). The English language needs of a technical vocational institution. Mary Joy Tividad (2024). The English Language Needs of a Technical Vocational Institution. Psychology and Education: A Multidisciplinary Journal, 16(3), 256-275.

[6] Wang, H. (2020). Teaching strategies for improving English effectiveness in higher vocational colleges based on the present situation of English teaching. Learning & Education, 9(4).

[7] Wu, D., Wang, S., Liu, Q., Abualigah, L., & Jia, H. (2022). An Improved Teaching-Learning-Based Optimization Algorithm with Reinforcement Learning Strategy for Solving Optimization Problems. Computational Intelligence and Neuroscience, 2022(1), 1535957.

[8] Wang, F. (2018). Reinforcement learning in a pomdp based intelligent tutoring system for optimizing teaching strategies. International Journal of Information and Education Technology, 8(8), 553-558.

[9] Fahad Mon, B., Wasfi, A., Hayajneh, M., Slim, A., & Abu Ali, N. (2023, September). Reinforcement learning in education: A literature review. In Informatics (Vol. 10, No. 3, p. 74). MDPI.

[10] Szepesvári, C. (2022). Algorithms for reinforcement learning. Springer nature.

[11] Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). Deep reinforcement learning: A brief survey. IEEE Signal Processing Magazine, 34(6), 26-38.

[12] Ding, Z., Huang, Y., Yuan, H., & Dong, H. (2020). Introduction to reinforcement learning. Deep reinforcement learning: fundamentals, research and applications, 47-123.

[13] Zhang, F. (2021). THE APPLICATION OF EDUCATIONAL PSYCHOLOGY IN THE REFORM OF CLIL BILINGUAL TEACHING MODE IN RAIL TRANSIT SPECIALTY GROUP IN HIGHER VOCATIONAL EDUCATION. Psychiatria Danubina, 33(suppl 6), 115-116.

[14] Li, T. (2025). Optimization of Reinforcement Learning in Personalized Teaching Mode of College English Classroom under the OBE Concept. International Journal of New Developments in Education, 7(3).

[15] Jayaranjan, M., Shekhar, G. R., Rafi, S., Ramesh, J. V. N., & Kiran, A. (2025, January). Optimizing English Learning with AI: Machine Learning Techniques and Tools. In 2025 International Conference on Intelligent Systems and Computational Networks (ICISCN) (pp. 1-6). IEEE.

[16] Lianmei Deng. (2023). A Study on Distance Personalized English Teaching Based on Deep Directed Graph Knowledge Tracking Model. International Journal of Advanced Computer Science and Applications (IJACSA),14(2).

[17] Ni Wang. (2024). Multimodal robot-assisted English writing guidance and error correction with reinforcement learning. Frontiers in Neurorobotics,18,1483131-1483131.

[18] Maqsood Shazia,Shahid Abdul,Nazar Fakhra,Asif Muhammad,Ahmad Muhammad & Mazzara Manuel. (2020). C-POS: A Context-Aware Adaptive Part-of-Speech Language Learning Framework. IEEE Access,8,30720-30733.