J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 97--112

Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Research on the Optimization Path of Multi-scene Teaching Strategies for English in Higher Vocational Railway Industry Based on Reinforcement Learning

Yao Li[1, ✉]

[1] *Hunan Technical College of Railway High-speed, Hengyang, Hunan, 421002, China*

ABSTRACT

The international development of the railroad industry puts forward higher requirements for the English application ability of senior railroad students, and reinforcement learning provides new ideas for the optimization of their teaching strategies. Based on reinforcement learning, the article constructs an adaptive learning path recommendation model (RL4ALPR). The model achieves application learning of multi-scenario knowledge of English in the railroad industry through railroad English knowledge level modeling, candidate learning item screening, recommender modeling, and reward calculation. The recommended effective value of the model in this paper is 0.581 at a learning path length of 60, which is 7.79% to 13.70% higher than the control model. The model realizes accurate recommendation of English exercises for the railroad industry based on the answers to the exercises. The evaluation scores of the students in the experimental class under the intervention of the model in this paper are improved to 24.26, 17.50, and 19.64 for speaking, reading comprehension, and translation of English in the railroad, respectively. Under the model of this paper, English teaching in the higher vocational railroad industry is highly recognized by students in terms of “content setting”, “teaching quality” and “teaching effect”. And the experimental class is better than the control class in terms of the level of knowledge about English for the railroad industry, the application of English for the railroad industry in multiple scenarios, and the comprehensive ability evaluation scores of 4-5 points more than the control class.

Keywords: Reinforcement Learning; Markov Decision Making; Q Learning Algorithm; RL4ALPR; English for Railway Industry

1. Introduction

✉ Corresponding author.
E-mail address: ly380512560@126.com (Y. Li).

Received 07 January 2024; Revised 02 August 2024; Accepted 21 November 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-007

© 2025 The Author(s). Published by Combinatorial Press. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

The content of basic English in the higher vocational level is more in-depth than that in secondary school, and it also plays an extremely important role in undertaking the learning part of professional English, which also plays a great role in the future development of the students [1-2]. Therefore, English teachers should fully recognize the importance of English teaching for railroad students and innovate the way of education and teaching to improve the situation of English teaching for railroad students [3-5].

From the actual situation of English teaching in railroad majors, based on the fact that most of the student groups have low scores in the college entrance examination, the English foundation is weak [6-7]. Coupled with the improper teaching methods of teachers, which can not attract students' attention to learning, and can not achieve the purpose of improving students' English proficiency, these are the problems that need to be solved by the team of English teachers in higher vocational colleges and universities at this stage [8-10]. It is necessary to strengthen the attention to enhance the teachers' teaching methods and students' learning interest, so as to strengthen the overall English level [11].

The teaching goal for railroad students is to cultivate applied talents with the ability of production, construction and management services for China's railroad business, so that the students will have good moral character and solid professional skills [12-13]. At the present stage, Chinese students in the class of passenger service will face a large number of foreign friends, and at the same time, there will be a large amount of English in the relevant materials of many new equipment for railroad transportation, which requires students of railroad majors to have strong communicative ability and reading ability in English, and promotes the common enhancement of the students' overall language ability [14-16].

Teachers should not only teach railroad students basic English, but also combine the relevant contents of railroad transportation, such as the contents of the dialogue of railroad passenger service, about railroad engineering construction, railroad transportation equipment and technology [17-19]. Therefore, teachers need to cultivate students' ability to read, translate and understand scientific articles related to railroad transportation, so that students can improve the language ability of railroad professional English on the basis of their daily English expression ability, deepen students' learning of English knowledge, cultivate students' English literacy and professionalism, broaden students' knowledge, and enhance students' cultural awareness [20-23].

The study incorporates a deflationary dot product attention mechanism on the basis of reinforcement learning algorithms to design a learning path recommendation (RL4ALPR) model as a multi-scenario teaching optimization strategy for English in the higher vocational railroad industry. The key of the model to achieve teaching optimization lies in SAKT, CN, A2C and reward computation, and the interaction embedding matrix with position encoding is obtained by SAKT, which is used to compute the learner's correct answer probability and improve the path recommendation effectiveness. The study designed a series of experiments, using a combination of simulation and practical experiments, to evaluate the application performance of the teaching strategy and its effectiveness in enhancing learners' English proficiency in the railroad industry.

2. Optimization Strategies for Multi-Scenario Teaching of English in the Railway Industry under Enhanced Learning

The application of reinforcement learning in multi-scenario teaching of English in the railroad industry in vocational high schools is reflected in personalized learning, intelligent tutoring, optimization of teaching resources, teaching assessment and feedback, and generation of teaching materials. In this section, the multi-scenario learning path recommendation model (RL4ALPR) based on reinforcement learning will be introduced in detail, aiming to optimize the teaching of multi-scenarios of English in the railroad industry by using this model.

2.1. Learning path recommendation

2.1.1. Enhanced learning. The structural model of reinforcement learning is shown in Figure 1, and the specifics of each element of reinforcement learning [24] are as follows:

Intelligence: an independent entity that is capable of thinking and can interact with the environment can then be abstracted as an intelligence.

Strategy: defines how an intelligent body behaves at a given moment. A strategy is a mapping from the currently perceived state of the environment to the action taken in that state.

Return Signal: defines the goal of the reinforcement learning problem. The feedback that the environment gives to the intelligent body at each time step is called the payoff.

Value Function: Unlike immediate payoffs, a value function portrays a preference for a state or behavior over a long period of time.

Environment model: the environment model is like an emulator, given a state and an action, the model predicts the next state and payoff resulting from that action.

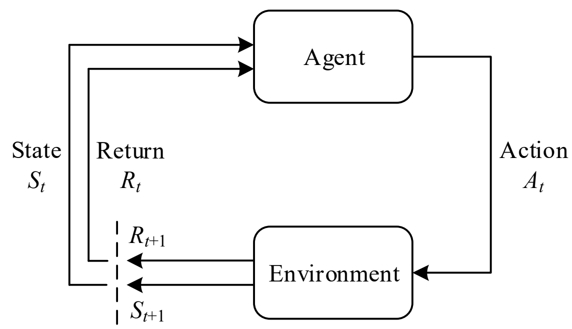

Fig. 1. Strengthen the learning structure model

2.1.2. Specific recommendation process:

1) Based on the learner's history of interactions X use a knowledge tracking algorithm to detect the learner's initial mastery E_{start} of the learning objectives, and set the learning objectives $T=\{t_1,t_2,\ldots\}$ to the learner.

2) At each time step t in the recommendation process, the state $s_t = y_t \oplus T$ consisting of the learner's knowledge level y_t and the learning objective T is used as an input to the reinforcement learning intelligences, and the set of candidate learning items filtered using the cognitive navigation algorithm based on the learning items k_{t-1} that have been recommended in the previous time step.

3) The learner answers k_t and gets the answer $score_t$, at which point a newly generated interaction record $f_t = \{k_t, score_t\}$ is available.

4) The interaction record f_t of each time step and all the historical learning records of the learner t before the time step together form a new interaction record group L, which L is inputted into the knowledge tracking model to get the status s_{t+1} and reward r_t of the next time step.

5) At the end of the path recommendation, the learner's final mastery of the learning target E_{end} is detected, which is used to calculate the learner's knowledge level improvement index IKL for the target learning item.

2.2. The RL4ALPR model

The general framework of the RL4ALPR model is shown in Figure 2. Assuming that the length of the learning path is N, i.e., the recommendation is divided into a total of N time-steps, the model details are next explained in terms of the t time-step.

RL4ALPR is divided into the following four sub-modules: knowledge level modeling (SAKT), candidate learning item screening (CN), reinforcement learning recommender (A2C), and reward computation.

At time step t, a state vector $s_t = y_t \oplus T$ consisting of the learner's knowledge level y_t, and the learning objective T serves as the input to the strategy network $\pi(a \mid s;\theta)$. The role of strategy network $\pi(a \mid s;\theta)$ is to act as an actor to recommend a learning item k_t based on the current strategy, and its output is a vector with each dimension of the vector indicating the probability of each learning item being selected. Then a learning item is randomly selected from the set of candidate learning items screened by CN according to the $\varepsilon - greedy$ policy k_t. The learner answers to k_t generating a new interaction record, and the environment generates a new level of knowledge y_{t+1} and a new state s_{t+1} based on the learner's interaction record. value network $v(s;w)$ acts as a critic for evaluating the action, i.e., the recommended learning item, and has as its input a vector s_t, with the output an scalar representing the predicted future cumulative discounted rewards for evaluating how good or bad the current recommended learning item k_t is and for improving the strategy for selecting the recommended learning item $\pi(a \mid s;\theta)$. The degree of change in the learner's knowledge mastery level at the t time step is considered as a reward r_t and is used for updating the strategy network $\pi(a \mid s;\theta)$, and the parameters θ and w of the value network $v(s;w)$.

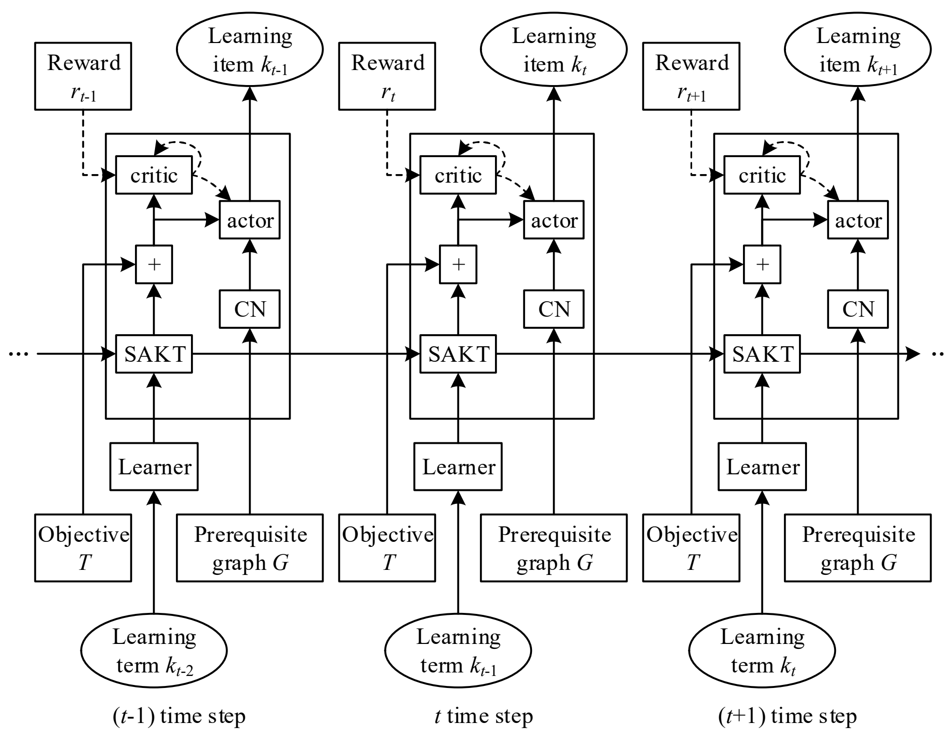

Fig. 2. The overall framework of RL4ALPR model

2.2.1. Knowledge level modeling. Firstly, each interaction tuple $x_i = \{e_i, score_i\}$ in X is operated $c_i = e_i + score_i \times E$ to transform the tuple into numerical form to obtain the input sequence $C = \{c_1, c_2, \ldots c_t\}$, where E is the number of exercises. The input sequence C is passed through the interaction embedding matrix $M \in R^{2E \times d}$, the position encoding matrix $P \in R^{n \times d}$, and the learning term embedding matrix $E \in R^{E \times d}$, respectively, to obtain the interaction embedding matrix $\hat{M} \in R^{n \times d}$

with position encoding and the learning term embedding matrix $\hat{E} \in R^{n \times d}$, where d is the hidden dimension, and n is the maximum sequence length that can be handled by the model, and if the length of C is greater than n, then C is divided into t/n sub-sequences, and if the length of C is less than n, iteratively in the left side of C fill the interaction x_i used for padding.

The embedding matrices $\hat{M}$, $\hat{E}$ are fed into the self-attention layer based on the deflated dot product attention mechanism. The deflated dot product attention mechanism is shown in Eq:

$$Attention(Q,K,V) = soft\max\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{1}$$

In this, the query and key-value pairs are obtained using equation (2):

$$Q = \hat{E}W^Q, K = \hat{E}W^K, V = \hat{M}W^V \tag{2}$$

W^Q, W^K, and $W^V \in R^{d \times d}$ are the query, key, and value projection matrices, respectively, which project the corresponding vectors to different spaces. The multi-head self-attention mechanism performs h linear projections as in Eq:

$$Multihead\left(\hat{M}, \hat{E}\right) = Concat\left(head_1, head_2, \cdots head_h\right) \tag{3}$$

where $head_i = attention\left(\hat{E}W_i^Q, \hat{M}W_i^K, \hat{M}W_i^V\right)$, $W^O \in R^{hd \times d}$.

In predicting the learner's knowledge acquisition level at moment $(t+1)$, only the first t interactions should be taken into account, i.e., when $j > i$, for query Q_i, the key K_i should not be taken into account.

The $S = Multihead\left(\hat{M}, \hat{E}\right)$ obtained from the multi-head self-attention layer is a linear combination of values and the feed-forward network was used to obtain the matrix F as in Eq:

$$F = FFN(S) = \mathrm{Re}LU\left(SW^{(1)} + b^{(1)}\right)W^{(2)} + b^{(2)} \tag{4}$$

where $W^{(1)}$, $W^{(2)} \in R^{d \times d}$, $b^{(1)}$, and $b^{(2)} \in R^d$ are the parameters in training.

Finally, the above matrix F is passed through the Sigmoid activation function to predict the probability of a student answering correctly to a learning item as in Eq:

$$\widehat{SCORE} = Sigmoid(Fw + b) \tag{5}$$

$\widehat{score}_l$ of each dimension value in $\widehat{SCORE}$ denotes the probability that the learner answers e_i correctly, $i \in [2,3 \ldots (t+1)]$, $Sigmoid(z) = 1/\left(1+e^{-z}\right)$, w are the weight matrices, and b is the bias.

The model trains the parameters by minimizing the cross-entropy loss between the predicted value $\widehat{score}_l$ and the actual answer $score_i$ as in Eq:

$$L = -\sum_i \left(score_i \log\left(\widehat{score}_l\right) + \left(1 - score_i\right)\log\left(1 - \widehat{score}_l\right)\right) \tag{6}$$

2.2.2. Candidate Learning Item Screening. Due to the complexity of the logical relationships between knowledge points, this paper designs a cognitive navigation algorithm based on a central node on the prerequisite graph for quickly filtering candidate nodes related to the central node. The set of learning items represented by the candidate nodes is the action space of the recommender on which the policy network in the recommender selects a learning item. The filtered set of candidate learning items prevents the recommended learning items from violating the logic of human cognition, and at the same time reduces the search space of the policy function in the intelligent body, which serves to accelerate convergence.

2.2.3. Recommender Modeling. The structure of the A2C algorithm is shown in Fig. 3. The A2C algorithm [25] uses a strategy gradient algorithm with BASELINE to update parameter θ of the strategy function in the actor, which allows the updated strategy to have smaller variance and faster convergence. Parameters w of the value function in the criterion are updated using the TD algorithm, and the parameters are usually updated using the TD algorithm with Multi-Step TD Target because of the large noise generated by the single-step update.

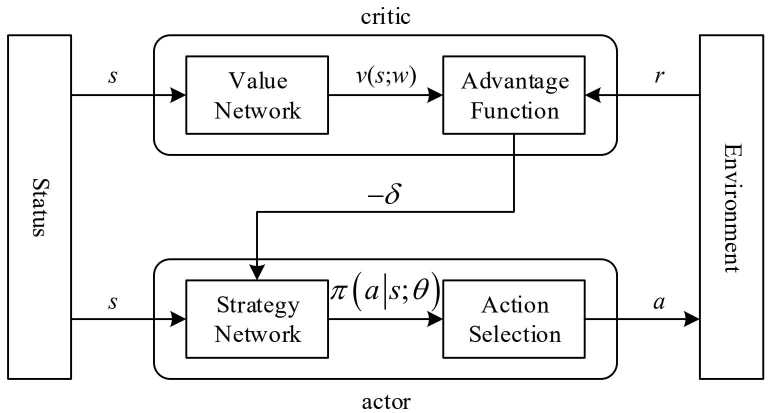

Fig. 3. A2C structure

Policy Gradient with Baseline [26] is the method used in reinforcement learning to update the parameters θ in the policy network $\pi(\cdot|s;\theta)$. Baseline is denoted as b and b is usually an expression that does not depend on any action A under which $\mathrm{E}_{A\sim\pi}\left[b\cdot\frac{\ln\pi(A|s;\theta)}{\partial\theta}\right]=0$ holds.

The TD algorithm is the method used in reinforcement learning to update the parameters w of the value network $v(s:w)$. The modeling idea of the TD algorithm is:

$$\begin{aligned} Model's\ Estimate\ Time\ T_{total} &= Actual\ Time\ T_1 \\ &+ Estimate\ Time\ T_2 \end{aligned} \tag{7}$$

The A2C algorithm, at each moment t during the learning path recommendation process, after observing the current state s_t, according to the policy network $\pi\left(\cdot|s_t;\theta_t\right)$, the intelligent body uses a $\varepsilon-greedy$ policy to randomly sample an action a_t. i.e., a learning term k_t, the learner answers to k_t, the environment gives a new state s_{t+1} and a reward r_t, and a trajectory $\left(s_t,k_t,r_t,s_{t+1}\right)$ is obtained at the moment of t. Afterwards, s_{t+1} will serve as an input to $\pi\left(\cdot|s_{t+1};\theta_t\right)$ recalculate the distribution of the new action Probability.

2.2.4. Calculation of incentives. After the execution of each action in the reinforcement learning algorithm, the environment provides an immediate reward r_t to guide the generation of the next action. At each time step in the recommendation process, the reward function is set as follows:

$$r_t = \Sigma\left(y_{t+1,k} - y_{t,k}\right) \tag{8}$$

where k denotes the index of the knowledge point included in learning objective T, and $y_{t,k}$ denotes t the learner's mastery of the knowledge point k at the moment.

2.3. *Evaluation of Learning Path Recommendation Effect of RL4ALPR Modeling*

2.3.1. Recommended performance test analysis. A simulation environment is constructed to compare the performance of the RL4ALPR model and commonly used recommendation algorithms for recommending learning paths of different lengths on the Mynereus dataset, in order to validate their effectiveness and feasibility in online learning path recommendation applications. The Mynereus dataset collects records of interactions on online learning platforms for students from many countries around the world, covering millions of learning events and learning trajectories, including answering questions, watching English learning videos, and participating in discussions.

Meanwhile, in order to evaluate the performance of the RL4ALPR model proposed in this paper in learning path recommendation, KNN (K-Nearest Neighbors Recommendation), LSTM-CNN (Long Short-Term Memory-Convolutional Neural Network Recommendation), LSTM-CNN (Long Short-Term Memory-Convolutional Neural Network), GRU4ReC (Gated Recurrent Unit for Recommender), and RL4ALPR Learning Path Recommendation algorithms are used as reference models for comparison. To validate the effectiveness of recommended learning paths through simulated learning, this paper adopts learning stage effectiveness E_p as an evaluation index to validate the effectiveness of the RL4ALPR model in learning path recommendation. The calculation formula is as follows:

$$\begin{aligned} E_p &= \frac{E_e - E_s}{E_{\sup} - E_s} \\ E_s &= \frac{1}{n}\sum_{t \in T} y_t \\ E_e &= \frac{1}{n}\sum_{t \in T} y_t{}' \end{aligned} \tag{9}$$

Where E_s is the mastery level of the learning objectives at the beginning stage of learning, E_e is the mastery level of the learning objectives at the end stage of learning, and $E_{\sup}$ is the maximum value of the mastery level of the learning objectives.

In the experiment, the historical learning records of students' English in the railroad industry were collected based on the online learning platform and divided into 3 parts: the first 70% was used for training and modeling the knowledge level, the middle 15% was masked, and the last 15% was used for target selection. In order to evaluate the effectiveness of each model based on the recommendation of learning paths of different lengths, the lengths of learning paths were set to 5, 15, 30, 45 and 60 in the experiment.

Figure 4 shows the effectiveness of each model based on different length learning path recommendation. From the figure, it can be seen that the Ep value of learning path recommendation of each model rises with the increment of learning path length. However, the RL4ALPR model in this paper has the highest Ep value for the effectiveness of learning path recommendation for different lengths on the Mynereus dataset. In particular, the effect is especially prominent when the learning path length is 60, with an Ep value of 0.581, which is 7.79% to 13.70% higher than the comparison model. This may be due to the fact that the recommendation of the comparison model tends to keep recommending similar exercises, which leads the learning path into some kind of localization. In contrast, the RL4ALPR model takes advantage of the potential correlation between students' multidimensional learning features, combines the time interval factor and the attention mechanism, solves the feature sparsity problem, optimizes the computational performance, and improves the prediction accuracy.

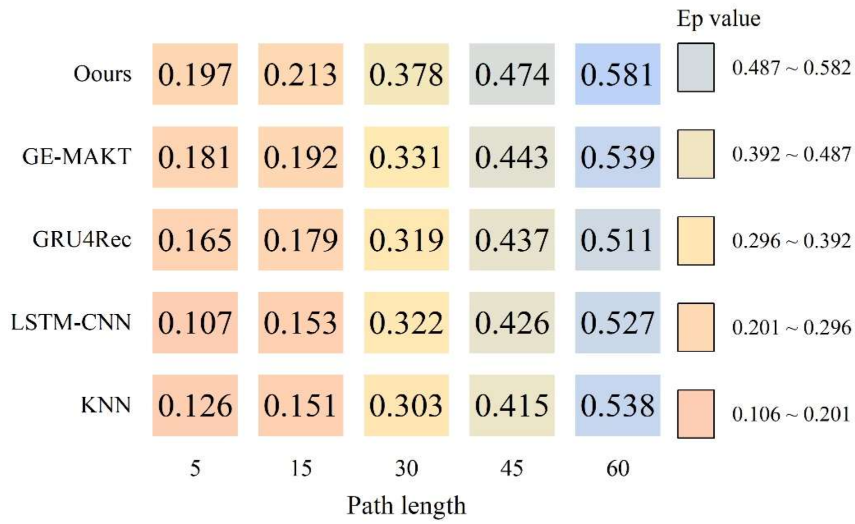

Fig. 4. The models are based on the effectiveness of different length learning paths

In conclusion, the RL4ALPR model verifies its superiority in learning path recommendation by simulating students' learning level changes in a simulation environment, which is a significant improvement over traditional methods.

3.3.2. Learning Path Recommendation Case. In this study, the adaptive learning path of a randomly selected learner of English in the railroad industry in a higher vocational college was selected for case study, and the trend of this learner's English learning in the railroad industry is shown in Figure 5. In the figure, the horizontal axis indicates the accuracy rate, the vertical axis indicates the path recommended by the exercises, the yellow dash line indicates the average accuracy rate of all learners who answered this exercise in the model, and the blue pentagonal symbols indicate the student's answers in each exercise. If the answer is correct, y = 1, and if the answer is incorrect, y = 0. As can be seen in the figure:

1) From the trend of the accuracy rate, the accuracy rate of the English exercises in the railroad industry recommended by the model in this paper for the learner is roughly wavy, when the student answers incorrectly to an exercise, the model recommends exercises with higher accuracy rate for the learner, and when the learner answers correctly to an exercise, the accuracy rate of the exercises recommended by the model for the learner is getting lower and lower, and this visualization shows that the model can recommend exercises for the learner according to the learner's This visualization shows that the model 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 model will recommend an exercise with comparable accuracy again. If the second answer is still wrong, it means that the exercise is the student's weak knowledge point concept, if both answers are correct, the previous error may be caused by mistakes and 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.

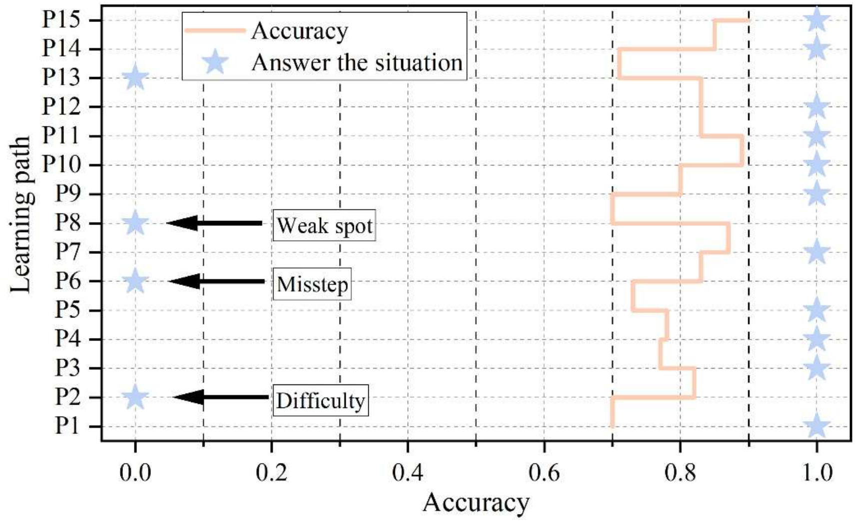

Fig. 5. The trend of English learning in railway industry

3. Practical Research on Optimizing Multi-scenario Teaching Strategies for English in Higher Education Railway Industry

3.1. *Subjects of study*

The subjects of the study are the students of the sophomore senior class 3 and class 5 of railroad management major in a university in a city, with class 3 students as the control class with 30 students and class 5 as the experimental class with 30 students. The experimental group adopts the teaching mode equipped with the online learning platform modeled in this paper, while the control class adopts traditional teaching. There is no significant difference between the English college entrance examination scores of the students in the two classes at the time of enrollment, and their English bases are similar. In order to eliminate the bias caused by different teachers' teaching, one teacher was fixed as the classroom teacher. At the same time, the teaching objectives, teaching contents, teaching requirements and teaching time of the two classes were identical.

3.2. *Evaluation tools and methods*

The evaluation utilizes a combination of quantitative and qualitative research methods to evaluate the effectiveness of effective teaching in the English classroom. Specific steps include:

3.2.1. Evaluation of the quality of teaching. The evaluation form was designed according to the purpose of the study. The table is divided into three dimensions: "content setting", "teaching quality" and "teaching effect". "Content Setting" consists of three items: "Teaching in English", "Organized Content", and "Integration with Care" (Question Nos. 1-3). "Teaching quality" includes four items: "Lectures are simple and easy to understand", "Lectures are passionate, lively and interesting", "Teacher-student relationship is harmonious", and "Combining different teaching strategies" (Question No. 4-7). "Teaching effectiveness" includes 4 items (question numbers 8-11) (question numbers 8-11), including "I can understand the teaching content", "my English reading and writing skills have improved", "my English listening and speaking skills have improved", and "my cooperation and communication skills have improved". The corresponding options for each item are "completely, mostly, sometimes, mostly, mostly, not at all", and the scores are 5 points, 4 points, 3 points, 2 points, and 1 point. This form is distributed to students in the experimental class at the end of the English course for the railway industry. The evaluation form was used to analyze the students' views on the English teaching in this semester and whether they had learned something, and whether there was any difference between the results of the two classes.

3.2.2. English language tests. The teaching cycle of English for the Railway Industry in this semester is 16 weeks, with a total of 20 lessons and 40 class hours. The first test and the second test were conducted at the beginning and the end of the semester respectively, and the two tests were of the same type, including oral English for the railroad industry (40 points), reading comprehension (30 points) and translation (30 points), with a total score of 100 points. Based on the test scores, we compared and analyzed whether there was any difference in the teaching effectiveness of the two classes.

3.3. Analysis of practice results based on teaching experiments

3.3.1. Assessment of teaching performance:

1) Pre-test. In this section, using the data obtained from the railroad program of a higher vocational school in a city in 2023 as a support, data on the English test scores of the railroad industry in 10 classes in the second year of college (including speaking, reading comprehension, and translation) were collected, as well as data on the gender ratio, age composition, and number of students of the control and experimental classes were collected. Through the summary analysis of the previous stage of the railroad industry English test scores of 10 classes in the second year of college. The mean and standard deviation of the data of the English quiz scores of the railroad industry of the 10 classes were obtained as shown in Table 1. After comparative analysis it can be seen that the mean as well as the standard deviation values of the initial scores of class 3 (15.46, 10.87, 13.74) and class 5 (15.53, 10.68, 13.55) are similar, and class 3 and class 5 can be approximated as statistically parallel groups.

Table 1. The average and SE of the English test results in the railway industry

Class	Oral English (40 score)	SE	Read comprehension (30 score)	SE	Translate (30 score)	SE
1	23.84	5.1	15.33	5.2	11.5	4.25
2	17.06	5	12.57	5.18	18.67	4.56
3	15.46	5.39	10.87	4.78	13.74	6.99
4	23.74	7.17	12.08	6.73	15.27	5.1
5	15.53	5.2	10.68	4.61	13.55	6.91
6	28.03	6.31	19.72	4.61	15.46	6.68
7	19.02	7.16	12.38	4.22	19.4	5.32
8	29.05	5.28	15.26	4.12	16.84	6.15
9	27.61	5.8	11.47	6.39	14.42	4.04
10	21.37	6.49	19.79	6.23	19.24	4.65

The gender ratio, age composition and number of students in classes 3 and 5 were analyzed. The irrelevant variables of experimental and control classes were avoided to have an impact on the results of the experiment. The total number of students in class 3 was counted as 30, of which 16 were boys and 14 were girls with an age range of 18-20 years.Class 5 had a total of 30 students, of which 15 were boys and 15 were girls with an age range of 18-20 years. The comparison shows that the difference in the number of students is small, the gender ratio is similar, and the age composition is basically the same.

In order to ensure the rigor of the teaching experiment, further validation of the two experimental subject classes is needed, which is used to determine whether there is a significant difference between the control class and the experimental class, in terms of the level of knowledge of English in the railroad industry before the experiment. Now for the experimental class and the control class of the previous stage of the railroad industry English test scores data, independent samples T-test, performance data processing results are shown in Table 2. According to the results of data analysis in the table, it can be seen that the sig values of the different modules in the table are all greater than 0.05, which indicates that there is no significant difference between the two groups of experimental data, i.e., there is a consistent variance chi-square. In addition, according to the sig (two-tailed) values in the t-test for equivalence of means are all greater than 0.05 and the values are close to 1, indicating that there is a significant uncorrelation between the two groups of performance data. In summary, it can be determined that after the pre-test of the experiment, the two parallel classes selected meet the needs of the experiment and can proceed to the next step.

Table 2. Independent sample t test

Pre-experiment grade		Oral English		Read comprehension		Translate	
		Assumed equal variance	Unassuming equal variance	Assumed equal variance	Unassuming equal variance	Assumed equal variance	Unassuming equal variance
Levine test	F	0.025		0.018		0.028	
	sig	0.793		0.882		0.874	
T test	freedom	153.215	153.117	132.378	132.416	125.630	125.579
	sig (Double tail)	0.997	0.997	0.995	0.995	0.996	0.996

The two classes selected through the experimental pre-test will be randomly assigned, with class 3 as the control class (CK) and class 5 as the experimental class (T). The experimental class will receive the classroom teaching of English for higher vocational railroads under the guidance of the reinforcement learning algorithm based on the study of this paper, while the control class will continue to be taught in the original way and learn the knowledge. The two classes will learn the same chapters and launch an assessment post-test after learning to test the effect of the teaching mode.

2) Back side. After the classroom teaching of English for the railroad industry, post-tests were conducted for students in the experimental class and the control class respectively, which were required to be completed independently online to test students' mastery of English for the railroad industry. The post-test paper was also divided into three parts, i.e. speaking, reading comprehension and translation, with the same score as the pre-test. Figure 6 shows the distribution of the post-test English for the railroad industry scores in the two classes. From the figure, it can be seen that the knowledge level of English for the railroad industry in both classes has improved after teaching, but the experimental class has made more progress, which is significantly higher than that of the control class. Statistically, the average scores of the students in the experimental class in speaking, reading comprehension and translation of railroad English are 24.26, 17.50 and 19.64 respectively, while the average scores of the control class are 16.86, 12.41 and 14.72 respectively.It can be seen from the distribution of the scores in the figure that the experimental class generally has a higher distribution of scores in speaking, reading comprehension and translation than the control class, which makes the two groups of students show a significant difference in their average scores show significant differences (sig < 0.05).

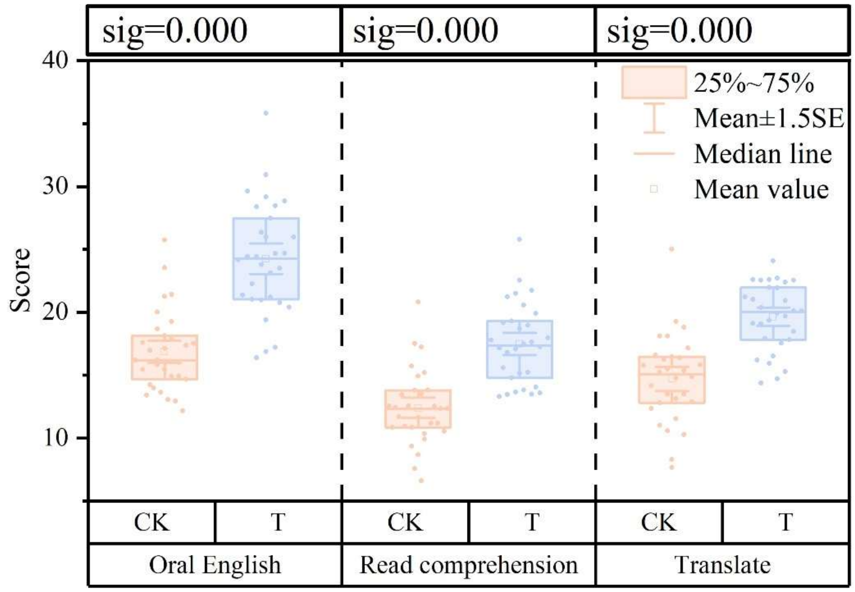

Fig. 6. The distribution of English grades in railway industry after two classes

3.3.2. Research on the quality of teaching. Based on the above theoretical design, this paper verifies the operability and effect of the recommended teaching mode based on reinforcement learning path in English classrooms in the higher vocational railway industry. At the end of the teaching experiment, the teaching quality of English in the railway industry was evaluated for the students in the experimental class. The evaluation indicators are the three levels of "content setting", "teaching quality" and "teaching effect" mentioned above.

Figure 7 shows the results of multi-scenario English teaching quality evaluation in the railway industry based on reinforcement learning algorithm. From the data in the figure, it can be seen that the students in the experimental class have a high degree of recognition of the above three levels. Most of the recognition scores for each question were 4 or 5 points, and only a few students scored 3 points, accounting for less than 13.33%. The statistical results showed that the average recognition of the three levels of "content setting", "teaching quality" and "teaching effect" of the class was 4.29, 4.46 and 4.53, respectively, indicating that the students had a high evaluation of the English teaching quality of the railway industry under the model in this paper. Among them, the "teaching effect" score is the highest, which reflects that under the intervention of the model in this paper, students can be proficient in English vocabulary, sentence patterns and expressions related to the railway industry, which lays a good English foundation for future railway work.

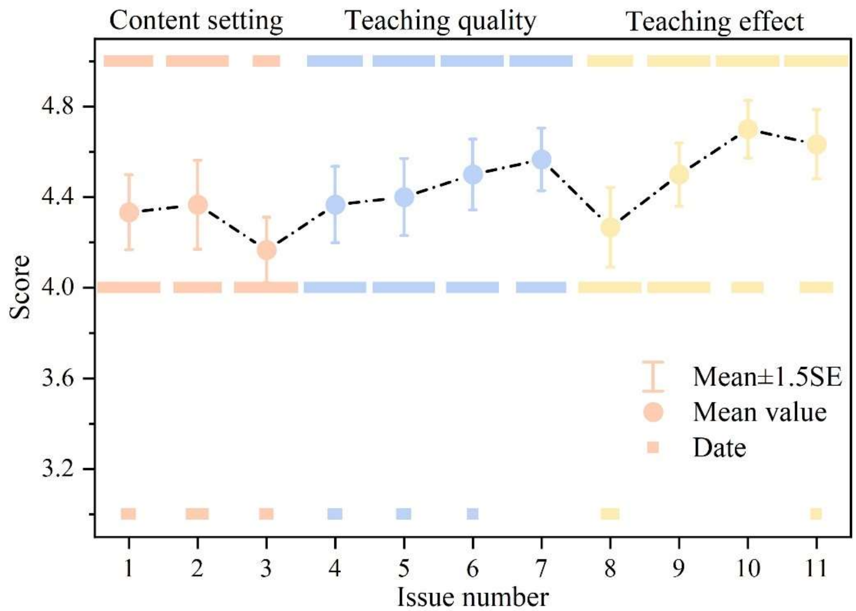

Fig. 7. The results of the evaluation of the teaching quality in the railway industry

In addition, the study made further sub-evaluations of the teaching evaluation of the two classes, which included learning attitudes, learning interests, learning behaviors, cognition of English in the railroad industry, multi-scenario application ability, and comprehensive ability. The same 5-level scale was used to quantitatively analyze the differences between the two classes in terms of evaluation indicators. Figure 8 shows the results of the comparison of the differences between the two classes on the teaching evaluation indicators such as learning attitude. From the figure, it can be clearly seen that the evaluation of learning attitude, learning interest, learning behavior, cognition of English in the railroad industry, multi-scenario application ability, and comprehensive ability in the control class after the end of the teaching is mostly concentrated in 2-3 points, accounting for more than 60%, which indicates that the students in the control class have a lower degree of satisfaction with the evaluation of the teaching of English in the railroad industry in the traditional classroom. In contrast to the experimental class, the evaluation scores of the students in the class are generally in the range of 4-5, and only 1-4 students have scores in the range of 3-4 on each index. This indicates that students show higher interest, attitude and action in learning when teaching English for the railroad industry under the intensive learning algorithm. And the cognitive level of English for the railroad industry, the multi-scenario application of English for the railroad industry and the comprehensive ability are much higher than the control class. That is to say, the teaching optimization strategy based on the reinforcement learning algorithm has a positive impact on students' English learning in the railroad industry.

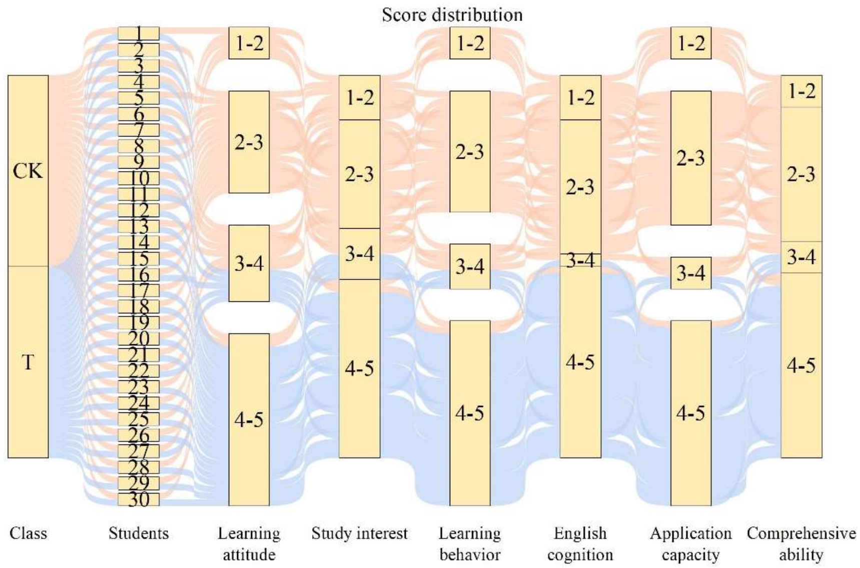

Fig. 8. The difference between the students' teaching and evaluation indicators

4. Conclusion

The research focus of this paper is to optimize the teaching strategy of English in the higher vocational railroad industry using reinforcement learning algorithms, for which an adaptive learning path recommendation model based on reinforcement learning is proposed. The knowledge tracking algorithm is used to model the knowledge of English in the railroad industry in different scenarios, the candidate learning item screening algorithm is used to screen the range of recommended learning items, while the reinforcement learning algorithm is used as a recommender, and the degree of change in the knowledge level of the learner is regarded as the reward of the reinforcement learning recommendation strategy.

The model in this paper has a recommendation effective value of 0.581 when the learning path length is 60, which shows a better learning path recommendation performance. The model can recommend exercises with appropriate difficulty according to the learners' answers to the English exercises in the railroad industry.

Based on the results of the pre-test, classes 3 and 5 were selected as the teaching experimental classes, and the SIG value of the English test scores of the two classes was greater than 0.05, and there was no significant difference. The teaching class using this model improved the railway English speaking, reading comprehension and translation faster, which was 4.92~7.4 points higher than that of the control class. The average scores of "content setting", "teaching quality" and "teaching effect" were 4.29, 4.46 and 4.53, respectively, and the students in the experimental class had higher learning interest and multi-scenario application ability of English in the railway industry.

References

[1] Pratiwi, D. I., Amumpuni, R. S., Fikria, A., & Budiastuti, R. E. (2023). Enhancing students' learning outcomes through MALL in TOEFL preparation class for railway mechanical technology. International Journal of Language Education, 7(2), 185-198.

[2] Hafner, C. A., & Pun, J. (2020). Introduction to this special issue: English for academic and professional purposes in the digital Era. RELC Journal, 51(1), 3-13.

[3] Luo, J. (2019, December). A Study on the Course Design of English for Urban Rail Transit in Independent Colleges from the Theory of ESP— Take the Order Class in Southwest Jiaotong University Hope College

as an Example. In 5th Annual International Conference on Social Science and Contemporary Humanity Development (SSCHD 2019) (pp. 216-220). Atlantis Press.

[4] Callies, M. (2019). Integrating corpus literacy into language teacher education. Learner corpora and language teaching, 245-263.

[5] Pratiwi, D. I. (2021). DIGITAL TECHNOLOGIES IN ONLINE ENGLISH CLASS FOR RAILWAY MECHANICAL STUDENTS DURING COVID19 OUTBREAK. Jurnal Penelitian Sekolah Tinggi Transportasi Darat, 12(2), 56-65.

[6] 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.

[7] Fikria, A., Puspitasari, A., & Lestianingsih, D. (2023). Probing the Needs of Injecting Technical Communication in English Module in Indonesia Railways Polytechnic: A Students' Perspectives. IJET (Indonesian Journal of English Teaching), 12(2), 142-150.

[8] Isaeva, T. E., Bessarabova, O. N., Savchenko, I. V., Shefieva, E. S., & Goryunova, E. R. (2016). Evaluation procedures of the university graduates' readiness for railway professional activities in the English-speaking countries. Indian Journal of Science and Technology, 9(31), 88593.

[9] Fraszczyk, A., Weerawat, W., Marinov, M., & Kirawanich, P. (2019). Exchange of Higher Education Teaching and Learning Practices Between UK and Thailand: A Case Study of RailExchange Courses. In Sustainable Rail Transport: Proceedings of RailNewcastle 2017 (pp. 241-256). Springer International Publishing.

[10] Hallen, D., & Sangeetha, N. (2015). Effectiveness of Mind Mapping in English Teaching among VIII Standard Students. Journal on English Language Teaching, 5(1), 45-50.

[11] Kholmanskikh, Y. S., & Bogatyreva, I. Y. (2023). Challenges Affecting Listening Comprehension in Professionally-Oriented English and the Strategies for Improvement (Railway Engineering). In Sustainable Development Risks and Risk Management: A Systemic View from the Positions of Economics and Law (pp. 543-547). Cham: Springer International Publishing.

[12] Prasitpornkun, N., & Chaturongakul, P. (2016). Needs analysis for English skills of train conductors and guards working at the State Railway of Thailand (SRT). Unpublished master‟ s thesis). Thammasat University, Bangkok, Thailand.

[13] Taghizadeh, M., Basirat, M., & Ejtehadi, A. (2020). THE EFFECT OF EDMODO ON RAILWAY ENGINEERING STUDENTS'ACADEMIC WRITING AND PERCEPTIONS. Iranian Journal of Engineering Education, 22(86), 117-131.

[14] Sharifovna, K. I. (2024, April). RAILWAY-RELATED ABBREVIATIONS IN ENGLISH: A COMPREHENSIVE OVERVIEW. In Proceedings of Scientific Conference on Multidisciplinary Studies (Vol. 3, No. 4, pp. 224-230).

[15] 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.

[16] Widiyanah, I., & Damayanti, A. T. (2017). Glossary Book For Cadets Of Indonesian Railway Academy In English Specific Purpose Program. Jurnal Perkeretaapian Indonesia (Indonesian Railway Journal), 1(1), 63-66.

[17] Pratiwi, D. I., & Ubaedillah, U. (2021). Digital vocabulary class in English for railway mechanical technology. Teaching English with Technology, 21(3), 67-88.

[18] Johnson, I. L., & Banda, F. (2021). Social Structuring of Languages on Commercial Advertising Signs at the Railways in South Africa. In 6th International Conference on Business and Management Dynamics, Book

Publisher, United Kingdom (pp. 378-394).

[19] Pratiwi, D. I., & Waluyo, B. (2023). Autonomous Learning and the Use of Digital Technologies in Online English Classrooms in Higher Education. Contemporary Educational Technology, 15(2).

[20] Vo, Q. P. (2017). Rethinking Intercultural Communication Competence in English Language Teaching: A Gap between Lecturers' Perspectives and Practices in a Southeast Asian Tertiary Context. Journal on English Language Teaching, 7(1), 20-29.

[21] Yekimov, S., Mishenina, T. M., Nianko, V., Ridel, T. M., & Lysytska, O. (2021, September). Application of the project method for teaching a foreign language to future railway transport engineers. In AIP Conference Proceedings (Vol. 2389, No. 1). AIP Publishing.

[22] Mendoza, S. (2017). Cultural and Parental Involvement Affect on English Learners Reading Development. Journal on English Language Teaching, 7(2), 1-12.

[23] 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.

[24] Xingyue Liang,Qiaoyun Wu,Wenzhang Liu,Yun Zhou,Chunyu Tan,Hongfu Yin & Changyin Sun. (2025). Intrinsic plasticity coding improved spiking actor network for reinforcement learning. Neural Networks107054-107054.

[25] Li Xinzhou,Chen Guifen,Wu Guowei,Sun Zhiyao & Chen Guangjiao. (2023). Research on Multi-Agent D2D Communication Resource Allocation Algorithm Based on A2C. Electronics(2),360-360.

[26] Christina Karafyllia,Jiří Kessler,Jana Hudecová,Josef Kapitán & Petr Bouř. (2025). Raman scattering of water in vicinity of polar complexes: Computational insight into baseline subtraction. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy125648-125648.