International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-15

https://doi.org/10.70917/ijcisim-2026-0068
Article

An Intelligent Teaching Aid System for English Linguistics Courses Based on Computer Generated Dialogue System

Li'ao Luo *

London Waterloo Campus, King's College London, London, SE1 9NH, UK; luoliaoapply@163.com

Abstract: Information technology is of great significance in the curriculum setting and implementation of college English. Traditional college English pays more attention to the extension of the application of multimedia facilities, which is characterized by the problems of untimely updating of content, heavy workload of teachers' preparation and insufficient reinforcement of grammatical context. In this paper, a reinforcement learning-based dialog generation model is proposed based on reinforcement learning theory, which fully considers the one-to-many relationship in dialog data, i.e., there may be multiple reasonable replies for each user input. Then on the basis of the model, an intelligent teaching dialogue system is further developed to realize assisted teaching for English language courses. Through the comparison experiments and case studies of different baseline models in the dataset, it is found that the dialogue generation model proposed in this paper performs better than the other baseline models in various indexes, while the intelligent system can still support about 55% of the users to complete the request when the users reach 700 in the functionality test. Finally, the statistical analysis of the performance in the study case shows that the intelligent dialog generation system proposed in this paper can affect the actual effectiveness of teaching and learning.

Keywords: reinforcement learning; response generation; intelligent teaching; dialog system

1. Introduction

Since the 1960s, computer-assisted language learning systems (CALL) have made significant progress as an interdisciplinary field at the intersection of information technology and linguistics. This field has produced a wealth of research findings and exhibited a trend toward disciplinary specialization, evolving from programmed linear instruction to multimedia interaction, and ultimately to today's adaptive learning systems. CALL systems have become increasingly intelligent and are widely applied in language education [1-4].

English learning, as one of the most closely watched educational fields globally, has significantly benefited from the introduction and application of intelligent teaching assistance models. With the support of intelligent technology, teaching efficiency can be improved, personalized instruction can be provided, customized learning plans can be offered, and students' motivation and enthusiasm for learning can be enhanced, thereby achieving a more effective, efficient, and student-centered teaching model [5-8]. Intelligent teaching assistance aims to combine advanced technology with teaching practice to create more innovative and practical teaching methods, bringing more possibilities and opportunities to the education field [9]. The "China Education Modernization 2035" plan explicitly states that the implementation of artificial intelligence to promote teacher team development should be carried out, and modern technology should be utilized to accelerate reforms in talent cultivation models. As such, the construction and practical exploration of intelligent teaching assistance systems, the promotion of innovative applications of intelligent technology in the education sector, and the advancement of modernization in English language education all hold significant practical significance in the new era [10-11].

Since Apple introduced the Siri virtual personal assistant in 2009, conversational robots have sparked a new wave in the industry. With the continuous development of natural language processing,

conversational systems have entered the GPT era, dominated by deep learning, and are gradually entering the education sector, serving as a valuable tool for achieving intelligent language teaching assistance [12-13]. Reference [14] established an intelligent teaching assistance system supported by machine learning, combining positioning algorithms and intelligent database analysis and processing functions to track students' learning progress. Reference [15] developed an English language learning intelligent tutoring system based on computers, utilizing collaborative filtering algorithms and convolutional neural network algorithms, which improved students' learning efficiency and significantly enhanced their exam scores. Literature [16] developed an online English speaking teaching assistance system with tag matching functionality, which assists students in self-directed learning, provides personalized resource recommendations, and supports teachers in monitoring student learning progress. Literature [17] designed an intelligent English teaching assistance system with the help of artificial intelligence and intelligent speech technology. In addition to basic teaching assistance, it can correct students' spoken pronunciation and promote self-directed learning. Literature [18] utilized a powerful language model to create an AI-enhanced intelligent educational assistance architecture with speech functionality, covering course content, student learning, teacher assistance, and policy dissemination to alleviate learning barriers for students. Literature [19] noted that while current AI dialogue systems enhance students' learning abilities in English instruction, these systems overlook functions related to culture, humor, empathy, debate, and problem-solving. In 2024, the AI dialogue system introduced in [20] was applied to English learning, integrating three factors—multiculturalism, humor, and empathy—to enhance students' motivation, engagement, and interactivity. [21] confirmed that dialogue-based computer-assisted language learning (CALL) facilitates second language learning through form- and goal-oriented approaches, guided interaction, corrective feedback, and gamification.

The study proposes a dialog generation model based on reinforcement learning, which consists of two modules, the implied word prediction network and the reply generation network. Then on the basis of the model, an intelligent teaching dialogue system is designed and developed for realizing the teaching aid function for English language courses. The PersonaChat dataset is selected to be compared with four baseline models, MemNet, PostKS, GLKPG and RL-GLKPG, respectively, to prove the reasonableness and coherence of the model proposed in this paper in multi-round conversations, and the DQN algorithm and this paper's algorithm are also compared for experiments. Finally, 81 students are selected as the study cases, and the normal test and t-test are conducted on the students' performance to test the effect of the intelligent teaching dialog system on the students' learning.

2. Intelligent Teaching Aid System for English Linguistics Courses

2.1. Enhanced Learning

2.1.1. Markov Decision-Making Process

Reinforcement learning is a learning method in which an intelligence continuously interacts with the environment, evaluates its behavior by obtaining rewards, and then guides its behavior, in which the intelligence continuously performs “trial and error” with the aim of obtaining the greatest rewards [22].

Reinforcement learning uses Markov Decision Process (MDP) [23] as the underlying framework. MDP is Markovian in the sense that the future state depends only on the current state and is independent of the historical state, i.e., $P(S_t+1 \mid S_t) = P(S_{t+1} \mid S_0, S_1, ..., S_t)$ where S_t denotes the state at the t th moment. The MDP is usually represented by a quintuple (S, A, P, R, γ), where each element represents the meaning as follows:

Element S denotes the set of all possible states observed by the intelligent body, $s_t \in S$, where s_t denotes the state of the environment observed by the intelligent body at the t th moment.

Element A denotes the set of all actions that can be selected by the intelligent body, $a_t \in A$, where a_t denotes the action selected by the intelligent body at the t th moment.

Element P denotes the state transfer probability matrix, $P^a_{s,s'} = \mathbb{P}(s_{t+1} = s' \mid s_t = s, a_t = a)$, which indicates that at state s the the probability that the intelligent body picks action a and the state shifts to s'.

Element R denotes the reward function, $R^a_s = \mathbb{E}(R_{t+1} \mid s_t = s, a_t = a)$, representing the reward R_{t+1} obtained by the intelligent body for picking action a in state s.

Element γ is a discount factor, $\gamma \in [0,1]$, which is used to regulate the weight between the current

reward and the long-term return; the larger γ is, the more attention is paid to the long-term return.

2.1.2. Optimal Strategy

If for all states $s \in S$, if there are two strategies π and π' with $\pi \le \pi'$, then $V^{\pi}(s) \le V^{\pi'}(s)$, i.e. if one strategy π' is better than another strategy π, then the payoffs on all states satisfy that the payoff under strategy π' is greater than or equal to the payoff under strategy π. Accordingly, define the optimal strategy π^*, there may be more than one optimal strategy, but under any optimal strategy, the payoffs at each state are unique, i.e.:

$$V^{\pi^*}(s) = \max_{\pi} V^{\pi}(s), \forall s \in S \tag{1}$$

Also, under the optimal policy, the value function of the action in each state is maximized, i.e:

$$Q^{\pi^*}(s,a) = \max_{\pi} Q^{\pi}(s,a), \forall s \in S \tag{2}$$

The purpose of reinforcement learning is to find the optimal policy π^* through continuous exploration, so as to maximize the long-term return in each state. Traditional reinforcement learning algorithms can be categorized into value estimation based methods and policy gradient based methods.

In the value estimation based approach, the intelligent body selects actions based on $Q(s,a)$ in decision making. For any state $s(s \in S)$, the intelligent body can select all actions $a(a \in A)$, and when the intelligent body selects an action, the decision is made based on which action in state s can obtain the largest long-term return $Q(s,a)$. Therefore, in the value estimation-based approach, the long-term payoff $Q(s,a), \forall s \in S, \forall s \in A$ that can be gained by each action in each state needs to be maintained, and the classical approaches are Q Learning and Deep Q Network (DQN).

In the policy-based approach, the intelligent body directly learns the policy $\pi(a \mid s, \theta)$, with θ as the parameterized representation of the policy, by obtaining the probabilities of all the actions under the state s, and then samples the probabilities of each action to determine the currently selected action. Among the strategy-based methods, a typical method is the Reinforce method, in which the parameter θ of the strategy is parametrically updated based on Eqs. (3) and (4):

$$G \leftarrow \sum_{k=t+1}^{r} \gamma^{k-t-1} R_k \tag{3}$$

$$\theta \leftarrow \theta + \alpha \gamma^t G \nabla \ln \pi(a_t \mid s_t, \theta) \tag{4}$$

That is, the parameter gradient is weighted according to the long-term return obtained by the policy $\pi(a \mid s, \theta)$ after executing the action in state s_t, and if more long-term return G is obtained, the better the action is, and the policy learns to focus on the action more in state s_t.

2.1.3. Advantage Actor-Critic (A2C) Algorithm

In the previous section, two methods for obtaining the optimal policy were introduced: the method based on value estimation and the method based on policy gradient, and the A2C algorithm is the algorithm generated by combining the two methods.The A2C algorithm can be seen from the word composition, which is divided into two parts, one part is Actor, and the other part is Critic.Actor is the meaning of performer, which here refers to the strategy of the intelligent body $\pi(a \mid s, \theta)$, which makes action decisions based on observed states [24].Critic means critic, which means to score the performer's performance, and here it refers to the evaluation of the actions selected by the intelligent body's strategy, which guides Actor's learning.

Actor models for the dialog strategy $\pi(a \mid s, \theta)$ and updates the parameter θ using the strategy gradient:

$$\theta \leftarrow \theta + \alpha A^{\pi}(s_t, a_t) \nabla \ln \pi(a_t \mid s_t, \theta) \tag{5}$$

where $A^{\pi}(s_t, a_t)$ is the Advantage Function, which denotes the advantage of picking the action a_t in

the state s_t compared to picking the other actions, and the Advantage Function is evaluated as a Critic, which is computed by Eq.(6):

$$A^{\pi}(s_t, a_t) = Q^{\pi}(s_t, a_t, \omega_Q) - V^{\pi}(s_t, \omega_V) \tag{6}$$

where ω is the parameter. By calculating the difference between the action value of the selected action a_t under the state s_t and the difference between the value of the action under the state s_t, the advantage of the action a_t under the state s_t is obtained, and the state value under the state s_t is the expectation of the action value when all actions are selected, so when the difference is the result of the dominant function, it means that the action a_t is selected Higher than expected, the higher the value, the better the selection action a_t.

Knowing that there is a conversion relationship between the action value function and the state value function, in order to reduce the amount of parameter calculation, the action value function is represented by the state value function, so the dominance function is further deduced as:

$$A^{\pi}(s_t, a_t) = R_{t+1} + \gamma V^{\pi}(s_{t+1}, \omega_V) - V^{\pi}(s_t, \omega_V) \tag{7}$$

Therefore, only the state value function needs to be modeled.

2.2. Reinforcement Learning Based Model for Dialogue Generation

2.2.1. Problem Definition and Model Overview

Given a set of training samples $\{(x, \{y\})\}$, where x is the input text and $\{y\}$ is its corresponding multiple target responses. We assume that each input text x has a hidden variable z. Before introducing the proposed model, this paper will first explore how to construct a suitable hidden variable space for the undialog generation task:

Any $z \in Z$ should be interpretable and able to show the relevance of the input text to each question. This makes it easier to assess whether the model assigns a suitable hidden variable to a sample $\{(x, \{y\})\}$.

Each $z \in Z$ should have the ability to capture features of differences between sentences. In this way, given different x a diverse z of hidden variables can be sampled, and the model can then generate diverse responses based on different z.

According to the above two points, we take the implied words in the responses as hidden variables, and the size of the hidden variable space is the size of the word list used in training. In this way, each hidden variable z corresponds to a word in the word list one by one. Based on this, we can directly assess the relevance of the display text represented by z, x and $\{y\}$. Also using the word list as a hidden variable space, for each input text x enough diverse hidden variables z can be selected to generate diverse responses.

2.2.2. Implicit Word Prediction Networks

The main purpose of an implied word prediction network is to approximate the probability distribution $p(z \mid x)$. We first encode the input text x using a bi-directional LSTM network to obtain a vector representation h_x of the input text. This is followed by calculating the probability distribution of the implied words using the following formula:

$$p(z \mid x) = softmax(W_2 \cdot \tanh(W_1 h_x + b_1) + b_2) \tag{8}$$

where W_1, b_1, W_2 and b_2 are the parameters of the LSTM network.

2.2.3. Response Generation Network

Suppose that the overall loss $\mathcal{L}(\{y\})$ can be approximated using a parameter-free differentiable function $f(\cdot)$ through independent losses $\ell(y \mid x, z)$ for each $y \in \{y\}$:

$$\mathcal{L}(\{y\} \mid x, z) = f_{y \in \{y\}}(\ell(y \mid x, z)) \tag{9}$$

The next section will first introduce the network used to estimate the loss $\ell(y \mid x, z)$, and then discuss in detail about the choice of $f(\cdot)$.

Given a hidden variable z and an input text x, another bi-directional LSTM network is employed as the input text encoder to encode x to obtain a vector representation of the input text h_x^g, while a fully-connected layer is used to take the word vectors of the implied words as inputs to obtain a vector representation of the hidden variable z, h_z^g. These two vector representations are subsequently used to decode the generated responses y. For decoding, the attention distribution between the input text x and h_z^g as well as the context c_z of h_z^g is first computed:

$$a_j = \frac{\exp(h_z^{gT} h_x^g(j))}{\sum_t exp(h_z^{gT} h_x^g(t))} \tag{10}$$

$$c_z = \sum_t a_t h_x^g(t) \tag{11}$$

where $\{h_x^g(t)\}$ is the sequence of input text vector representations obtained by the encoder for each word in the input text x, and c_z is the weighted sum of these vector representations.

For the generation of the responses, we use a unidirectional LSTM network as the decoder and define its implicit vectors as $\{h_t(t)\}$. In the decoding phase, we compute the context vector $c_y(t)$ of $h_y(t)$ in the same way by replacing h_z^g with $h_y(t)$ in equations (10) and (11). We then splice the vectors $h_y(t), c_y(t)$ and c_z to pass into the fully connected layer and subsequently the softmax output layer to obtain the probability distribution $p(y_t \mid y_{<t}, x, z)$ of the current output word $y(t)$. Here we set $\ell(y \mid x, z)$ as the negative log-likelihood function of $p(y \mid x, z)$. Compared to the traditional SEQ2SEQ model decoding, we incorporate more information related to the implied word to further help the model predict the current output word $y(t)$.

Next, explore how to design an appropriate $f(\cdot)$ function. The general practice is to set $f(\cdot)$ as a mean value function:

$$\mathcal{L}_{avg}(\{y\} \mid x, z) = \frac{1}{|\{y\}|} \sum_{y \in \{y\}} \ell(y \mid x, z) \tag{12}$$

where $|\{y\}|$ is the base of the set $\{y\}$. But for our setting, $\mathcal{L}_{arg}$ may not be appropriate. The loss term $\ell(y \mid x, z)$ will have a small value for those responses that have a corresponding hidden variable z, while its value will be large for responses that do not have a corresponding hidden variable z. This leads to the overall loss $\mathcal{L}_{avg}(| y \mid x, z)$ will still have a large value. Therefore, we propose to use the minimal value function most our $f(\cdot)$ function:

$$\mathcal{L}_{\min}(\{y\} \mid x, z) = \min_{y \in \{y\}} \ell(y \mid x, z) \tag{13}$$

In this way, if a given sampled hidden variable z with input text x as model input can generate better results, the overall loss value of the model also becomes small.

2.2.4. Reinforcement Learning Based Joint Training Methods

Usually, in order to update the model parameters, we can calculate the gradient of $J(\Theta)$ by using Eq. (14):

$$\begin{aligned}\nabla_{\Theta} J_{\Theta} &= \nabla_{\Theta} \sum_{z \in Z} p(z \mid x) \mathcal{L}(\mid y \mid \mid x, z) \\ &= \sum_{z \in Z} [\mathcal{L}(\{y\} \mid x, z) \nabla_{W} p(z \mid x) + p(z \mid x) \nabla_{G} \mathcal{L}(\mid y \mid \mid x, z)] \\ &= E_{p(z \mid x; W)} [\mathcal{L}(\{y\} \mid x, z) \nabla_{W} \log p(z \mid x) + \nabla_{G} \mathcal{L}(\{y\} \mid x, z)]\end{aligned} \tag{14}$$

The large hidden variable space Z makes the above derivation difficult. Therefore, we propose to use reinforcement learning for optimization. First, we use an existing keyword extraction tool to extract the keywords corresponding to the responses for each input text as implied words, and use these implied words as the target outputs to pre-train the implied word prediction network, and then we use the implied words with the highest probability of being predicted by the implied word network as the inputs to pre-train the response generation network. In the training phase, we use a reinforcement learning algorithm to train the implied word prediction network and a standard backward propagation algorithm to update the reply generation network.

To address both problems, we design a second-order sampling method that can restrict the sampling of hidden variables to a smaller space. For the second problem, we propose to replace the $\mathcal{L}(\{y\} \mid x, z)$ term in the strategy gradient with the reward $\mathcal{R}(\{y\} \mid x, z)$ with boundary constraints.

2.2.5. Diversity Response Generation

Next, we explore how to generate K semantically diverse responses for a test input x. Given a trained model, we select the top 1000 hidden variables z as a candidate set based on the probability $p(z \mid x)$, and then perform the same clustering operation on the candidate set as in the training phase, where we select the cluster centers as outputs from K clusters with large drums. The final K hidden variables z selected along with the test input x are used as inputs to the generative network, which subsequently outputs K different responses.

2.3. Multi-Round Dialogue System Implementation for Intelligent Teaching and Learning

2.3.1. System Design

The Multi-Round Dialogue System for Intelligent Teaching and Learning simulates real teaching interactions and is a communication platform between virtual teachers and students. On this platform, both parties engage in useful two-way communication in the form of questioning and responding, thus helping students to grasp and digest what they have learned more effectively. In this section, the overall design of the Intelligent Teaching Dialogue System will be analyzed in detail, revealing its comprehensive working principles and outlining the composition of each level by detailing the overall architecture and data flow of the system.

The overall architecture of the Intelligent Teaching Dialogue System is shown in Figure 1, which contains four layers: the front-end display layer, the back-end basic service layer, the Intelligent Teaching Dialogue processing layer and the data storage layer. Users interact with the system through the front-end display layer, either through the Web interface or through the mobile APP. This layer provides the user interface and the interaction entrance of the system.

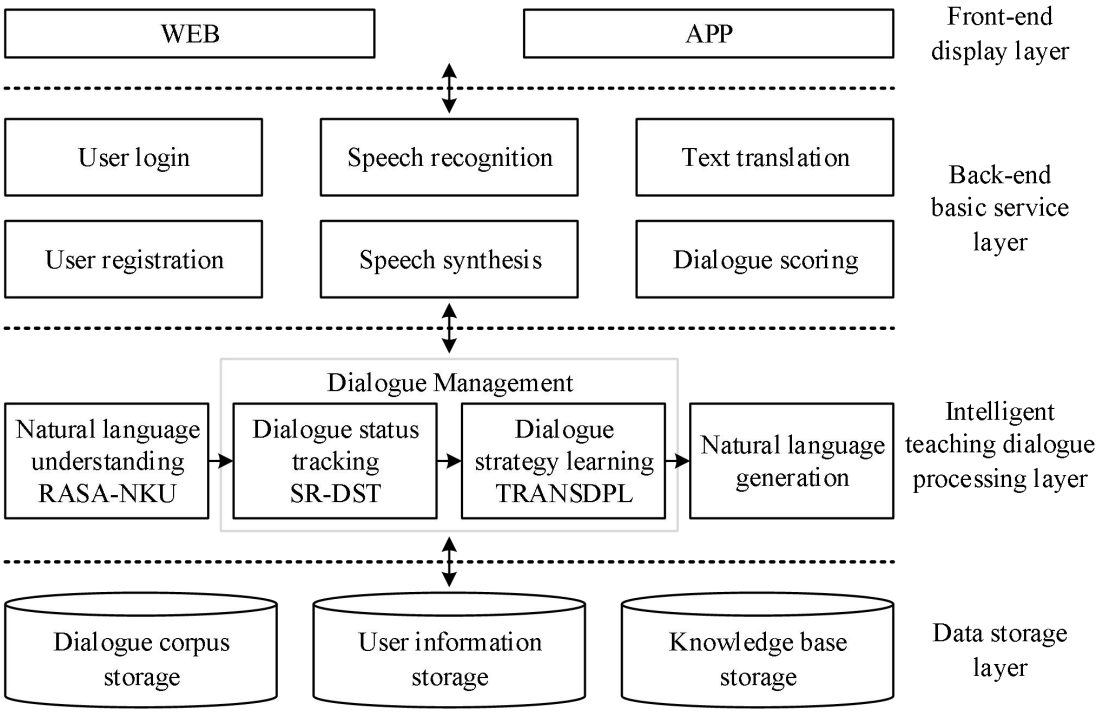


Figure 1. The overall architecture diagram of the intelligent teaching dialogue system.

2.3.2. Back-end service design and implementation

This layer mainly focuses on the design of some user experience related functions, and the main user experience functions include user login and registration, voice processing and text translation. Golang's gin framework is used for the development of the back-end basic service layer, and automated deployment is realized by combining Gogs, Docker, and Drone to support an efficient integration development process.

3. The Effect of Intelligent Teaching Aid in English Linguistics Courses3.1 Experimental Setup

3.1. Experimental Setup

3.1.1. Introduction to the Data set

The experiments in this chapter are conducted on the PersonaChat dataset, which is a multi-round conversation dataset based on the background knowledge of both parties in a conversation, constructed through crowdsourcing. In constructing a conversation, the crowdsourced workers are paired two by two, and then given the background knowledge of each of the two people, which is through is sentences describing personal experiences and personality information, and then the two crowdsourced workers start a conversation and utilize the provided background knowledge information during the conversation.The PersonaChat dataset contains a total of 9,897 multilateration conversations, and the average number of conversational rounds of this dataset is 7-8 rounds. The number of knowledge sentences contained in the background knowledge set is about 4-5, so these statistics were referred to when conducting the dialog simulation in the experiment, and the maximum number of rounds for the dialog simulation was set to 8.

3.1.2. Baseline Model

In order to validate the effectiveness of the methods proposed in this chapter, MemNet, PostKS, GLKPG, and RL-GLKPG dialog generation models are selected as comparison models for this experiment.

3.1.3. Experimental Environment and Parameter Settings

The sentence encoder for all models in the experiments is a single-layer bi-directional GRU network

with the hidden state dimension set to 200, and the decoder is a uni-directional GRU with the hidden state dimension set to 400. The size of the vocabulary list is set to 15,000. When encoding the words in the dataset, all the words that do not appear in the vocabulary list are uniformly replaced with the characters "UNK" will be used instead. The Adagrad algorithm is used to update the parameters of the model during the training process, and the optimizer is initialized with a learning rate of 0.0005, which is gradually reduced by a decay factor as the training progresses.

3.1.4. Evaluation Indicators

(1) Automated Evaluation Metrics

This experiment mainly uses two automatic evaluation metrics, DISTINCT and KnowledgeF1, to automate the scoring of the response statements generated by each model on the test set.DISTINCT is used to measure the degree of word diversity of the response statements generated by the model.KnowledgeF1 is used to measure the amount of information in the responses generated by the model, which is computed as shown in Equation (15), Eq. (16) and Eq. (17) show. W_γ and W_K denote the set consisting of non-discontinued words in the model-generated replies and background knowledge sentences, respectively.The higher value of KnowledgeF1 indicates that more information in the background knowledge sentence appears in the reply statement, reflecting the higher utilization of knowledge by the model.

$$Precision = \frac{|W_Y \cap W_K|}{|W_Y|} \tag{15}$$

$$Recall = \frac{|W_Y \cap W_K|}{|W_K|} \tag{16}$$

$$KnowledgeFl = 2 \times \frac{Recall \times Precision}{Recall + Precision} \tag{17}$$

(2) Manual evaluation index

This chapter focuses on the knowledge selection strategy of the dialog generation model in multi-round dialog, while the automatic evaluation index only evaluates from the point of view of a single reply statement, i.e., a single round of dialog, which cannot accurately reflect the model's ability of multi-round dialog, for this reason, this experiment generates multi-round dialog data by making the model carry out a dialog simulation, and then carries out a manual evaluation. First, 100 samples are randomly selected from the test set, and the background knowledge sentences provided in the samples and the starting sentences in the dialog data are taken out as the test data.

3.2. Experimental Results and Analysis

3.2.1. Automated Evaluation of Results

The DISTINCT-1 and DISTINCT-2 values of each baseline model and the proposed reinforcement learning-based diverse response generation method (Ours) on the PersonaChat dataset are shown in Table 1. From the table, it can be seen that GLKPG achieves the highest values for the DISTINCT- and DISTINCT-2 metrics, which are 0.052 and 0.142, respectively. Whereas, the DISTINCT-1 and DISTINCT-2 values of the proposed model in this paper are slightly lower than those of the GLKPG model but still significantly higher than those of the other baseline models. The MemNet and PostKS models did not adopt the knowledge replication mechanism, which resulted in their inability to effectively utilize the knowledge information in generating responses, and still suffered from the generic response problem of the traditional Seq2Seq model, so the diversity of responses was lower. In addition, by comparing GLKPG and RL-GLKPG, it can be seen that after optimizing the decoder of the GLKPG model through reinforcement learning, the performance of the model decreases significantly, and the DISTINCT metrics are close to MemNet. It indicates that the application of reinforcement learning to the optimization of reply generation may adversely affect the decoder.

Table 1. DISTINCT automatic evaluation results.

Model	DISTINCT-1	DISTINCT-2
Mem Net	0.019	0.033
PostKS	0.027	0.088
GLKPG	0.052	0.142
RL-GLKPG	0.021	0.037
Ours	0.041	0.117

Figure 2 shows the evaluation results of Knowledge Precision/Recall/F1 metrics, and it can be seen that the model proposed in this paper outperforms the baseline model in Knowledge Precision and Knowledge F1 metrics, and is lower than the GLKPG model but outperforms the other models in Knowledge Recall metrics. GLKPG has the highest Knowledge Recall value, which indicates that the GLKPG model introduces more words related to background knowledge in the reply statements, which explains why the GLKPG model has the highest Distinct metric value. With the advantage of knowledge replication mechanism, the model proposed in this paper utilizes more background knowledge than MemNet and PostKS models, thus Knowledge Precision/Recall/F1 is higher.

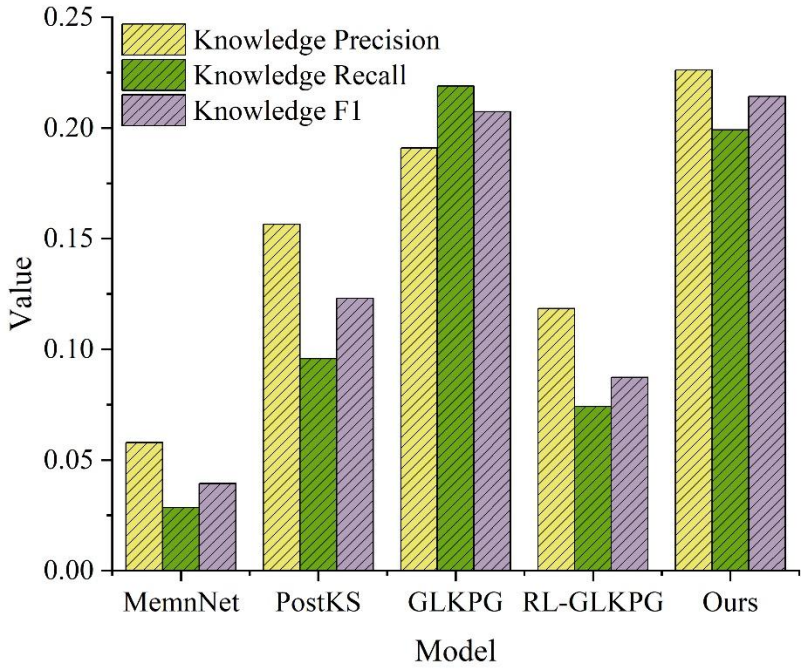


Figure 2. Knowledge precision/ Recall/F1 automatic evaluation results.

3.2.2. Manual Evaluation Results

Table 2 shows the results of the manual evaluation, which were put together with the conversation data simulated by the baseline model based on the same background knowledge and starting sentences for comparison, respectively, and the evaluators made the following judgments for each of the four metrics: (1) the conversations produced by the model proposed in this chapter are better-win; (2) the conversations of the baseline model are better-lose; and (3) tie-tie.

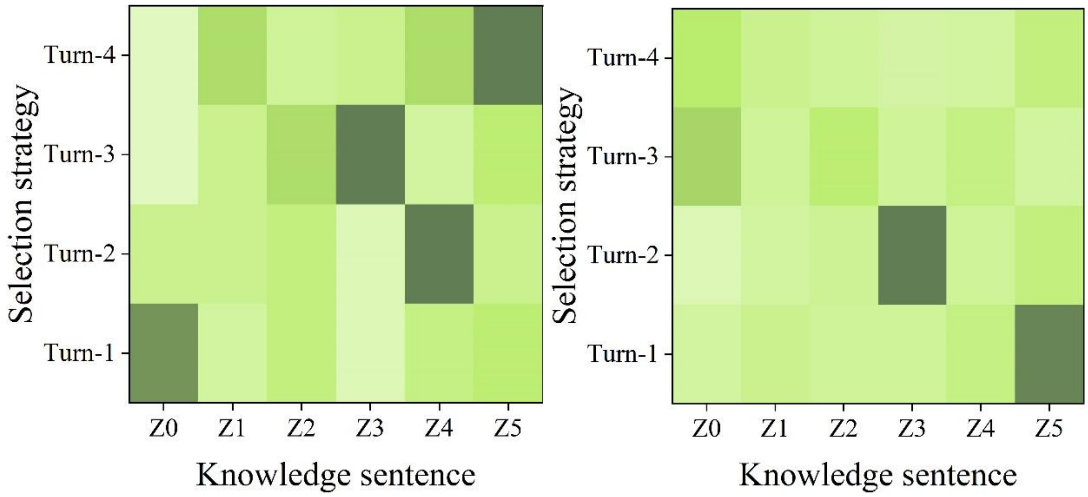
As can be seen from the table, the win rate of the model proposed in this chapter is higher than 50% in all the indicators when comparing with each baseline model, which indicates that the model proposed in this chapter is more advantageous in the global perspective of multi-round conversations. For the coverage metrics, the table shows that the advantage of the model proposed in this chapter over GLKPG is not obvious (the winning rate is only 53%), indicating that there is not much difference between the two models in terms of the amount of background knowledge used in the perspective of the whole conversation. The method proposed in this paper encourages the model to try to select previously unused knowledge during knowledge selection by rewarding information, so it can greatly improve the knowledge coverage while avoiding the repetition of dialog content, which is more advantageous in terms of conciseness metrics.

Table 2. Artificial evaluation results.

Control group	Ours vs MemNe			Ours vs PostKS			Ours vs GLKPG			Ours vs RL-GLKPG		
	win	lose	tie	win	lose	tie	win	lose	tie	win	lose	tie
Coherence%	86	9	10	79	16	11	83	13	9	88	9	7
Coverage ratio%	94	6	4	89	10	7	53	29	22	77	12	15
Simplicity%	71	19	17	74	12	19	81	10	14	86	8	11
Consistency%%	79	6	18	72	14	19	60	26	21	75	17	14

3.2.3. Analysis of Model Generation Results

In order to have a deeper understanding of the knowledge selection strategy of dialogue agents, the knowledge choice distribution of each dialogue agent is visualized here, as shown in Figure 3, the figure gives the probability of two conversation agents selecting each knowledge sentence Z_i in each round of dialogue, the darker the color represents the higher the probability of choosing Z_i, where Z_0 represents the probability of choosing empty knowledge "no_knowledge_use", indicating that no background knowledge is used to generate responses. It can be seen that as the dialogue progresses, the two dialogue agents adjust their knowledge selection strategies in each round, select different knowledge for different conversation contexts and express knowledge information in the reply statement, which indicates that the knowledge selection strategy of multi-round dialogue agents is effective through reinforcement learning.



(a) Person 1 (b) Person 2

Figure 3. Knowledge selection distribution.

3.2.4. Diversity Analysis

For qualitative evaluation of the models, some dialogues generated by different models are shown later. Due to space constraints, two dialogs generated by RL and RL-AHRED+LaRL are randomly selected for comparison. We exemplify two conversations (within 5 turns) for Sw DA and one conversation for SinaWD. the first turn is entered by the user. Then the two agents take turns in a dialog with the other agent's previously generated discourse as input. The results are shown in Tables 3 and 4. From the two tables, it can be observed that RL-AHRED+LaRL tends to generate more specific and meaningful dialogs for factual issues.

Table 3. Dialogue simulation output of two models on Twitter data.

Turn	DRL	Ours
A	Where do you live?	Where do you live?
B	I'm Dallas	I lived in Chicago
A	Really?	I love Chicago
B	Yes,you will?	That's right
A	All right	yeah
A	What do you like to cook	What do you like to cook?

B	I like to cook	I like to cool Italian food
A	I have. have you?	It is good for a dinner party
B	Yeah	We had a party
A	Uh,I don't know	It's really good

Table 4. Dialogue simulation output of two models on Weibo data.

DRL	Ours
I want to quit smoking. Do you believe it?	I want to quit smoking. Do you believe it?
Mutual concern	Really, you sure?
I sleep	Tired and happy
Okay	Be confident
Thank you	Haha, ok

By documenting the training process, the proposed model generates more informative replies if the model explores strategies in the potential word space. Figure 4 shows the training process of word level strategy on inputs with reinforcement learning, it can be seen that as the number of rounds increases, the reward value is increasing, which means that the responses are reasonable, but the diversity of responses is missing. However, strategy exploration of the potential word space is shown in Figure 5, where the responses are both reasonable and diverse as the number of rounds increases.

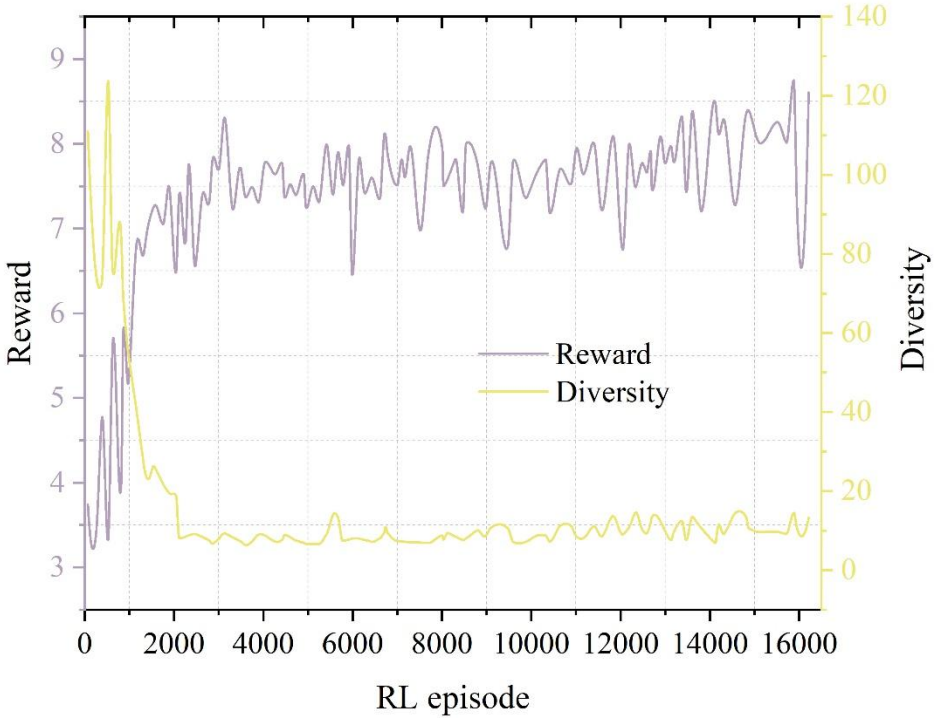


Figure 4. Diversity and Task Reward Learning Curves in Level Word RL Training.

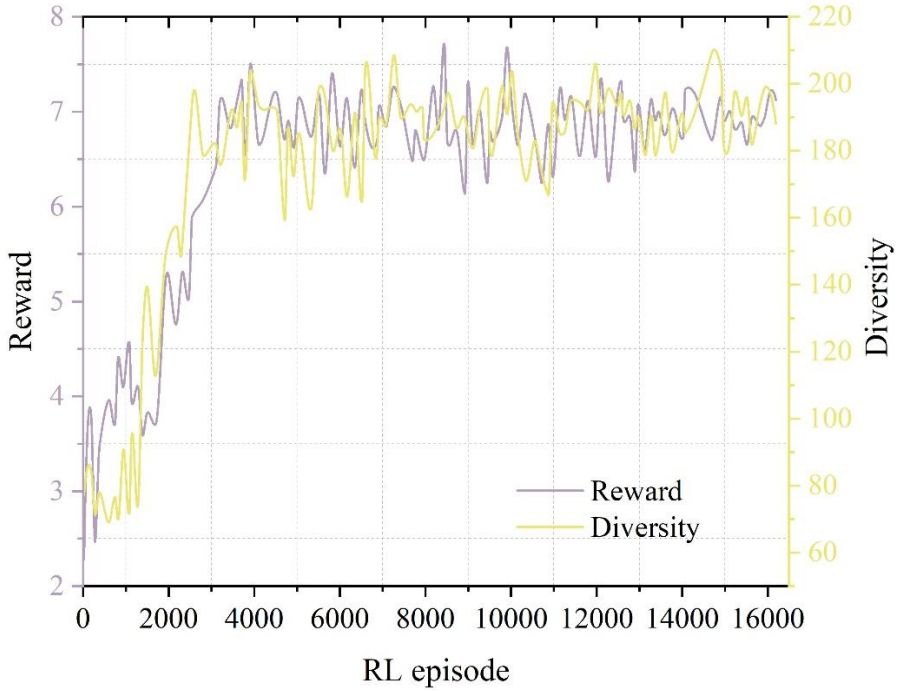


Figure 5. Diversity and task reward learning curve during latent word RL training.

3.3. System Non-Functional Testing

Functional testing of the system is mainly based on the functional requirements of the system to verify whether each functional point meets the expectations. The following are the test cases designed in the three major functional modules of instant messaging communication, classroom teaching, and backend management. This subsection mainly describes the testing of the concurrency of the system. The results of the system concurrency test with different number of users are shown in Table 5. From the results, it can be concluded that the system basically operates normally when the number of concurrently operating users of the system does not exceed 310. When the number of simultaneous users exceeds 310, the system is unable to respond successfully to all requests, and when the number of users is 400, the system can support 96.8% of the users to complete the operation. When the number of users increases further, the availability of the system decreases significantly, and when the number of simultaneous operations reaches 700, the system can only support about 55% of the users to successfully complete the request.

Table 5. Concurrency test results.

The number of users	Mean response time /ms	Successful response number
20	48	20
60	75	60
160	145	160
210	243	210
270	326	270
310	472	310
400	785	387
510	1203	436
600	2031	492
700	2281	387

3.4. Effectiveness of English Language Teaching

Two classes of a foreign language school in Zhengzhou City with a total of 81 students were selected for the experiment, respectively class A (39) and class B (42), considering whether the distribution state of the intelligent teaching system on the student's performance is reasonable, i.e., whether it is normally distributed or not, this paper carried out a K-S test on the data as shown in Table 6. As can be seen from

the table, the probability of significance of the overall achievement of the sample P = 0.0089, P < 0.05 and close to 0, indicating that the sample basically obeyed the normal distribution. This test on the normal distribution of grades facilitates teachers to select the best and screen the laggards. The test results show that more than 60% of the students scored more than 80 points, which can be seen that this course basically meets the teaching requirements.

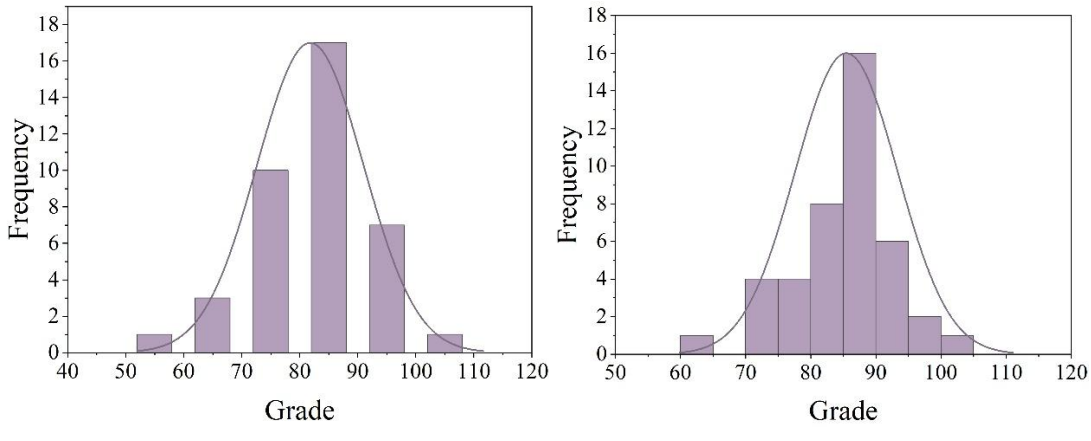
Table 6. The single sample, kormogoov, was tested.

Test type		Serial number	Final grade
Case number		81	81
Normal parametera,b	Mean value	41.24	81.7351
	Standard deviation	23.743	8.27641
Extreme difference	Absolute	0.067	0.158
	Positive	0.067	0.113
	Negative	-0.067	-0.162
Inspection statistics		0.067	0.158
Asymptotically significant (double tail)		0.205c,d	0.001c

a. test that the distribution is normal; b. calculated from the data; c. corrected for Riley's significance; d. this is the lower limit of true significance.

Considering the differences in the professional skills courses of the two classes in the English program: when Class A took two courses of Financial Management and Investment, Class B took two courses of International Trade and Fundamentals of Management, this paper intends to conduct a t-test on the final grades of Practical Speech Rhetoric of the two classes in the sample in order to compare the means of the grades of Practical Speech Rhetoric of the two directions of the English majors in the program in order to test whether the grades of the two classes there was a significant difference and whether the final grades were influenced by intelligent teaching aids.

First of all, a basic analysis of the two classes' scores was carried out, and it was found that the mean score of Class A was 81.4, with a standard deviation of 9.77; the mean score of Class B was 85.35, with a standard deviation of 7.24. The difference between the lowest scores of the two classes was 25, the difference between the highest scores was 10, and the difference between the mean scores was 4, so it can be initially decided that Class B's scores were better. The comparison of standard deviation also reveals that the performance of class B is more stable. Then the sample t-test for the two classes' grades, the results show: Significance (two-tailed) probability p=0.000<0.05, that is, there is a significant difference between the final grades of English majors in two different directions. The histogram normal distribution of the final grades of the two classes is shown in Figure 6. From the histogram distribution of the two classes in the figure and the shape of the normal distribution curve, we can intuitively see the differences in the grades of the two classes; Class A is obviously more dispersed and has lower grades, while Class B is obviously more stable and has higher grades. The objective factors such as lecturers, classroom materials and teaching methods are the same in both classes, but there is a significant difference in the grades, which indicates that the learning efficiency of the two classes in the course of Practical Speech and Rhetoric may have been affected by the learning of related professional skills courses.



(a) A class (b) B class

Figure 6. The histogram and normal distribution curve of the last grade of the two pe.

4. Conclusion

This paper proposes a dialogue generation model based on reinforcement learning, which can be optimized by the policy gradient reinforcement learning algorithm, and can effectively solve the problem of large hidden variable space faced by the model, and designs an intelligent dialogue system for English language courses on the basis of the dialogue generation model. It is tested and analyzed by selecting datasets as well as actual cases. After the experiments, it is shown that the dialog generation model proposed in this paper has a better performance compared with other baseline models, with a winning rate of more than 50% in all indicators. The system can respond positively when facing different numbers of users, and it can be seen in the actual cases that the intelligent teaching system is of great help to teaching, and the data-based analysis of test papers and grades can be evaluated objectively.

References

1. Tafazoli, D., María, E. G., & Abril, C. A. H. (2019). Intelligent language tutoring system: Integrating intelligent computer-assisted language learning into language education. International Journal of Information and Communication Technology Education (IJICTE), 15(3), 60-74.
2. Chaib, S. O., Joti, I., & Khoulji, S. (2023). Evaluation of a computer-assisted language learning system based on adaptive learning designed for self-training in scientific French language. International Journal of Information and Education Technology, 13(8), 1284.
3. Lim, M. H., & Aryadoust, V. (2022). A scientometric review of research trends in computer-assisted language learning (1977–2020). Computer Assisted Language Learning, 35(9), 2675-2700.
4. Prastikawati, E. F. (2019). Dyned programme as Computer Assisted Language Learning (CALL) for university students: A perception and its impact. International Journal of Emerging Technologies in Learning (Online), 14(13), 4.
5. Tang, J., & Deng, Y. (2022). The Design Model of English Graded Teaching Assistant Expert System Based on Improved B/S Three-Tier Structure System. Mobile Information Systems, 2022(1), 4167760.
6. Sun, Z., Anbarasan, M., & Praveen Kumar, D. J. C. I. (2021). Design of online intelligent English teaching platform based on artificial intelligence techniques. Computational Intelligence, 37(3), 1166-1180.
7. Fu, Y., Zhang, Z., & Yang, H. (2023). Design of Oral English Teaching Assistant System based on deep belief networks. Soft Computing, 27(22), 17403-17418.
8. Yuan, S. (2024). Personalized College English Learning Experience Assisted by Artificial Intelligence: An Algorithm-Driven Adaptive Learning Approach. International Journal of High Speed Electronics and Systems, 2540157.
9. Chanda, R. C., Vafaei-Zadeh, A., Hanifah, H., & Ramayah, T. (2025). Artificial intelligence teaching assistant adoption in university education: Key drivers through the ability, motivation and opportunity framework. Education and Information Technologies, 1-42.
10. Ma, Y. (2025, January). Design of intelligent teaching assistant for college English based on human-computer collaboration. In Fifth International Conference on Signal Processing and Computer Science (SPCS 2024) (Vol. 13442, pp. 668-674). SPIE.
11. Meng-yue, C., Dan, L., & Jun, W. (2020). A study of college English culture intelligence-aided teaching system and teaching pattern. English Language Teaching, 13(3), 77-83.
12. Paladines, J., & Ramirez, J. (2020). A systematic literature review of intelligent tutoring systems with dialogue in natural language. IEEE Access, 8, 164246-164267.
13. Razumovskaia, E., Glavas, G., Majewska, O., Ponti, E. M., Korhonen, A., & Vulic, I. (2022). Crossing the conversational chasm: A primer on natural language processing for multilingual task-oriented dialogue systems. Journal of Artificial Intelligence Research, 74, 1351-1402.
14. Yan, F. (2025). Machine Learning-assisted Intelligent Teaching System (ML-ITS) for College Online English Learning. International Journal of High Speed Electronics and Systems, 2540224.
15. Li, X. (2024, October). Intelligent Tutoring System for English Language Acquisition Based on Computer Algorithms. In 2024 3rd International Conference on Data Analytics, Computing and Artificial Intelligence (ICDACAI) (pp. 225-230). IEEE.
16. Li, X., Long, X., Long, Y., Chen, S., & Chen, Z. (2020, August). An Intelligent System of Oral English Assistant Teaching Based on Tag Matching. In The International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (pp. 588-595). Cham: Springer International Publishing.
17. Wu, Q. (2021, October). Design Scheme of Intelligent English Teaching Aid System Based on Artificial Intelligence Technology. In 2021 3rd International Conference on Artificial Intelligence and Advanced Manufacture (pp. 2726-2729).
18. Sajja, R., Sermet, Y., Cwiertny, D., & Demir, I. (2023). Platform-independent and curriculum-oriented intelligent assistant for higher education. International journal of educational technology in higher education, 20(1), 42.
19. Zhai, C., & Wibowo, S. (2023). A systematic review on artificial intelligence dialogue systems for enhancing English as foreign language students' interactional competence in the university. Computers and Education: Artificial Intelligence, 4, 100134.

20. Zhai, C., Wibowo, S., & Li, L. D. (2024). Evaluating the AI dialogue System's intercultural, humorous, and empathetic dimensions in English language learning: A case study. Computers and Education: Artificial Intelligence, 7, 100262.
21. Bibauw, S., Van den Noortgate, W., François, T., & Desmet, P. (2022). Dialogue systems for language learning: A meta-analysis. Language Learning & Technology, 26(1), 1-24.
22. Ying'an Wei,Jingjing Fan,Qinglong Meng,Kumar Biswajit Debnath,Yuqin Yang,Jiao Liu & Yu Lei. (2025). EOLD: A reinforcement learning-based energy-optimised load disaggregation framework for demand-side energy management. Renewable Energy,252,123536-123536.
23. François Dufour,Alexandre Génadot & Romain Namyst. (2025). The bearing only localization problem via partially observed Markov decision process. Mathematical Methods of Operations Research,101(2),1-39.
24. Ming Sun,Zexu Jiang,Erhan Dong & Tianyu Lv. (2025). A distributed multi-agent joint optimization algorithm based on CERL and A2C for resource allocation in vehicular networks. Vehicular Communications,53,100919-100919.