Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 7876504, 8 pages
https://doi.org/10.1155/2022/7876504

Research Article

English Language Learning Pattern Matching Based on Distributed Reinforcement Learning

Hua Zhao

College of Arts, Xinxiang University, Xinxiang, Henan 453000, China

Correspondence should be addressed to Hua Zhao; zhaohuaamy99@xxu.edu.cn

Received 16 April 2022; Revised 24 May 2022; Accepted 25 May 2022; Published 7 September 2022

Academic Editor: Zaoli Yang

The rapid development of a new generation of information technology, the promotion of network technology, and the emergence of complex and diverse requirements for control objects make the structure of language learning models more and more distributed. Distributed learning theory emphasizes the central position of learners in the learning process and the universality of learning scenes. This paper explores the significance and value of various learning modes to improve students' learning effect. By analyzing the research data and explaining various effective language learning models, this paper aims to establish a theoretical framework of English language learning models and explore more effective language model matching schemes. This paper analyzes the adaptive multiagent, reward function, Markov model, probability function model, etc. and conducts experiments on the basis of the designed model. The linear correlation parameters of the model and the English language pattern matching efficiency are analyzed and judged on several important indicators. Because the algorithm designed in this paper has a good effect on the control of error, the error reduction rate has reached 85.6%.

1. Introduction

Cooperation is an important trend in the development of education in the future, and collaborative learning has gradually become the main way for educators to carry out teaching and learning activities. In this context, it has quickly attracted widespread attention from experts and scholars in various fields [1]. In the cultivation of the four basic abilities of English listening, speaking, reading, and writing, the effect of English language teaching has always lagged behind that of other subjects [2]. From the perspective of the teaching process, teaching is dominated by the language communication between teachers and students in a suitable educational environment, so the way and frequency of interaction between teachers and students play a vital role in oral English teaching [3]. With the in-depth development of technical education applications, the scale and complexity of information-based learning resources and learning systems are increasing. We regard distributed learning as a learning method that breaks the boundaries of time and space through computer networks and information technology and provides learners with rich information-based learning resources and a good network learning support environment [4]. Its technical means, design ideas, and system architecture have undergone profound changes, and the teaching system is developing in the direction of distributed, collaborative, and intelligent way [5]. On the one hand, with the help of technology, the scale of learning system is expanding day by day, and it has distinct distributed characteristics in structure; on the other hand, people hope to realize the unified sharing, reuse, and interoperability of these distributed learning resources and systems.

The ultimate goal of English teaching is to cultivate students' comprehensive language application ability. In fact, the individual's cognition needs to be continuously developed for a long time, but learning is completed through assimilation or adaptation to external stimuli and finally reaches a state of equilibrium, then beginning a new stage of learning [6]. The core of this way is to emphasize cognitive conflict. In teaching, it is particularly important for teachers to trigger students' cognitive conflict. In language teaching, we usually measure the development level of students' language ability by their four skills of listening, speaking, reading, and writing [7]. In order to study the effect of

students' English language learning, improve learning efficiency, construct dynamic knowledge structure, and measure the role of each mode, the modeling method of distributed learning system is a set of systematic engineering methods, which analyze the requirements of distributed learning system, designs distributed learning system, and establishes the software model of distributed learning system from two aspects of model abstraction and representation [8]. The distributed learning system modeling method is based on object-oriented method, constructivist learning environment design method, and software process method.

Since entering the 21st century, with the continuous development and maturity of computer science and technology and digital media technology, new technologies such as mobile Internet, artificial intelligence, big data, and cloud computing are strongly changing the social culture and economic form at an unprecedented speed and momentum [9]. Distributed learning environment is a kind of learning environment based on distributed cognitive theory, which aims to call all technical means to provide the same learning place and communication place for geographically dispersed learners. Learning is any improvement in a system that makes the system do better or more efficiently when repeating the same work or doing similar work [10]. Reinforcement learning is a learning mechanism for matching learning through the interaction between agent and dynamic environment. It is a trial and error learning method. Trial and error and delayed reward are the two most important features of reinforcement learning [11]. Generally, the states and actions of the reinforcement learning system are considered as discrete and finite sets, and the value function can be expressed by look-up table method [12]. In practical applications, there are a large number of system states or actions that are continuous or both [13]. This paper will adopt the research method of combining qualitative research and quantitative research. Through the research, the connotation of English language learning and the research status of mobile learning are sorted, analyzed, and refined.

This paper discusses the significance and value of various learning modes to improve students' learning effect. The innovative contribution of analysis and research lies in establishing the theoretical framework of English learning model and exploring more effective language pattern matching schemes. Various effective language learning models are explained based on the data level. Reinforcement learning technology and agent technology are introduced into the research of adaptive system. Based on the dynamic binding mechanism, the adaptive mechanism based on reinforcement learning is proposed, and the corresponding learning algorithm is proposed to support the learning process of adaptive agent.

2. Methodology

2.1. Research on Distributed and Reinforcement Learning. As a cognitive theory including cognitive subject and environment, distributed cognition advocates placing individual cognitive activities in context and social culture and emphasizes that cognitive phenomena are widely distributed within individuals, between individuals, media, learning environment, social culture, and time, that is, the analysis element system covering cognitive subject and environment and all things involved in cognition. Distributed learning is a teaching mode. It allows teachers, students, and content to be distributed in different noncentral places. This makes teaching and learning independent of time and space. Therefore, distributed learning seems to have similar characteristics or some connection with distance education. Distributed learning is not a new term to replace "distance learning." It comes from the concept of "distributed resources." Therefore, distributed learning is a teaching model based on noncentral storage of learning resources. Its pancentralization reflects the independence of teaching and learning, so that learning will not be depressed in a single form of learning, showing a trend of diversification. Teaching interaction is the interaction between learners and learning environment in the process of learning. At present, the combination of reinforcement learning and other technologies is also one of the focuses of research [14]. In the single agent environment, the most common technologies combined with reinforcement learning are genetic algorithm and neural network. The reason is that genetic algorithm and neural network also have strong white adaptability, so it is easier to combine with agents that emphasize initiative [15]. The current combination of reinforcement learning and other techniques is also one of the focuses of research. In a single agent environment, the most common technology combined with reinforcement learning is genetic algorithm and neural network. The reason is that genetic algorithm and neural network also have strong white adaptability, so it is easier to combine with agents that emphasize initiative [16]. Figure 1 shows the basic model of reinforcement learning.

Because of learning algorithms, they are generally divided into three categories: unsupervised learning, supervised learning, and reinforcement learning [17]. Since unsupervised learning is usually the same as Pavlov's conditioning principle, the learning system will adjust parameters and distribution characteristics according to the data provided by it, and it is not a closed loop [18]. Supervised learning is a learning method with feedback mechanism, as shown in Figure 2. Generally, there will be error signals to express the feedback content [19]. Learning is mainly manifested in the signal provider, and there will be signals generated by the environment to evaluate the quality of the generated actions, which are usually listed as standard parameters.

In addition to the agent and the environment, a comprehensive reinforcement learning system should also have other components, such as reward and value functions and a model of the environment [20]. The value set as s_t is the expectation of accumulated reward obtained during the execution of action a_t and subsequent strategy π, which is generally expressed as $V(s_t)$. Then, there are

$$V(s_t) = E\left(\sum_{i=0}^{\infty} \gamma^i r_{t+i}\right). \quad (1)$$

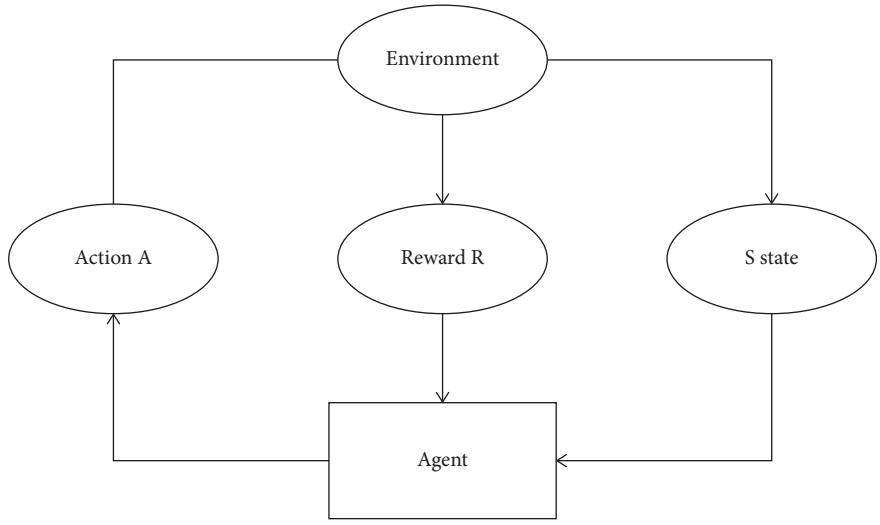

Figure 1: Basic model of reinforcement learning.

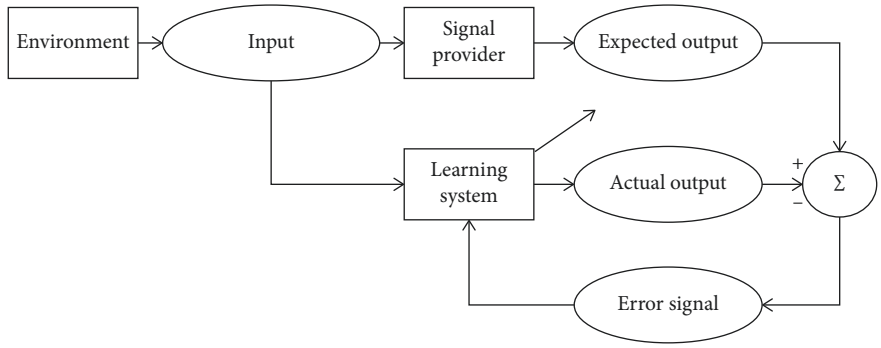

Figure 2: Learning framework diagram of supervised learning.

$r_t = R(s_t, a_t)$ is the reward of t moment. Then, for any strategy π, there will be a defined value function whose expected value is

$$V_\pi(s) = E_\pi\left(\sum_{t=0}^{\infty} \gamma^t r_t | s_0 = s\right). \quad (2)$$

Among them, r_t, s_t are the immediate reward and state at time t, respectively, and the decay coefficient $\gamma(\gamma \in [0, 1])$ makes the adjacent reward more important than the future reward. Figure 3 is a schematic diagram of an adaptive multiagent system.

Figure 3 shows a high-level abstract view of adaptive multiagent. An adaptive multiagent system is generally composed of multiple adaptive agents. Each adaptive agent is relatively independent and resides in the corresponding environment [21]. Adaptive agents will interact with each other. For example, when sharing a common resource, adaptive agents will negotiate with each other.

2.2. Analysis of English Language Matching Models. The language matching system is a typical distributed control system. Generally, it is considered that each language matching system utilizes local and neighbor information to design a coordinated control protocol, so that multiple language matching systems can obtain specific collective behaviors, such as formation keeping and assembly [22]. Based on different needs and applications, learning systems are rich in types and forms, and the specific technologies that they rely on are also different. For the distributed linear control system with unknown disturbance, firstly, the distributed output cooperative control protocol is designed when the system matrix is known, and then the minimization performance index is introduced to obtain the distributed optimal solution combined with the optimization method [23]. Then, the output coordinated optimal control problem with unknown system matrix is considered. The interactive process of collaborative learning in cloth learning environment reflects a multidimensional interaction, which can be summarized as two aspects: the interaction between subjects and the interaction between subjects and objects [24]. The collaborative interaction skills that run through them are a key point that cannot be ignored. Consider a multiagent system consisting of N linear subsystems, each matching A_i can be represented as a subsystem A_i, regarded as a tracker, and described by the following dynamic equations:

$$\dot{x}_i(t) = A_i x_i(t) + B_i u_i(t) + \Delta_i(t), \tag{3}$$

$$y_i(t) = C_i x_i(t) + D_i u_i(t), \tag{4}$$

where $x_i \in R^{ni}, u_i \in R^{mi}, y_i(t) \in R^{pi}$ represent the status, input, and output of subsystem A_i respectively. $\Delta_i(t)$ indicates that the subsystem A_i is subject to external disturbance. $x_{i0}, i = 1, \ldots, N$, indicates the initial state of subsystem A_i. At this time, an additional autonomous system can be described, which can be expressed as

$$\dot{\omega}(t) = S\omega(t), \omega(0) = \omega_0. \tag{5}$$

The autonomous system is used to generate the external disturbance to be suppressed and the rated reference output signal A_i of subsystem $y_{ri}(t) \in R^{pi}$, which are, respectively, expressed as follows:

$$\Delta_i(t) = E_i \omega(t), \tag{6}$$

$$y_{ri}(t) = F_i \omega(t). \tag{7}$$

There is A_i when subsystem $E_i = 0$ is not subject to disturbances from external systems. At this point, a multigroup interactive data table about distributed reinforcement learning can be obtained (Table 1).

The task of agent is to learn a strategy π: $S \longrightarrow A$ to maximize the return value of the action selected by agent from the environment. How to quantitatively define the learning strategy π is the primary factor that agents need to consider when learning. In addition, agents should also consider the long-term impact of choosing actions, that is, whether the actions of agents are optimal in the long run. Therefore, three objective functions of reinforcement learning need to be obtained:

$$V^{\pi}(s_i) = \sum_{k=0}^{\infty} \gamma^k r_{i+k}, \tag{8}$$

$$V^{\pi}(s_i) = \sum_{k=0}^{h} r_{i+k}, \tag{9}$$

$$V^{\pi}(s_i) = \lim_{h \longrightarrow \infty} \left(\frac{1}{h} \sum_{k=0}^{h} r_{i+k} \right). \tag{10}$$

The objective function $V^{\pi}(s_i)$ in (8) is called the converted cumulative return, and γ is the discount factor. However, it is found through calculation that γ in the above formula reflects the degree of importance that agent places on the future. The larger the value is, the more important the future return is. The objective function in (9) becomes a finite level return. At this time, only the cumulative return of finite steps in the future is considered. The objective function in (10) represents the average return. At this time, the average return of the whole cycle is considered [25]. For individuals, language has dual functions. On the one hand, individuals can communicate with the outside world through language. On the other hand, language can promote the development of its own internal language. Therefore, language unifies the individual's social development and thinking development [26]. When the learner produces the correct words, reinforcement is carried out; when the words produced by the learner are wrong, they are corrected. Strengthening and correction are actually the purpose of feedback, but the behavioral view ignores the cognitive function of the learners themselves. Analyzing the environment of adaptive agent is an important work to describe the behavior of adaptive agent [27]. The environment change of adaptive agent is a continuous process. In order to simplify the description, this paper regards the environment as a series of discrete finite states; namely,

$$S = \{s_1, s_2, \ldots, s_n\}, \tag{11}$$

where S represents the state set of the environment, which includes all kinds of environments. Setting this parameter is beneficial for this paper to make a reasonable and accurate judgment on the environmental changes of adaptive agent. A distributed system is to integrate some application systems with limited functions into a more powerful application system to meet the application requirements that any single application system cannot meet. Because of this modularity, the structure of the whole system is very flexible and the system has strong adaptability. It can be flexibly combined, constructed, and even automatically adjusted according to the needs of the actual system, which brings great convenience to the functional design and implementation of the system.

Assuming that there are c_i kinds of changes in the state component sc_i, the environment in which the agent is located can be divided into $c_1 c_2 \ldots c_m$ kinds of states. If different states of state components are represented by integers in $[0, c_i - 1]$, then the state set of the environment can be represented as follows:

$$S = \left\{ \begin{bmatrix} 0 \\ 0 \\ \ldots \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ \ldots \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} c_1 - 1 \\ c_2 - 1 \\ \ldots \\ c_m - 1 \end{bmatrix} \right\}. \tag{12}$$

The above state components are related to specific applications. When designing a specific match, it is necessary to carry out the corresponding state components according to the definition and define the corresponding function to collect the current state of these state components to ensure the correctness of the data for subsequent calculations and source reliability. The reinforcement function defines the quantitative evaluation of agent action by environment. The design of reinforcement function is based on the specific application. Generally, the actions that have a positive impact on the learning process are given a larger return value, and the actions that have a negative impact are given a smaller return value. The adaptive agent will gradually tend to choose the actions with a larger return value in the learning process.

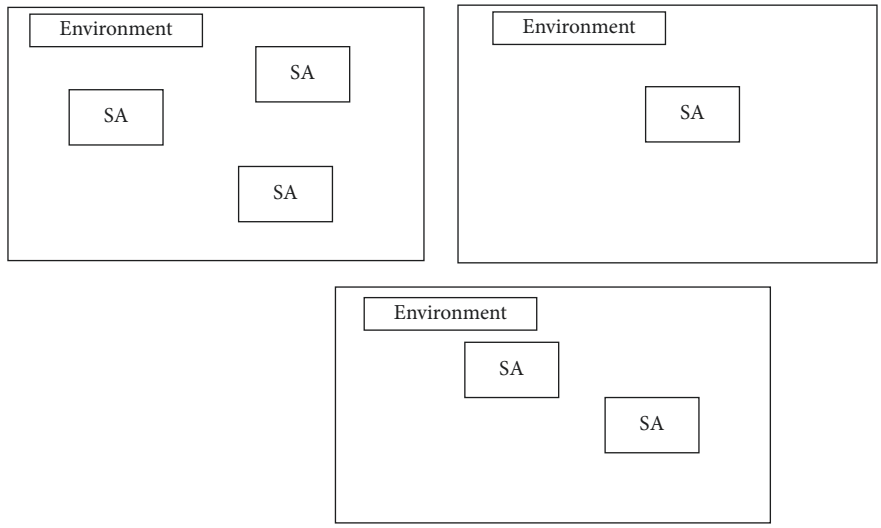

Figure 3: Schematic diagram of an adaptive multiagent system.

Table 1: Multigroup interaction data table for distributed reinforcement learning.

	Utilization ratio	Frequency
Sample group 1	0.422	0.657
Sample group 2	0.351	0.555
Sample group 3	0.236	0.243
Sample group 4	0.574	0.274
Sample group 5	0.236	0.658

In reinforcement learning system, the key assumption based on English language matching is that the interaction between agent and environment can be regarded as a Markov matching model. Therefore, on the basis of the above research, this paper continues to analyze the problem processing method for Markov and set the time point on the time series as t. Therefore, Markov's matching process can be composed of five elements:

$$\langle S, A(s), P_{ss}^{a}, R_{ss}^{a}, V | s, s \in S, a \in A(s) \rangle. \tag{13}$$

When the system is in the state s at the matching time point t , the matching a is executed; the system will get the probability P_{ss}^{a}, when the next matching is carried out. At this time, all actions will get a matrix composed of transition probability, which can be expressed as

$$P_{ss'}^{a} = \Pr\{s_{t+1.} = s | s_t = s, a_t = a\}. \tag{14}$$

When the system is in the state s at the matching time point t, after the execution state a, the system will get timely reward R_{ss}^{a}, which is generally a reward function:

$$R_{ss}^{a} = E\{r_{t+1} | s_t = s, a_t = a, s_{t+1} = s\}. \tag{15}$$

If the transition probability function P_{ss}^{a} and the reward function R_{ss}^{a} have nothing to do with the matching time t, at this time, the amount does not change with the amount at the time point, and it is in a stable state. The state at this time can be expressed as

$$Pr(h_1) = Pr\{s_{t+1} = s, r_{t+1} = r | s_t, a_t, r_t, s_{t-1}, a_{t-1} \ldots r_1, s_0, a_0\}. \tag{16}$$

For any parameter, it can be said that the environment has Markov property, if there are

$$Pr(h_1) = Pr\{s_{t+1} = s, r_{t+1} = r | s_t, a_t\}. \tag{17}$$

At this time, the comparison parameter tables under different models can be obtained, as shown in Table 2.

Many teaching aids in traditional classrooms are visualization tools, such as chalk and paper. For students, they can use visualization tools to extract information from images, understand the deeper meaning of words, and exchange ideas with each other. Taking activity theory, distributed cognitive theory, and conversation theory as the theoretical basis and taking into account the core idea of situational cognitive theory, this paper proposes a conversation activity distributed theoretical model, which is called CAD for short. At the same time, CAD applies conversation theory, activity theory, and distributed cognitive theory. The three theories support CAD from different angles. Conversation theory and activity level can provide the basis for the concrete interaction of CAD, and activity theory can provide support for the system design based on CAD and how to make the system more effective. Creating a harmonious, reciprocal, and orderly group culture in a distributed English language environment is a prerequisite for the effective occurrence and development of collaborative English language interaction. At the same time, it plays a vital role in narrowing the matching relevance between English language learners and between students and teachers, stimulating students' initiative to participate in interactive communication, and maintaining the orderly development of collaborative interaction process. Carry out rich and colorful extracurricular activities, such as actively carrying out English speech competitions and other activities. Combining the two teaching channels is an effective way to improve the effect. In short, in the teaching of new

Table 2: Comparison of parameters of Markov model, reinforcement learning model, and distributed model.

	Markov model	Reinforcement learning model	Distributed model
Matching efficiency	0.57	0.63	0.53
Error parameter	0.45	0.35	0.54

textbooks, creating context can cultivate students' communicative ability and consolidate students' basic knowledge and skills. It can stimulate students' interest in learning English and stimulate their initiative. It is an effective means to achieve teaching success.

3. Result Analysis and Discussion

In order to establish a scientific, feasible, and high-efficiency English language matching model, based on the above research and analysis, this paper makes further experimental analysis, so as to confirm the reliability of pattern design and observe whether the model can match English language to the corresponding pattern in practice. To this end, this paper will analyze and judge several important indicators such as distributed self-adjustment efficiency, interactive correlation degree of reinforcement learning, linear correlation parameters of Markov model, and English language pattern matching efficiency. This paper selects three English learning sample sets from ordinary colleges and universities for efficiency analysis. Figures 4 and 5 are the analysis diagrams of the interactive correlation between distributed self-moderation efficiency and reinforcement learning on sample sets A, B, and C.

Therefore, it can be observed that in the distributed self-regulation efficiency and interactive correlation of reinforcement learning, it has a good impact on the self-regulation and correlation of English language patterns, which provides good reliable data and correlation coefficient for the matching with the next analysis. In the distributed self-adaptive efficiency, it is observed that it has a good test effect on sample set C, and, in the overall trend, it can also be known that the efficiency is gradually improved with the increasing of the numerical value on the quantization axis, which also reflects the advantages of the model in data processing and also has a good analysis effect for a large number of data. For the interactive relevance of reinforcement learning, in this paper, reinforcement learning is an important indicator of the entire English language learning model, and it is also of great significance to the matching model, because the interactive relevance is important for the model to push a reasonable and correct English language. The learning model has a decisive influence. Therefore, it can also be found in the experiment that the test results on the sample set C have always shown good results. This is also because the sample set C also has a good advantage in distributed self-adjustment. Therefore, it will also achieve good results in terms of interaction. When measuring the linear correlation parameters of Markov model and the

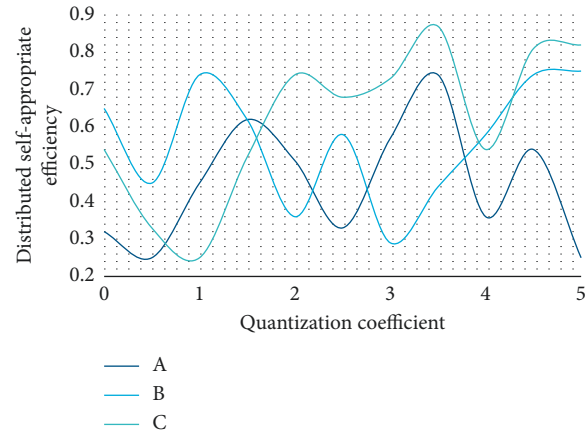

Figure 4: Analysis diagram of distributed self-moderation efficiency.

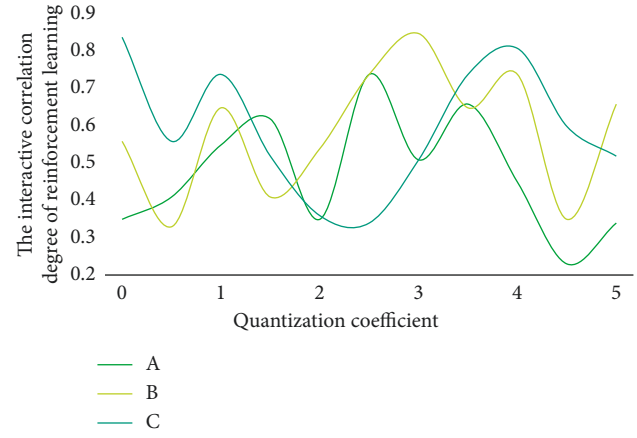

Figure 5: Analysis diagram of the interaction degree of reinforcement learning.

efficiency of English pattern matching, another two sets of sample sets Q1 and Q2 are used in this paper. Figures 6 and 7 are the analysis charts of linear correlation parameters of Markov model and English pattern matching efficiency on sample sets Q1 and Q2.

The above two parameters are a direct comparison of the model. Through the experiment, it can be found that the linear correlation parameters of Markov model are generally stable in Q1 but relatively large in Q2. This may be because the sample set in Q1 is concentrated and the correlation in English language is strong, This will lead to the Markov model's judgment that it has high linear correlation, and the transformed graphics will be stable, while, in Q2, on the contrary, due to the lack of concentration of language, the resulting graphics will be divergent and unstable. In the efficiency of English language pattern matching, it can be found that, in the last 4-5 stages, Q1 and Q2 have assimilation phenomenon. This is because the algorithm designed in this paper has a good effect on the control of errors, and the error reduction rate has reached 85.6%. This will have a great impact on the results. Although the results are somewhat deviated due to the difference of sample sets and

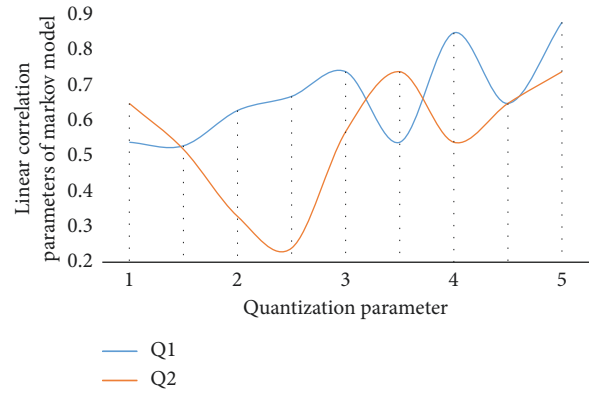

Figure 6: Linear correlation parameter analysis diagram of Markov model.

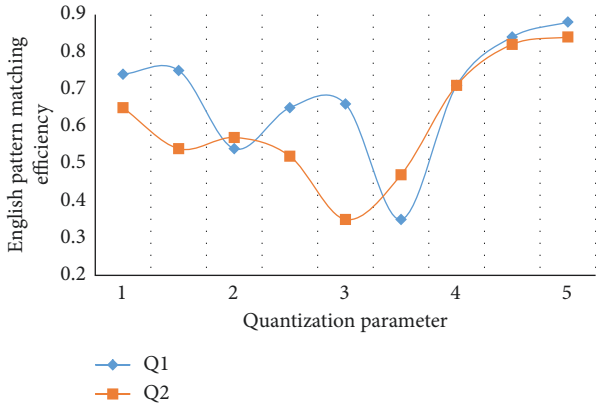

Figure 7: Analysis of English language pattern matching efficiency.

the inconsistency of calculation flow in the comparison of previous experiments, the error function is designed at the end, and even the data that originally deviated far away can get a certain reference value.

4. Conclusions

The development of students' language ability is a gradual and spiraling increasing process. Among them, the autonomous learning process is a process in which students exert their subjective initiative to learn actively, and it is also a teaching process in which teachers play an important role. By analyzing the relevant theories of learning theory and learning environment, this paper establishes the abstract model of "role activity environment" of distributed learning system and establishes the problem framework of system modeling from the two dimensions of system elements and modeling level, that is, according to the matrix model of "role activity environment" and "theoretical basis analysis method design model." The reinforcement learning technology and agent technology are introduced into the research of adaptive system. On the basis of dynamic binding mechanism, the adaptive mechanism of adaptive system in uncertain environment-reinforcement learning-based adaptive mechanism is proposed, and the corresponding learning algorithm is proposed to support the learning process of adaptive agent. The reinforcement learning algorithm that builds the environment model through the shared experience strategy can reduce the training and speed up the learning process by constructing the environment model between the agents through the shared experience strategy. Finally, the experimental simulation in the grid environment proves that the algorithm is effective and convergent. Because the algorithm designed in this paper has a good effect on the control of error, the error reduction rate has reached 85.6%. However, this paper needs further modification. There are still some problems in the research. For example, we need to return to the goal of reinforcement learning and find better decisions. It is necessary to jump out of the decisions and data that have been tried before, so that it is possible to find a better decision. This needs further modification in future research.

Data Availability

The data used to support the findings of this study are available from the author upon request.

Conflicts of Interest

The author declares he has no conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] K. Manchella, A. K. Umrawal, and V. Aggarwal, "Flexpool: a distributed model-free deep reinforcement learning algorithm for joint passengers and goods transportation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 4, pp. 2035–2047, 2021.

[2] G. Yan, T. Kan, Y. Yang, and W. Zhang, "Distributed electric heating participation in demand response optimization scheduling based on deep reinforcement learning," *Power Grid Technology*, vol. 44, no. 11, 8 pages, 2020.

[3] B. Li and J. I. Wei, "Design of distributed intelligent intrusion prevention scheme based on reinforcement learning," *Computer Technology and Development*, vol. 29, no. 1, 6 pages, 2019.

[4] T. Liu, Y. Luo, and C. Yang, "Distributed interference coordination based on multi-agent deep reinforcement learning," *Journal of Communications*, vol. 41, no. 7, 11 pages, 2020.

[5] Y. Li, Z. Zhang, K. Meng, and H. Wei, "Optimal control strategy for energy storage in island microgrid based on hierarchical control," *Energy Storage Science and Technology*, vol. 11, no. 1, p. 176, 2022.

[6] M. Zhang, L. Wang, and Y. Feng, "Distributed cooperative spectrum sensing method based on reinforcement learning and consensus fusion," *Systems Engineering and Electronic Technology*, vol. 41, no. 3, 7 pages, 2019.

[7] W. Zhang, L. Ma, and X. Wang, "Research on event-driven multi-agent reinforcement learning," *Journal of Intelligent Systems*, vol. 12, no. 1, 6 pages, 2017.

[8] M. Mahbub, "Unmanned aerial vehicle-collaborative 5G: a cooperative technology for enhancement of 5G NR," *International Journal of Information Technology*, vol. 13, no. 2, pp. 793–799, 2021.

[9] A. Wu, R. Yang, X. Liang, J. Zhang, D. Qi, and N. Wang, "Visual range maneuver decision of unmanned combat aerial vehicle based on fuzzy reasoning," *International Journal of Fuzzy Systems*, vol. 24, no. 1, pp. 519–536, 2022.

[10] H. Wei, G. Zheng, V. Gayah, and Z. Li, "Recent advances in reinforcement learning for traffic signal control: A survey of models and evaluation," *ACM SIGKDD Explorations Newsletter*, vol. 22, no. 2, pp. 12–18, 2021.

[11] A. Sharif, J. P. Li, M. A. Saleem et al., "A dynamic clustering technique based on deep reinforcement learning for Internet of vehicles," *Journal of Intelligent Manufacturing*, vol. 32, no. 3, pp. 757–768, 2021.

[12] L. Lu, G.-W. Shen, G. Chun, Y.-H. Cui, C.-H. Jiang, and D.-Y. Wu, "A spark streaming parameter optimization method based on deep reinforcement learning," *Computer and Modernization*, no. 10, p. 49, 2021.

[13] H. Li, G. Li, and K. Wang, "Real-time scheduling strategy for electric vehicles based on deep reinforcement learning," *Automation of Electric Power Systems*, vol. 44, no. 22, 7 pages, 2020.

[14] H. Yang, W. Huang, and F. Ao, "Simulation of fish school self-organization behavior based on reinforcement learning," *Journal of National University of Defense Technology*, vol. 42, no. 1, 9 pages, 2020.

[15] Y. I. Liu and J. He, "The application of reinforcement learning in the control method of urban traffic lights," *Science and Technology Review*, vol. 37, no. 6, 7 pages, 2019.

[16] Z. Xu, L. Cao, Y. Zhang, X. Chen, and C. Li, "Research on deep reinforcement learning algorithm based on dynamic fusion target," *Computer Engineering and Applications*, vol. 55, no. 7, pp. 157–161, 2019.

[17] H. Li and P. Zhang, "Transient stability emergency control strategy of power system based on deep reinforcement learning," in *Proceedings of the 2022 5th International Conference on Energy, Electrical and Power Engineering (CEEPE)*, pp. 365–369, IEEE, 2022.

[18] P. Fan, S. Ke, S. Kamel et al., "A frequency and voltage coordinated control strategy of island microgrid including electric vehicles," *Electronics*, vol. 11, no. 1, p. 17, 2021.

[19] J. Zhong, T. Wang, and L. Cheng, "Collision-free path planning for welding manipulator via hybrid algorithm of deep reinforcement learning and inverse kinematics," *Complex & Intelligent Systems*, vol. 8, no. 3, pp. 1899–1912, 2022.

[20] Z. Zhao and J. Wang, "Simulation of ship automatic collision avoidance path based on reinforcement learning in various encounter states," *Science Technology and Engineering*, vol. 18, no. 18, 6 pages, 2018.

[21] L. Xu and Z. Zhao, "Channel and power allocation algorithm based on distributed collaborative Q-learning," *Computer Engineering*, vol. 45, no. 6, 6 pages, 2019.

[22] J. Liu, F. Gao, and X. Luo, "A review of deep reinforcement learning based on value function and policy gradient," *Journal of Computer Science*, vol. 42, no. 6, 33 pages, 2019.

[23] S. Manna, T. D. Loeffler, R. Batra et al., "Learning in continuous action space for developing high dimensional potential energy models," *Nature communications*, vol. 13, no. 1, pp. 1–10, 2022.

[24] Q. Jiang and B. I. Zeng, "Research on mobile robot navigation strategy based on deep reinforcement learning," *Computer Measurement & Control*, vol. 27, no. 8, 6 pages, 2019.

[25] K. Zhang and Y. U. Yang, "A review of teaching and learning methods based on inverse reinforcement learning," *Computer Research and Development*, vol. 56, no. 2, 8 pages, 2019.

[26] C. Lu, Y. Li, and Y. Feng, "Data-driven optimal stabilization control and simulation based on reinforcement learning," *Pattern Recognition and Artificial Intelligence*, vol. 32, no. 4, 8 pages, 2019.

[27] X. Zhang, L. Sun, Z. Kuang, and M. Tomizuka, "Learning variable impedance control via inverse reinforcement learning for force-related tasks," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2225–2232, 2021.