

<https://doi.org/10.70917/ijcisim-2025-0323>
Article

Adaptive Learning Path Optimization Based on Reinforcement Learning in Chinese Teaching and Learning

Guo Ruifeng^{1,2,*} and Astri Yulia³

¹ Faculty of Education Science, Jining Normal University, Ulanqab 012000, Inner Mongolia, China

² Faculty of Education and Social Science, Universiti Selangor, Batang Berjuntai, 45600, Selangor, Malaysia

³ School of Education, Sunway University, Bandar Sunway, 47500, Selangor, Malaysia.

* Correspondence author: guoruifeng313@163.com

Abstract: In order to improve the teaching effect of Chinese courses, this paper focuses on the construction and assessment of adaptive learning paths, aiming to realize personalized teaching and meet students' diversified learning needs by using artificial intelligence technology. The article establishes a learner model from the learner's individualized needs, and designs a domain knowledge structure model in combination with the knowledge characteristics of the Chinese curriculum. The DINA model is used to diagnose students' mastery of Chinese knowledge points, and then a gated loop unit is introduced to design a knowledge tracking model for exploring students' knowledge level. Combined with the Q-Learning algorithm of reinforcement learning, the adaptive learning path recommendation method for Chinese knowledge was designed, and the data analysis was carried out through application practice. The results show that the performance of students' Chinese knowledge tracking level and path recommendation is better than the current mainstream methods, and the overall language learning performance of the experimental class students in the teaching practice has been improved by 14.69 points, and the mean value of the learning path score is 6.47 points (7-point scale). Therefore, relying on intelligent algorithms to assist the teaching of language courses helps to improve the level of language teaching and also provides guidance for innovative language teaching methods.

Keywords: learner model; DINA model; gated loop unit; reinforcement learning; Q-Learning; adaptive learning paths

1. Introduction

The integration and development of information technology and education system has become the core driving force to promote educational change, and this process not only provides technical support for the modernization of education, but also is an important practical path for the implementation of the national education strategy [1-2]. Chinese, as a core subject in the basic education stage, is of great value in cultivating students' language use ability and cultural literacy [3]. However, current Chinese teaching still faces many practical challenges [4-5], first of all, at the level of knowledge transfer, the content of the textbook is mostly linked by unit themes, and it is difficult for students to integrate fragmented knowledge and form a systematic knowledge network. At the level of grasping the learning situation, it is difficult for teachers to accurately understand the differences in the accumulation of knowledge and thinking styles of different students, which leads to the fact that layered teaching is a mere formality. At the level of ability cultivation, some teachers lack knowledge about the cultivation of students' core abilities such as logical thinking and creative expression. In this context, it is of great significance for teachers to use digital intelligence technology to integrate multidisciplinary knowledge



in their Chinese teaching practice, to construct a knowledge system of the language discipline, and to provide services for students' personalized learning, which is important for improving teaching efficiency and promoting students' personalized learning [6-7].

Personalized teaching arose with the development of computers, and later developed into adaptive learning to take care of individualized learning needs [8]. Adaptive learning emphasizes adjusting the content or mode of instruction according to the characteristics of the learners to achieve truly tailored instruction to improve learning effectiveness and motivation [9]. Scholars generally believe that e-learning is easier to achieve adaptive learning than traditional teaching, in which learning systems and artificial intelligence, which play an important role in the implementation of adaptive learning, will promote the further realization of lifelong learning and personalized learning in future education [10-11].

Adaptive learning can effectively make up for the shortcomings of traditional teaching, and is an important way to help learners achieve personalized learning [12]. Adaptive learning is mainly composed of three parts: measurement model, learning model and recommendation strategy. The measurement model describes the learner's mastery of knowledge, the learning model is the quantification of the relationship between learning materials and the learner's knowledge changes, and the recommendation strategy refers to the selection of appropriate learning materials for the learner based on the existing information [13]. Among them, recommendation strategy is the core part of adaptive learning, how recommendation strategy is constructed directly affects how to provide learners with suitable learning materials, so the research on recommendation strategy is crucial in adaptive learning [14-15]. A good recommendation strategy cannot be separated from the measurement of learners' attribute characteristics and efficient recommendation techniques, and the application of reinforcement learning can solve these two problems well. Reinforcement learning is a branch of machine science, and reinforcement learning emphasizes that learners find the relationship between learning states and learning behaviors by interacting with the environment in order to obtain the maximum cumulative rewards, so that the learners' learning effects can be maximized [16-18]. Reinforcement learning provides theoretical and technical support for the realization of adaptive learning paths [19].

Technology not only promotes the balanced development of education, but also makes the adaptability and personalization of education possible. Based on the learners' needs in Chinese teaching, this paper designs the learner model and domain knowledge model respectively, diagnoses students' mastery of knowledge points using the DINA model, and introduces the GRU model to design the evolutionary knowledge tracking model for assessing students' knowledge level. Based on the reinforcement learning algorithm, the Q-Learning algorithm is combined to construct an adaptive learning path recommendation algorithm for Chinese course knowledge. This study provides a new research perspective for the innovation of Chinese teaching, and also provides a theoretical practice to promote the optimization of adaptive learning paths in the language curriculum.

2. Relevant theoretical and technical foundations

With the rapid development of digital technology, information technology and network technology, the new generation of technological revolution represented by big data, cloud computing, artificial intelligence and so on is profoundly changing social production and daily life style, and at the same time, it also brings new challenges and opportunities to the field of education. It is necessary to give full play to the advantages of new technologies and explore the in-depth integration of online and offline of the teaching mode, which is used to serve personalized learning. The data-driven, intelligent recommendation and customizable curriculum and other educational advantages of the rapidly changing artificial intelligence technology can provide adequate technical support and resource guarantee for personalized education.

2.1. Intensive study of relevant theories

2.1.1. Markov process

Reinforcement learning (RL) is one of the paradigms and methodologies of machine learning, where an intelligent body obtains feedback by interacting with its environment. Specifically, the intelligent body makes an action based on the current state, the environment gives a reward based on the intelligent body's action, and the intelligent body learns a strategy to reach the problem of maximizing the reward or achieving a specific goal.

(1) Markov property

Markov property means that the state of the current moment is only related to the state of the previous moment, and has no relationship with the state of any other moment. Describing Markov property in terms of probabilistic formulas for state transfer is expressed as follows:

$$p(S_{t+1} | S_t) = p(S_{t+1} | S_1, \dots, S_t) \quad (1)$$

(2) Markov process

A Markov process (MP) is a memoryless stochastic process, i.e., the conditional probability of a Markov process is related only to the current state of the system. A Markov process can be represented by a dichotomy $\langle S, P \rangle$, S denoting a finite set of states, and P denoting a matrix of state transfer probabilities, where $p_{ss'} = p(S_{t+1} = s' | S_t = s)$ represents the transfer probability from the current state S_s to the next state $S_{s'}$.

(3) Markov reward process

The purpose of reinforcement learning is to maximize the reward during the interaction with the environment, and it is necessary that the environment has given the reward back to the intelligent body. Therefore, the Markov Reward Process (MRP) adds the reward function R to the Markov process, and the MRP extends the MP into $\langle S, P, R, \gamma \rangle$. Where the reward function R is defined as:

$$R_s = E[R_{t+1} | S_t = s] \quad (2)$$

denotes the expected reward that can be obtained at the next moment S_{t+1} in state S_t at moment t , and the reward discount factor γ denotes the sum of all decaying rewards (discounted rewards) from moment t onwards.

For the whole process, the intelligent body can receive rewards multiple times, so the final reward of the intelligent body is the cumulative reward G_t , denoted as follows:

$$G_t = R_{t+1} + \gamma R_{t+2} + \dots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \quad (3)$$

where G_t is the cumulative reward and the decay coefficient γ takes values between $[0,1]$ and is used to adjust the weights of the intelligence on current and future rewards.

For a Markovian reward process, the value function is defined as the expected reward for the state s , i.e.:

$$\begin{aligned} V^x(s) &= E[G_t | S_t = s] \\ &= E[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots | S_t = s] \end{aligned} \quad (4)$$

(4) Markov Decision Process

Markov Decision Process (MDP), a Markov reward process with decision making, where the next state of the intelligent is not only related to the current state but also to the action currently taken, thus the state transfer probability of the MDP has an additional condition, viz:

$$p_{ss'}^a = p(S_{t+1} = s' | S_t = s, A_t = a) \quad (5)$$

and the MDP can be represented by a quintuple $\langle S, A, P, R, \gamma \rangle$, where A is a finite set of actions and the reward function becomes:

$$R_t = E[R_{t+1} | S_t = s, A_t = a] \quad (6)$$

Intelligent bodies need to select actions based on observations of the environment, so MDP introduces policies that represent the distribution of all possible actions. The policy π is a mapping from each state space and action space to the action probability distribution, denoted as $\pi(a|s)$, $\pi(a|s)$ represents the probability of choosing the action as a if the state is s at the

moment t , which can be expressed as:

$$\pi(a | s) = p(A_t = a | S_t = s) \quad (7)$$

2.1.2. Q-Learning Algorithm

Q-Learning algorithm is a planning algorithm that seeks the optimal policy by updating the value function through continuous iteration, which is an off-policy learning method based on Time Difference (TD), and is also a typical model-free reinforcement learning algorithm that does not need to rely on any environment model. The core idea of the Q-Learning algorithm is to create a table of Q values based on the set of states S and the set of actions A (the vertical coordinate of the Q value table represents the state, and the horizontal coordinate represents the action corresponding to a state). a_t denotes the action corresponding to a state, and then update the $Q(s, a)$ in the Q table, i.e., the state-action value function, based on the reward R and Q estimates after each iteration. When the update of the Q value table is completed, the action corresponding to the maximum value of Q in that state, i.e., the maximum value of the state-action value function, is selected from it to be executed, so that the intelligent body can get the maximum cumulative rewards from the external environment, and thus the optimal strategy can be found. The specific steps of the algorithm are as follows:

- (1) Initialize the $Q(s, a)$ data in the Q-value table. Generally, all Q values are initialized to 0.
- (2) Determine the state s_t at the current t moment.
- (3) Select the action a_t based on the Q-value table and combined with a certain action strategy π .
- (4) Perform the action a_t and then perform a state transfer to migrate to a new state s_{t+1} at the next moment, i.e., the moment of $t+1$, and obtain the corresponding reward R_{t+1} .
- (5) Update the correction $Q(s_t, a_t)$ by iterating the recursive formula. The relevant recursive formula is as follows:

$$Q(s_t, a_t)^{new} = Q(s_t, a_t)^{old} + \alpha \left[R_{t+1} + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)^{old} \right] \quad (8)$$

Where α denotes the learning rate, which satisfies $\alpha \in [0, 1]$, and if α takes a larger value, the fraction of Q-values prior to the retained iteration will be smaller, and the $Q(s, a)$ value will converge faster, but too fast convergence will lead to oscillations. γ denotes the discount factor, which satisfies $\gamma \in [0, 1]$ and represents the degree of influence and importance of subsequent rewards on the current moment, a value of 0 for γ indicates that only short-term rewards at the current moment are emphasized, and a value of 1 for γ implies that more emphasis is placed on distant rewards in the future. $Q(s_t, a_t)^{old}$ is the value function for the t moment before the iteration, while $Q(s_t, a_t)^{new}$ refers to the value function after the iteration, $\max_{a'} Q(s_{t+1}, a')$ then represents the maximum $Q(s, a)$ value corresponding to the state at the next moment, i.e., the $t+1$ moment.

(6) When the smart body reaches the goal state, i.e., the end point or the restricted area, it stops iterating and returns to the start state to start a new cycle. Otherwise, return to step (2) and continue iterating to update the corrected $Q(s, a)$ value.

(7) If the recursive formula is satisfied, it means that the iteration of $Q(s, a)$ values converges and the update of the table of Q values is completed. Then the action corresponding to the largest $Q(s, a)$ value is selected according to the greedy strategy, and the system will be able to obtain the largest cumulative reward, so as to find the optimal strategy to find the optimal path.

2.2. DINA model with GRU

2.2.1. DINA model basis

The DINA model is essentially a mathematical model that can be used to describe the probability of a subject answering an item correctly, and thus functions as a cognitive diagnostic. Therefore, the

DINA model is also often regarded as a cognitive diagnostic model. The full name of the DINA model is Deterministic input, noisy and gate model, which involves only two parameters, is simple and flexible, and can be described by the following mathematical language, namely:

$$P(x_{ij} = 1 | \alpha_i, q_j, \eta_{ij}) = (1 - s_j)^{\eta_{ij}} g_j^{1 - \eta_{ij}} \quad (9)$$

where P is used to denote the probability that subject i can answer question j correctly. Where x_{ij} is the score of subject i on question j , when subject i answers question j correctly $x_{ij} = 1$, otherwise $x_{ij} = 0$. $\alpha_i = [\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{iK}]$ is known as the cognitive attribute vector, and if $\alpha_{ik} = 1$, it means that the subject i mastered the cognitive attribute α_{ik} , and if $\alpha_{ik} = 0$, it means no mastery. $q_j = [q_{j1}, \dots, q_{jK}]$, i.e., the j th row of the Q-matrix, is used to reflect the examination of cognitive attributes by question j . $\eta_{ij} = \prod_{k=1}^K \alpha_{jk}^{q_{jk}}$, by the expression η_{ij} for $0 \leq \eta_{ij} \leq 1$, and if $\eta_{ij} = 1$, it means that the subject i has mastered all the attributes examined in the question j without failing to do so, and if there are attributes that have not been mastered, there will be $\eta_{ij} = 0$. s_j is the miss parameter, which indicates the probability that the subject would have mastered all the examined attributes but answered the question j incorrectly; and g_j is the guess parameter, which is the probability that one has not fully mastered all the attributes examined by j but still answered the question correctly.

Using the DINA model for cognitive diagnosis, two parameters are first estimated based on the students' response data, and the values of the parameters g_j and s_j reflect the reliability of the model and the cognitive diagnostic process. Secondly, the DINA model is used to summarize the attribute mastery patterns of individuals and groups, and the commonly used estimation methods include great likelihood estimation, Bayesian posterior estimation, and great posterior estimation.

2.2.2. Door-controlled circulation units

Gated Recurrent Unit (GRU) is a variant of recurrent neural network, which is an improved network model proposed to solve the problems of classical recurrent neural network (RNN), such as the existence of gradient explosion and difficult to memorize long-term memory. In the GRU model, there are two gates, the update gate and the reset gate, the update gate determines how much state information from the previous moment is sent to the current state, and the reset gate determines how much state information from the previous moment is sent to the hidden layer state.

The role of the reset gate is to control the proportion of state information from the previous moment that goes to the hidden layer, and its expression is:

$$r_t = \sigma(W_r [h_{t-1}, x_t]) \quad (10)$$

Where W_r is the reset gating weight, σ is the sigmoid function, and the final output is mapped to $[0, 1]$, where 0 means that no amount is allowed to pass and 1 means that any amount is allowed to pass, and the larger the value of the output means that the more information about the state of the previous moment enters the hidden layer.

After getting the reset gating signal, reset the information of the last hidden layer to get $\tilde{h}_{t-1} = h_{t-1} * r_t$, at this time, we have discarded some of the unneeded information, and then consider the user's current input information, and reset the information \tilde{h}_{t-1} and the input of the moment x_t are spliced, i.e:

$$\tilde{h}_t = \tanh(W_{\tilde{h}} \cdot [r_t * h_{t-1}, x_t]) \quad (11)$$

Where $W_{\tilde{h}}$ is the hidden layer state weights, $[\]$ denotes the splicing operation (addition), $*$ denotes the dot product operation, and the tanh activation function maps the obtained data to $[-1, 1]$.

The update gate acts to control the proportion of state information from the previous moment that is sent to the current state, and the update gate formula is:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (12)$$

Where W_z is the update gate weight and σ is the sigmoid function that maps the result to $[0,1]$, with 0 indicating that no amount is allowed to pass, and 1 indicating that any amount is allowed to pass, the larger the value of the update gate, the more previous-moment information has been brought into the current moment.

z_t can both forget and select memory, through $z_t * \tilde{h}_t$ can discard part of the information in \tilde{h}_t , through $(1-z_t) * h_{t-1}$ can select part of the information in h_{t-1} , $z_t * \tilde{h}_t$ and $(1-z_t) * h_{t-1}$ form a mutually complementary relationship. If the value of z_t is larger, it indicates that less information is retained in the hidden layer in the previous moment, which is mainly supplemented by \tilde{h}_t information, and if the value of z_t is smaller, it indicates that more information is retained in the hidden layer in the previous moment, which is less supplemented by \tilde{h}_t information. Therefore, the state of the current layer can be represented as:

$$h_t = (1-z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (13)$$

GRU retains some of the state information through the reset gate and update gate functions, so that the important information is preserved after a longer time sequence, which solves the long dependency problem in RNN more effectively.

3. Recommendations for optimizing adaptive learning paths

In the era of big data, the explosive growth of online learning resources, information overload and network confusion have become obstacles to learners' personalized learning, and there are many bottlenecks in regular education that restrict the effective occurrence of personalized learning, such as the diagnosis of learners' learning needs, the search for common learning paths, and the guidance of individualized learning paths. Therefore, planning adaptive personalized learning paths for learners based on reinforcement learning is an effective method and key way to promote their personalized development.

3.1. Problem definition and preparatory knowledge

3.1.1. Modeling Learner and Domain Knowledge

The learner is the main participant and experiencer of adaptive learning, as well as the main body of personalized resource acquisition, so the design of adaptive learning should firstly focus on the characteristics of the learner's personalized needs. In order to more clearly characterize the properties of learner objects in the system, we need to instantiate them, i.e., model the learners. It can be said that the learner model is the foundation of adaptive learning implementation, which is mainly achieved by collecting and processing learner personalization information in real time with the help of user modeling components or third-party agent software. In this study, given the IMS LIP standard, the learner model is characterized in the form of a tetrad in terms of four dimensions: basic characteristics, learning styles, cognitive level, and learning records. Namely:

$$LearningModel = \left\{ \begin{array}{l} BaseInformation, LearningStyle, \\ CognitiveLevel, AccessRecords \end{array} \right\} \quad (14)$$

Among them, BaseInformation is used to represent some basic static information of the learner, LearningStyle represents the learning style of the learner, which can be constructed by drawing on Felder's learning style model, CognitiveLevel represents the current cognitive level of the learner, which can be characterized from the primary, intermediate, and advanced levels, and AccessRecords represents the access records of the learner throughout the learning process. AccessRecords represents the access records of the learner during the whole learning process.

Domain knowledge refers to all the clusters of empirical, theoretical and methodological knowledge units in a professional field, while in the computer world we define it as the knowledge clusters which

are easy to operate in terms of computer storage, system organization and management by using a certain (or a number of) representations to materialize and structure the knowledge for the needs of a specific domain.

For teaching the Chinese curriculum, we characterize the structural model of its domain knowledge as:

$$Object = \{Id, Name, Level, Style, Content, OR\} \quad (15)$$

Among them, Id denotes the unique identification of the Chinese knowledge object, $Level$ denotes the level of difficulty of the Chinese knowledge object, $Style$ denotes the style of bias of the Chinese knowledge object, $Name$ denotes the name of the knowledge object, $Content$ denotes the information of the content of the knowledge object, and OR denotes the set of relationships to which the knowledge object belongs. Based on the above three types of relations, the model of relations between knowledge objects in the Chinese curriculum domain in this study OR can be expressed as:

$$OR = \{< Aid, Bid >, Type, Weight\} \quad (16)$$

Where $Type$ denotes the type of relationship and $Weight$ the weight value of the relationship, the higher the value represents the greater the correlation between the knowledge.

3.1.2. Students' mastery of knowledge points

After completing the modeling of learners and language domain knowledge, in order to further get the students' learning status, this paper introduces the DINA model to diagnose the students' mastery of Chinese knowledge points.

The traditional DINA model models the students by uniting the students' question doing R matrix and the test question knowledge point association Q matrix. Each student P_u is described as a vector of knowledge point mastery $\mathbf{a}_u = \{\alpha_{u1}, \alpha_{u2}, \dots, \alpha_{uK}\}$, where each dimension corresponds to one knowledge point, $\alpha_{uk} = 1$ denotes that student u has mastered knowledge point k , $\alpha_{uk} = 0$ denotes that student u has not mastered knowledge point k .

With the knowledge point mastery vector \mathbf{a}_u of student P_u known, for the test question J_v that student P_u has not answered, one can obtain the potential response of student P_u to test question J_v , i.e.:

$$\eta_{uv} = \prod_{k=1}^K \alpha_{uk}^{q_{vk}} \quad (17)$$

Where $\eta_{uv} = 0$ indicates that the student P_u cannot answer the test question J_v correctly, and if $\eta_{uv} = 1$, it is assumed that the student P_u can answer the test question J_v correctly.

In addition, the DINA model introduces test parameters (error rate and guessing rate) to model the student's responses in the real situation. Specifically, the response of student P_u to test question J_v is represented as:

$$P_v(\mathbf{a}_u) = P(R_{uv} = 1 | \mathbf{a}_u) = g_v^{1-\eta_{uv}} (1-s_v)^{\eta_{uv}} \quad (18)$$

Since the interrelationships between knowledge points are considered to be "connected" in the DINA model, the failure rate s_v is defined as the probability that a student who has mastered all the skills required for a test question will still be unable to answer Question v correctly, and the guess rate g_v is defined as the probability that a student who has not mastered all the skills required for a test question will be able to answer Question v correctly.

The DINA model uses the EM algorithm to maximize the marginal likelihood of Eq. (18) to obtain parameter estimates for $\hat{s}_1, \hat{s}_2, \dots, \hat{s}_V$ and $\hat{g}_1, \hat{g}_2, \dots, \hat{g}_V$. And the knowledge point mastery vector \mathbf{a}_u of a student P_u can be determined by maximizing the posterior probability of the student's test score

to obtain the student's dichotomous knowledge point mastery vector. Then:

$$\begin{aligned}
\hat{\alpha}_u &= \arg \max_{\alpha} (\alpha | \mathbf{R}_u) \\
&= \arg \max_{\alpha} L(\mathbf{R}_u | \alpha, \hat{s}_v, \hat{g}_v) P(\alpha) \\
&= \arg \max_{\alpha} L(\mathbf{R}_u | \alpha, \hat{s}_v, \hat{g}_v) \\
&= \arg \max_{\alpha} \prod_{v=1}^V P(\mathbf{R}_u | \alpha, \hat{s}_v, \hat{g}_v)
\end{aligned} \tag{19}$$

After obtaining the students' knowledge point mastery, it can be combined with the students' knowledge point mastery and the knowledge point examination of the test questions to be recommended as the a priori of the RALP model proposed in this paper to make the prediction of students' answering situation and be used for recommendation.

3.2. Adaptive Learning Path Recommendations

3.2.1. Path Recommendation Framework

Figure 1 shows the model framework of adaptive learning path recommendation (RALP). In the adaptive learning path optimization recommendation for Chinese teaching, the learning path includes the sequence of learning activities and the learning object, and the learning style and knowledge level are two important individual difference characteristics of the learners, which predict the performance of different learning behaviors, and are the important basis for realizing adaptive push.

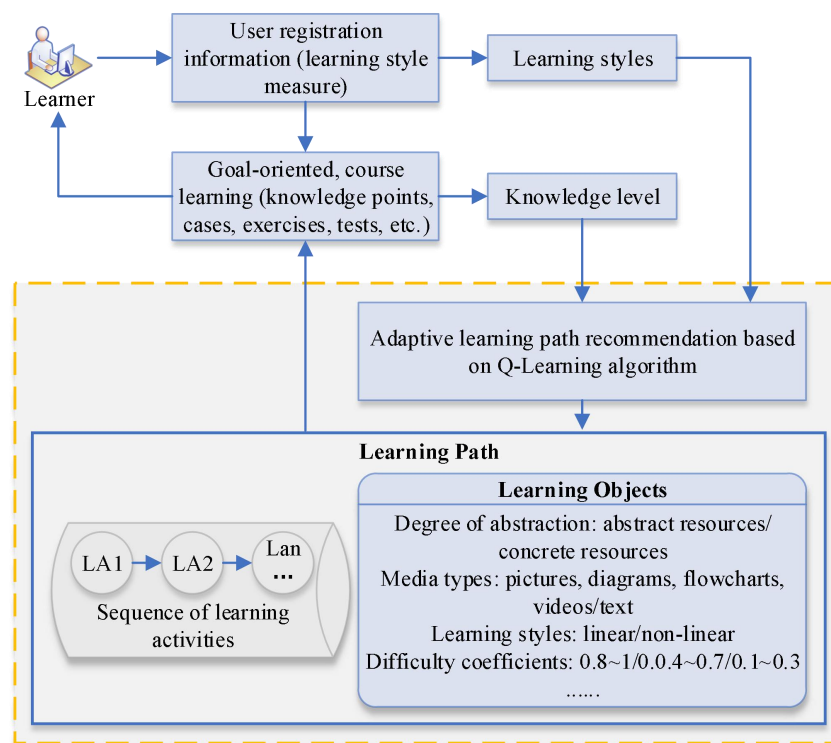


Figure 1. Adaptive learning path recommendation framework

The objectives of adaptive learning path optimization recommendation are as follows:

(1) Learning style determination. Using the explicit subjective determination of the questionnaire survey and the implicit method of mining learning behavior patterns by DINA model to infer the learning style, the combination of the two methods can realize the accurate push of personalized learning path.

(2) Knowledge level estimation. Measuring students' knowledge ability is a constant topic in the field of education, and the knowledge level tends to change over time with the accumulation of learning. Using the knowledge tracking model, we can track and diagnose students' knowledge level in

real time based on learning behavior data (case study time, number of clicks and clicks, question and answer time and number of attempts, etc.), and dynamically present the learning objects in terms of conceptual knowledge comprehension level and difficulty level.

(3) Learning path mining and personalized recommendation. By using the reinforcement learning Q-learning algorithm, the best learning path is mined from the group learning behavior. Meanwhile, personalized push is achieved based on characteristics such as learning style and knowledge level to solve problems such as "learning disorientation" and "cognitive overload", and improve the internal motivation and learning demand for learning.

3.2.2. Evolutionary Knowledge Tracking

In order to realize the adaptive learning path recommendation of Chinese course knowledge, this paper designs the GRU-based Evolutionary Knowledge Tracking model (G-EKT), whose specific framework is shown in Fig. 2, which mainly consists of an embedding layer, a knowledge extraction layer and a knowledge evolution layer.

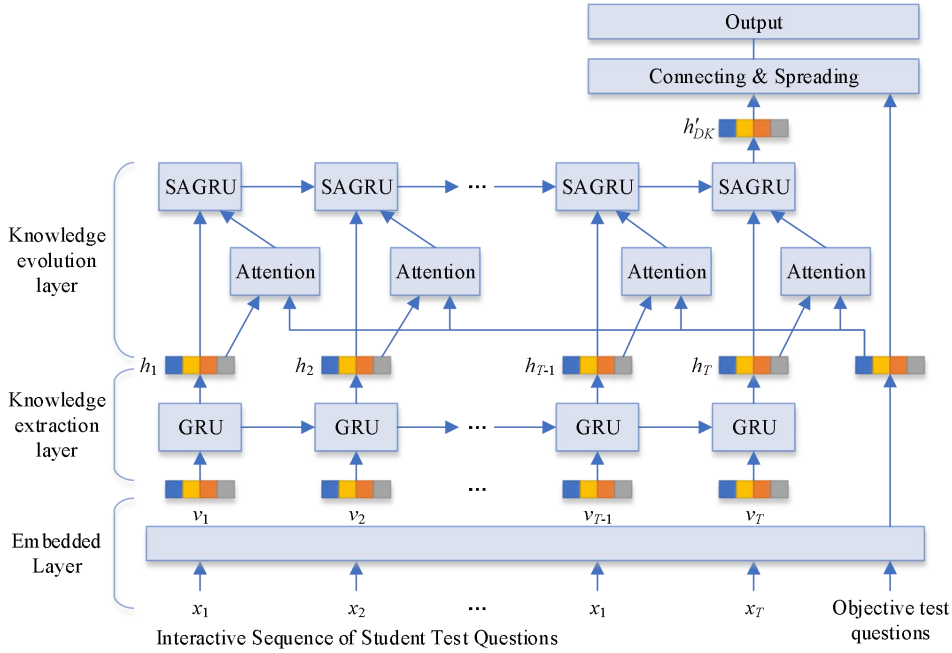


Figure 2. G-EKT Model structure

First, the data of students' test question interaction sequences are input into the embedding layer, which converts the large-scale sparse features into low-dimensional dense features suitable for model input, and sorts the students' test question interactions by time, and converts the unique heat encoding x_i into the embedding vector v_i . Then in order to solve the problem that it is difficult to model the potential knowledge state, the gated recurrent unit (GRU) in deep learning is used in the knowledge extraction layer to model the dependency relationship between the test question interaction sequences, following the principle that the knowledge mastery of the students will directly affect the results of their successive questions. That is, if a student has not mastered a certain knowledge point, he will be continuously wrong when doing that question, and vice versa. For this reason, an auxiliary loss function is introduced to impose additional supervision on the test-question interaction at the current moment, which both provides more semantic information for the knowledge representation of the input data and yields better embedding vectors, and helps to minimize the negative impact of the student's external environment and internal cognition on the representation of the knowledge state, so that the hidden state of the GRU can effectively represent the knowledge state. Next, in order to solve the problem of the influence of the order of test questions on the prediction effect, a gated recurrent unit (SAGRU) with a self-attention updating gate is designed at the knowledge evolution layer to model the knowledge evolution process related to the target test questions, which can calculate the correlation between the knowledge state and the target test questions by means of the self-attention mechanism, dynamically perceive the related knowledge and weaken the influence of the relative order of the questions, so as to guarantee the smooth evolution of the target test questions and to ensure the success of the target test questions. SAGRU can calculate the correlation between the knowledge state and the

target test questions with the help of self-attention mechanism, dynamically perceive the related knowledge, and weaken the influence of the relative order of test questions to ensure the smooth evolution of the target test questions. Finally, the dynamic knowledge states h'_{DK} and the embedding vectors of the target questions are concatenated and fed into a multilayer perceptual machine (MLP) for the final judgment, which outputs in real time the students' mastery of each knowledge point.

3.2.3. Description of the recommendation algorithm

After obtaining the students' mastery of each knowledge point, this paper designs the Q-Learning algorithm based on reinforcement learning for adaptive learning path recommendation of Chinese knowledge. It specifically contains the following links:

(1) Define the state, action, reward and Q-table in the RALP model. Among them, state is the learner's mastery state of the target knowledge point in the learning task at a certain moment, i.e., the level of mastery of the target knowledge point is predicted by the G-EKT model after the intelligent body chooses an action, which takes the value of the number in the interval of $[0,1]$, denoted by s . Actions are the two outcomes of a learner learning a learning item, i.e., mastery or non-mastery, which are regarded as two actions, and the action is denoted by $a(a \in A)$, with a_{i0} indicating that the student has not mastered the i th learning item, and a_{i1} indicating that the student has mastered the i th learning item. Assuming that there are N learning items in a learning task, this corresponds to $2N$ actions. Reward means that the environment from the current state s_j after executing an action a_{i0} or a_{i1} reaches the learning objective of the target knowledge point, i.e., the target state value s_i is reached, and the reward value r is 1; otherwise, r is 0. The reward $R(s,a)$ can be expressed as:

$$R(s,a) = \begin{cases} 1 & \tilde{s} \geq s \\ 0 & \tilde{s} < s \end{cases} \quad (20)$$

where (s,a) is the current state and action, and \tilde{s} is the current state s_j after performing the action.

The RALP model calculates the rewards for different behaviors of an intelligent body by building a Q-table. $Q(s,a)$ denotes the reward received by the intelligent body for performing behavior a in state s . The intelligent body takes the value of $Q(s,a)$ as the basis of the choice of the next behavior. The matrix Q can be represented as:

$$Q = \begin{bmatrix} Q(s_1, a_{10}) & Q(s_1, a_{11}) & Q(s_1, a_{20}) & Q(s_1, a_{21}) & \cdots & Q(s_1, a_{n0}) & Q(s_1, a_{n1}) & \cdots & Q(s_1, a_{n0}) \\ Q(s_2, a_{10}) & Q(s_2, a_{11}) & Q(s_2, a_{20}) & Q(s_2, a_{21}) & \cdots & Q(s_2, a_{n0}) & Q(s_2, a_{n1}) & \cdots & Q(s_2, a_{n0}) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Q(s_j, a_{10}) & Q(s_j, a_{11}) & Q(s_j, a_{20}) & Q(s_j, a_{21}) & \cdots & Q(s_j, a_{n0}) & Q(s_j, a_{n1}) & \cdots & Q(s_j, a_{n0}) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Q(s_m, a_{10}) & Q(s_m, a_{11}) & Q(s_m, a_{20}) & Q(s_m, a_{21}) & \cdots & Q(s_m, a_{n0}) & Q(s_m, a_{n1}) & \cdots & Q(s_m, a_{n0}) \end{bmatrix} \quad (21)$$

Where rows denote states, columns denote actions, and the element $Q(s_j, a_{i0})$ denotes the expectation that the learner will be able to gain by choosing the action a_{i0} in state s_j .

During model training, after performing each action, the model updates the expectation $Q(s,a)$ using reward feedback. The updated expectation value $Q'(s,a)$ is given by:

$$Q'(s,a) = Q(s,a) + \alpha \left[R(s,a) + \gamma \cdot \max_{\tilde{a}'} \{Q(\tilde{s}', \tilde{a}')\} - Q(s,a) \right] \quad (22)$$

where (\tilde{s}', \tilde{a}') denotes the next state and its corresponding behavior, $\max_{\tilde{a}'} \{Q(\tilde{s}', \tilde{a}')\}$ denotes the maximum Q value corresponding to all actions in the \tilde{s}' state, the learning rate $\alpha = 0.1$, and the discount factor $\gamma = 0.9$.

(2) The model is trained by obtaining rewards for the behavior of the intelligent body in different

environments to update the parameters. In this process, the state and the behavior of the intelligent body keep changing, and the behavior exhibited by the intelligent body from a certain initial state to the target state can be regarded as a round, and the next round is entered after a round is completed. Under the state s , the rule for selecting behaviors a is as follows: determine whether the set of Q-values of the behaviors that are not exhibited in the state s is all 0. If it is all 0, then a behavior is randomly selected. Otherwise, 90% probability is to choose the behavior with the largest Q value, and 10% probability is to choose a behavior randomly.

(3) When the model training is finished, the training result Q-Table is outputted for personalized recommendation of learning items within the learning task.

4. Adaptive learning path recommendation application validation

Educational informatization has promoted changes in the concept of education, and education in the information age pays more attention to the individual demands of learners. Promoting adaptive development of learners is one of the basic features of education in the information technology era and an important means of cultivating innovative talents. Therefore, focusing on adaptive learning and promoting learners' individual development has become an important part of future educational development.

4.1. Knowledge Tracking and Recommendation Effectiveness

4.1.1. Predicted effects of student performance

In this paper, experiments are conducted on two real online education datasets to demonstrate the effectiveness of the G-EKT model by comparing its prediction performance with other knowledge tracking models and knowledge tracking performance. The online education datasets mainly include the ASSISTments and Slepemapy.cz datasets, and the average AUC, average ACC, and average RMSE are used as metrics to evaluate the prediction performance. AUC is defined as the area enclosed by the ROC curve and the lower axis, and 50% of the AUC value indicates the prediction performance obtained by random guessing, and a high AUC value indicates that the model has high ACC is the percentage of correctly predicted results over all results, and a high ACC value indicates that the model has a high predictive performance. RMSE is defined as the root mean square error between the predicted value and the true value, and the lower the value of RMSE, the better the predictive performance of the model.

For each model, 50 tests are conducted in this paper and the average AUC value, average ACC value and average RMSE value are taken for comparison to get the comparative results of model knowledge tracking prediction performance as shown in Table 1.

From the table, it can be seen that the AUC, ACC and RMSE values of the G-EKT model are better than the other three comparative methods on both datasets, which suggests that the performance of the G-EKT model in predicting the future performance of the students is superior to the existing models. From the experimental comparison results, it can be seen that the BKT model has the lowest predictive performance on both data sets, which suggests that the BKT model has limitations in modeling students' potential level of knowledge about a knowledge point as a binary variable. The DKT model uses the latent vectors of the recurrent neural network to model the overall knowledge level of the students, and cannot model the students' mastery of the individual knowledge points, so the prediction performance of the DKT model on the two datasets is lower than that of the DKVMN and the G-EKT model. Both DKVMN and G-EKT models can model students' mastery of individual knowledge points, but the DKVMN model ignores students' forgetting behavior during the learning period and defaults to the fact that students' mastery of unreviewed knowledge points remains the same, which is a limitation, and thus the predictive performance of the G-EKT model is stronger than that of the DKVMN model.

Table 1. Prediction results of models

Index	Datasets	BKT	DKT	DKVMN	G-EKT
AUC	ASSISTments	0.672	0.725	0.734	0.762
	Slepemapy.cz	0.701	0.784	0.793	0.815
ACC	ASSISTments	0.636	0.681	0.687	0.736
	Slepemapy.cz	0.675	0.734	0.745	0.774
RMSE	ASSISTments	0.443	0.416	0.402	0.319
	Slepemapy.cz	0.421	0.408	0.399	0.348

4.1.2. Path recommendation performance effects

The state-of-the-art adaptive learning path recommendation frameworks selected for comparison in this study are Cognitive Structural Enhancement Model (CSEAL), Knowledge Point Concept Coverage Based Prediction Model (KCP-ER), and Bi-Layer Multi-Objective Recommendation Model (MulOER-SAN), with consistent parameter settings of the deep learning algorithms involved in the baseline model. In order to highlight the learning path recommendation effect, this paper uses effectiveness and adaptability metrics to evaluate the quality of the resulting adaptive learning paths. Effectiveness is used to assess the improvement of learners' learning performance in a session, and adaptability reflects whether each recommended exercise has an appropriate difficulty level. Table 2 shows the results of learning path recommendation for different models on different datasets.

It can be observed that the adaptive learning path recommended using the CSEAL model performs the worst on two different educational datasets. After analysis, it is found that the CSEAL model obtains the exercise candidate set directly through the successor knowledge point concepts of the current knowledge point concepts in the knowledge graph, instead of obtaining it by using the knowledge point concept coverage prediction, and it also does not take into account the difficulty characteristics of the exercises, which may be the reason for the poor recommendation effect. The MulOER-SAN model outperforms the KCP-ER model in all metrics on both datasets, and the RALP model outperforms all comparison models in both effectiveness and adaptability. The technology comparison experiments show that the dynamic learning environment representation that incorporates learner behavior and domain knowledge features makes the adaptive learning path recommendation based on reinforcement learning algorithms of higher quality and exhibits superior performance.

Table 2. Experimental results of different models

Model	ASSISTments		Slepemapy.cz	
	Validity	Adaptability	Validity	Adaptability
CSEAL	0.421	0.684	0.438	0.709
KCP-ER	0.457	0.701	0.452	0.743
MulOER-SAN	0.478	0.715	0.489	0.776
RALP	0.521	0.779	0.556	0.814

4.2. Adaptive Learning Pathway Applications

4.2.1. Experimental Procedures for Applied Objects

To verify whether the RALP model in this paper can push more matching learning paths for learners with different ability levels in Chinese teaching practice, language teaching practice is carried out. The following aspects are involved:

(1) Application object and content. In this study, 80 first-year high school students in the class of 2023 in a middle school affiliated to a teacher training university were used as the research subjects to carry out the application activities of Chinese teaching for a period of 18 weeks. The 80 first-year high school students were divided equally according to the random number table method, in which there were 40 students in each of the experimental class (YW1) and the control class (YW2), and the average age of the students in the two groups was 15.7 years old, with a male-to-female ratio of about 2:3. The textbooks in the syllabus of the first year of high school were selected as the teaching contents, and the Chinese teaching practice activities were carried out as a result of this.

(2) The teaching practice lasted for 18 weeks, and a pre-test was administered to both groups of students to test their initial proficiency levels. At the same time, the Ministep software was used to analyze the data of the previous students and the test questions of the question bank, compare the difficulty of the test questions and the ability level of the learners within the same range, eliminate abnormal questions, and determine the difficulty level of each test question. The test questions were randomly pushed for the students in the experimental group from the question bank, and the initial ability level of the students in the experimental group was obtained by Ministep software. The Chinese teaching activities were carried out separately for the two groups of students, with the control group adopting the traditional classroom teaching mode and the experimental group's teaching practice arranged as follows:

When the online teaching platform was used as the basic support to carry out the 18-week Chinese teaching practice, one test was carried out every three weeks, totaling six post-test experiments. At the end of each test, a learner diagnostic report was obtained and the learner's ability level was updated, the cognitive deficits of the students were counted, and remedial teaching was carried out after the class

based on the diagnostic report and the learner's ability level. Test questions are pushed to the learners based on the latest ability level calibration results for each test. Questionnaires were distributed and collected at the end of the 18th week of teaching practice to investigate the experimental group's satisfaction with the application of the RALP model, and finally, the relevant data and questionnaire results were collated and analyzed. The experimental group and the control group kept the same irrelevant variables except for the different teaching methods.

4.2.2. Performance of Students in Academic Achievement

With the initial knowledge level of the control group before carrying out the RALP model-based application of teaching Chinese. The analysis shows that the difference between the pretest scores of the experimental group ($M=72.16, SD=3.27$) and the control group ($M=73.09, SD=4.76$) is not significant ($t=0.571, p=0.193$). It indicates that the initial knowledge level of the two groups of students is comparable before carrying out the application of RALP model-based Chinese teaching and learning, which can avoid the result error due to the difference in the initial knowledge level of the students.

As the experimental group used the RALP model for the first time, there may be phenomena such as unfamiliarity with the operation process and unadaptation to the teaching method, which may affect the results of the teaching application, therefore, in this study, during the 18-week teaching practice, we carried out one teaching test every three weeks and recorded the test scores of the two groups of students. The article on the six post-test scores of paired samples t-test, comparing the experimental group and the control group scores of each test, and its specific comparison results are shown in Table 3.

The difference in performance between the experimental group ($M=74.28, SD=4.15$) and the control group ($M=74.39, SD=3.27$) was not significant ($t=0.273, p=0.135$) in the posttest 1 test. There may be two reasons for this, one is that the experimental group was not familiar with the ARLP model and did not adapt to the new teaching method. The second is that the experimental group did not use the ARLP model for personalized learning in the pre-test. After the first teaching in the posttest, students gradually adapted to the teaching mode supported by the ARLP model and completed the Chinese course with its assistance. With $p=0.032<0.05$ in the second test, $p=0.007<0.01$ in the third test, and $p<0.001$ in the results of the fourth to sixth tests, the posttest scores of the experimental group and the control group differed significantly, and the differences gradually increased. The control group lacks the support of cognitive diagnosis and related personalized teaching aids, so the ambiguous goals and time-consuming access to information in Chinese course learning are the main reasons for the lower problem solving rate and learning efficiency. The diagnostic results given by the ARLP model can help the experimental group to accurately grasp their own knowledge structure, and the adaptive learning path recommended with the support of the diagnostic results can help the experimental group to make up for the knowledge deficiencies. The ARLP model can help the experimental group accurately grasp their own knowledge structure, and the adaptive learning path recommendation supported by the diagnostic results can assist the experimental group to make up for the knowledge deficiencies, so as to carry out more targeted learning of Chinese courses, and ultimately improve the problem solving rate and learning efficiency of the experimental group. As a result, the teaching application of ARLP model promotes the adaptive learning of the experimental group, and the students' performance is significantly improved.

Table 3. Analysis of the post-test scores of the experimental and control group

Test	Class	Means	SD	SE	t value	p value
Test1	YW1	74.28	4.15	0.916	0.273	0.135
	YW2	74.39	3.27	0.943		
Test2	YW1	79.87	3.89	0.781	1.765*	0.032
	YW2	75.01	4.13	0.854		
Test3	YW1	82.19	3.72	0.821	2.914**	0.007
	YW2	76.53	3.95	0.804		
Test4	YW1	83.07	0.81	0.804	3.562***	<0.001
	YW2	76.98	3.49	0.783		
Test5	YW1	85.72	2.65	0.751	6.298***	<0.001
	YW2	77.16	2.57	0.742		
Test6	YW1	86.85	2.16	0.736	5.741***	<0.001
	YW2	78.92	2.33	0.729		

Note: * $p<0.05$, ** $p<0.01$, *** $p<0.001$

4.2.3. Learning path quality scores

At the end of the 18-week teaching practice, a comprehensive rating was conducted for the two groups of students' learning of the knowledge points of the Chinese curriculum, mainly through the online learning platform to collect students' ratings of the effectiveness of the recommended adaptive learning paths and students' ratings of the quality of the recommended learning paths. The ratings were characterized by a 7-point Likert scale, with higher branches indicating that the adaptive learning paths were of higher quality and better able to meet students' learning needs of Chinese course knowledge. At the end of the practice of the entire teaching course, the quality ratings of all the learning paths were averaged out on a student-by-student basis, and their results were obtained as shown in Fig. 3, with the colored area in the figure as the margin of error.

From the figure, it can be seen that the experimental group students' ratings of adaptive learning paths are generally higher than the control group students' ratings of independent learning paths, indicating that the adaptive learning paths constructed in this study are of better quality relative to the independent learning paths, due to the fact that the latter recommends suitable learning paths for students and reduces the learning load. An independent samples t-test on the scores of the two groups of students showed that the mean values of the experimental and control group students' scores on the quality of the learning paths were 6.47 and 5.89, respectively, and the experimental group's scores were 9.85% higher than those of the control group students, with a significant difference in the composite scores ($p < 0.01$).

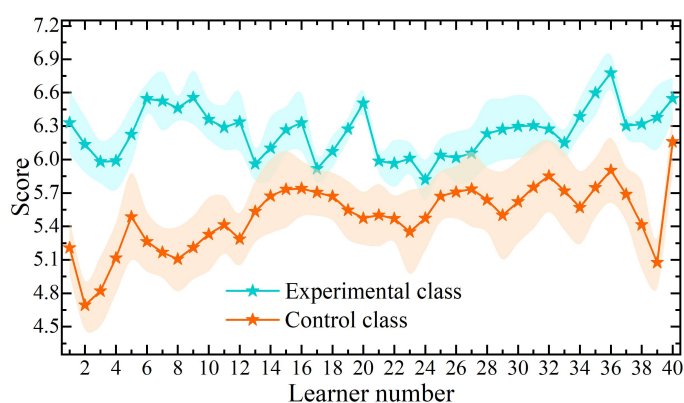


Figure 3. Learning path quality scoring

At the end of the Chinese course, the log of learning behaviors recorded by the online learning platform can be used to derive the average learning time for students to learn a single knowledge unit, and the average learning time after normalization is shown in Figure 4.

As can be seen from the figure, most of the students in the experimental group used a shorter learning time than the students in the control group, which is due to the fact that when learning according to the recommended adaptive learning paths, the students spend less time on searching and selecting learning resources, and learn more efficiently. However, there are some exceptions, which may be due to individual differences in students' learning ability and learning efficiency. In addition, the difference in the average study time of students in the control group was large, while the difference in the average study time of students in the experimental group was small. This is due to the different searching and learning abilities of students, which caused the time spent by students in the control group in searching for learning resources to be different, whereas students in the experimental group used the adaptive learning paths recommended by the platform, which narrowed down the difference in the average learning time to a certain extent. The results of the independent sample t-test on the average learning time of the two groups of students show that the difference in the average learning time of the two groups of students is significant ($p < 0.05$), and the adaptive learning path effectively reduces the cognitive load of the experimental group of students, so that their learning efficiency is improved.

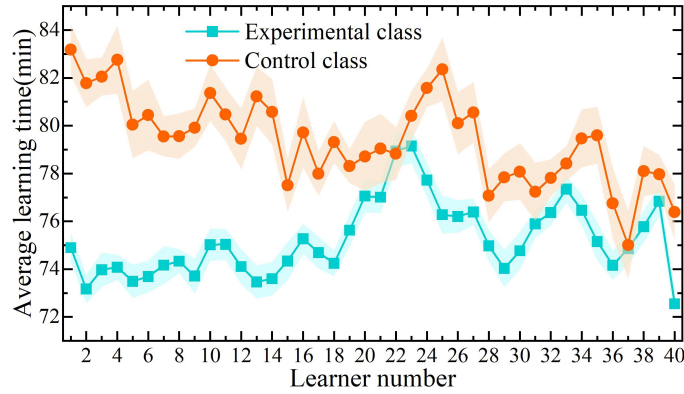


Figure 4. Comparison of average study time

4.2.4. Learning and mastery of knowledge points

Based on the analysis of student performance and learning path quality scores, for students in the experimental class, data can also be analyzed to evaluate how many knowledge points a student has learned and how many more have been mastered, and visualize the learning status and implement personalized intervention guidance to visualize the learning and mastery of knowledge points as shown in Figure 5. In the figure, the right circle represents the number of knowledge points learned, the left square represents the number of knowledge points mastered, and the difference in distance between the two points represents the number of knowledge points not mastered by the student.

From the figure, it can be seen that the 33rd student corresponds to the longest line between the two points, which indicates that the student's learning situation is worse. In this paper, when building the ARLP model, it is designed for the student's learning situation, i.e., when the distance difference between the two points is lower than the set threshold (20), the system will automatically send an alert according to the student's learning situation. And at the same time, it will push the knowledge points appropriate to its difficulty for learning according to the cognitive ability of the learner, informing the students of the previous knowledge and related knowledge that they should learn. After completing the learning, the subsequent knowledge points that can be learned only, this personalized intervention guidance helps to promote teaching and learning. In the figure, the number of corresponding knowledge points for 3 students, such as 5th, 9th and 11th, is only shown as circles, indicating that the knowledge of the Chinese they have learned has been fully mastered.

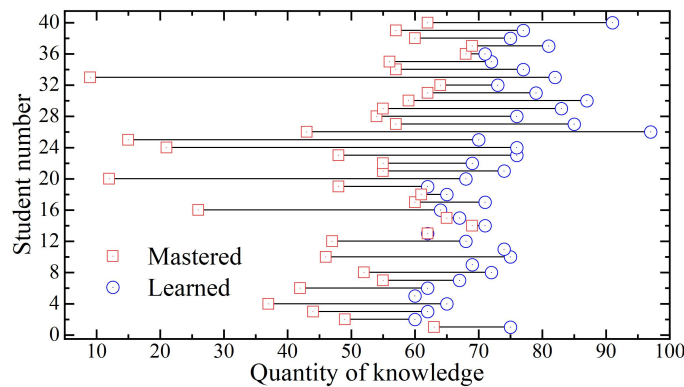


Figure 5. Visualize the learning and mastery of knowledge points

5. Conclusion

In order to improve the effectiveness and quality of Chinese teaching, the article proposes an adaptive learning path optimization recommendation framework that combines reinforcement learning and knowledge tracking, and carries out application practice for the effectiveness of the framework. It is found that the effectiveness and adaptability of the RALP model in recommending adaptive learning paths reach 0.556 and 0.814 respectively, and in the teaching practice supported by the RALP model, the experimental class students' language learning performance is significantly improved, and the quality of the learning paths scores are higher, and the mastery of Chinese course knowledge can be visually manifested. Therefore, the adaptive learning path optimization recommendation can

significantly enhance the effect of language teaching and provide a decision-making path to promote the innovative reform of language course teaching.

About the Author

Ruifeng Guo (1984.9-), female, born in Shangdu, Inner Mongolia, holds a master's degree from Inner Mongolia Normal University. Currently pursuing a PhD in Social Sciences at Universiti Selangor, Malaysia, lecturer at Jining Normal University, main research areas: primary school Chinese teaching and children's literature.

References

1. Duan, J., Xiong, W., Li, J., Xu, M., & Jia, Y. (2025, March). Thoughts and Research on the Blended Teaching Model in Military Academies under the Condition of Information Technology. In 2024 4th International Conference on Education, Language and Art (ICELA 2024) (pp. 41-47). Atlantis Press.
2. Ergashi, M. G., & Muhidin, N. G. (2025). The Role of Interactive Pedagogy for Educational Reform. *ASEAN Journal of Educational Research and Technology*, 4(2), 139-146.
3. Li, N., & Liang, Y. (2025). Teachers' AI readiness in Chinese as a Foreign Language education: Scale development and validation. *System*, 129, 103597.
4. Wang, Y. (2025). The Development and Challenges of Chinese Language Education: A Global Perspective. *Journal of Education, Humanity and Arts Research*, 9-9.
5. Siddiqui, M. G., Khan, F., Ibrahim, M., & Fawad, R. M. (2025). PERCEPTION OF TEACHERS ABOUT THE IMPORTANCE OF STUDYING CHINESE LANGUAGE: PROSPECTS, POSSIBILITIES AND CHALLENGES FOR FUTURE DYNAMICS. *International Journal of Social Sciences Bulletin*, 3(9), 688-706.
6. Yu, W., Huili, S., & ZhangYi, P. (2025). TECHNOLOGY INTEGRATION, DIGITAL PEDAGOGY, AND MOTIVATION: A REVIEW OF CORE ISSUES IN INTELLIGENTLY EMPOWERED INTERNATIONAL CHINESE LANGUAGE EDUCATION. *TECHNOLOGY*, 8(05).
7. Xie, W., Li, J., Mu, Y., Zhang, H., Zhao, S., Zheng, X., & Song, K. (2025). The Power of Personalized Datasets: Advancing Chinese Composition Writing for Elementary School through Targeted Model Fine-Tuning. *International Journal of Asian Language Processing*, 35(02), 2450017.
8. Shanthi, D., Ashok, G., Biswal, C., Udharika, S., Varshini, S., & Sindhu, G. (2025). Ai-driven adaptive it training: A personalized learning framework for enhanced knowledge retention and engagement. *Metallurgical and Materials Engineering*, 136-145.
9. Yvonne Khomo, F. L., Millham, R., & Mustafa, A. (2025). Personalized Adaptive Learning Management System Based on Bioinspired Algorithm: A Systematic Review. *Engineering Innovations*, 14, 139-146.
10. Rodriguez, M., Rizky, A., Tanjung, Y. I., & Sumliyah, S. (2025). Optimizing e-learning platforms through adaptive learning for diverse demographics. *Jurnal MENTARI: Manajemen, Pendidikan dan Teknologi Informasi*, 4(1), 30-39.
11. Umaeswari, P., Patil, S., Divakarachari, P. B., & Falkowski-Gilski, P. (2025). Smart adaptive learning and optimized feature clustering for enhanced image retrieval. *Scientific Reports*, 15(1), 24324.
12. Pulungan, N. K., & Nora, Y. E. (2025). Students' Experiences and Meaning-Making in AI-Based Adaptive Learning in Higher Education. *Journal of Educational Innovation and Research*, 1(10), 424-434.
13. Zhang, Y., Zhang, L., Zhang, H., & Wu, X. (2025). Development of a Computerized Adaptive Assessment and Learning System for Mathematical Ability Based on Cognitive Diagnosis. *Journal of Intelligence*, 13(9), 114.
14. Troussas, C., Krouska, A., Mylonas, P., & Sgouropoulou, C. (2025). Personalized Instructional Strategy Adaptation Using TOPSIS: A Multi-Criteria Decision-Making Approach for Adaptive Learning Systems. *Information*, 16(5), 409.

15. Venice, J. A., Arivazhagan, D., Suman, N., Shanthi, H. J., & Swadhi, R. (2025). Recommendation Systems and Content Personalization: Algorithms, Applications, and Adaptive Learning. In *AI for Large Scale Communication Networks* (pp. 323-348). IGI Global.
16. Student, R. L. A. T. E. (2025). Optimizing Educational Content With Q-Learning: A Reinforcement Learning Approach To Enhance Student. *Proceedings of the E-Learning and Smart Engineering Systems (ELSES 2024)*, 36, 193.
17. Liu, H. Personalized learning support system for special education: a real-time feedback mechanism based on deep reinforcement learning. *Frontiers in Psychology*, 16, 1658698.
18. Kappagantula, S., & Mannayee, G. (2025). Reinforcement Learning in Educational Robotics: A Framework for Human-Robot Interaction in 21st Century Classrooms. *Foro de Educacion*, 23(1).
19. Ruan, S., & Lu, K. (2025). Adaptive deep reinforcement learning for personalized learning pathways: A multimodal data-driven approach with real-time feedback optimization. *Computers and Education: Artificial Intelligence*, 100463.