Research on Data-driven College English Teaching Model Based on Reinforcement Learning and Virtual Reality through Online Gaming

Jin Yan[1*]

[1]Basic Teaching and Research Department of Shaanxi Police Vocational College, Xi an 710061, China, jinyan2022609@163.com

Corresponding author: Jin Yan, jinyan2022609@163.com

Abstract. With the rapid development of corpus linguistics, the concept of DDL came into being. DDL (Data driven learning) makes full use of the network and corpus resources, changes the traditional teaching mode of conclusion based one-way indoctrination, creates a real language environment for learners, and reduces the phenomenon of interlanguage fossilization that is easy to occur in the process teaching method. It can effectively stimulate learners' interest in learning, cultivate their ability to learn independently and solve problems themselves, and achieve the ultimate goal of assisting English teaching. In view of this, this paper, based on the concept of DDL and taking the practice of English teaching reform as an example, designs a data-driven teaching model for college English, and discusses its specific implementation plan in the teaching process. The practical application of multi-agent deep RL (Reinforcement learning) algorithm is summarized and compared with other RL algorithms. The research shows that the algorithm in this paper improves the utilization rate of learning experience by 10.55%, thus greatly improving the learning performance. This study provides a new way for English teaching reform to improve students' ability to explore independently, and enriches the research on college English teaching models.

Keywords: Reinforcement learning; College English; Data-driven; Corpus; Teaching model; Virtual reality through Online Gaming
DOI: https://doi.org/10.14733/cadaps.2024.S5.197-210

1 INTRODUCTION

English teaching mode has experienced decades of development and change, and has formed dozens of teaching modes, including grammar translation method, listening and speaking method, cognitive method, communicative method, etc. [15]. But generally speaking, all these teaching models are no more than two categories, namely, the result method and the process method. The difference of teaching models reflects the difference of teaching theoretical thinking and the way of organizing teaching activities. Therefore, it has become a research trend of teaching reform to transform traditional teaching models and whether new teaching models can replace them [14]. The new teaching mode should be supported by modern information technology, especially the network technology, so that English teaching and learning can be developed towards personalized and autonomous learning without being limited by time and place to a certain extent [7]. Corpus DDL and teaching mode has changed the traditional knowledge acquisition and information processing mode in the teaching process, and the relationship between teachers and students, as well as between students and students, from teacher led to student led curriculum learning mode [1]. It was proposed by Tim Johns in 1991. It refers to "language material learning method" or "data-driven language learning method". It is an auxiliary means and tool for language learning. It is based on corpus, especially the method of learning language based on retrieval of language materials [20]. With some questions, language learners use retrieval software to discover rules and draw conclusions based on observation and analysis of a large number of real language materials, and master the use of certain language knowledge through real-time practice [4]. This is an advanced learning mode that directly applies corpus and its retrieval tools. Teachers organize students to analyze corpus data and encourage them to explore language phenomena and functions independently. DDL uses unlimited real examples to enable learners to transition from results to processes along the "continuum of teaching methods", thus building a bridge between results based teaching methods and process based teaching methods [8]. The advantage of corpus teaching is that it can change the way we learn text, and enable students to better find the scope and environment of language use. Get effective information to the maximum extent in limited time, help students quickly learn specific language points and their application methods, and help students form concepts and transform output [19].

Reinforcement Learning (RL) is an important branch of machine learning. Its essence is to describe and solve the problem that agents learn strategies to maximize returns or achieve specific goals in the process of interacting with the environment. Different from supervised learning, RL doesn't tell agents how to produce correct actions. It only evaluates the quality of actions and corrects action choices and strategies according to feedback signals. Therefore, the reward function of RL requires less information, is easier to design, and is suitable for solving complex decision-making problems [6]. RL is a kind of learning technology based on environmental feedback. RL agent identifies its own state, decides its actions according to a certain strategy, and adjusts the strategy according to the reward provided by the environment until it is optimal. Its main idea is to simulate the learning mechanism of human beings and higher animals, and to improve the original strategies through constant interaction with the environment and feedback from the environment [16]. The biggest feature of RL is that it can realize unsupervised learning without depending on specific models, making RL an effective method to solve sequential decision-making problems [5]. The process of RL takes a lot of computing time, and it slows down rapidly with the increase of the size of the state space and the action space of the learning system. The problem of slow convergence speed severely limits the application scope of RL in practical problems, so there is an urgent need for accelerated algorithms of RL. How to control the balance between "exploration" and "utilization" is of great significance to ensure that the algorithm can quickly converge to the optimal strategy [22]. The introduction of any other technology or method in the process of RL will inevitably affect the proportion of the two in the learning process. Therefore, it is more meaningful to design a reasonable balance control mechanism for specific RL algorithms than to study the problem only from the perspective of basic RL.The use of RL in online gaming also presents challenges. The computing time required for RL algorithms can be substantial, particularly as the size of the state and action spaces increases. This slow convergence speed can limit the practical application of RL in online gaming. Therefore, there is a need for accelerated RL algorithms to overcome this challenge and enhance the learning efficiency in complex gaming environments.

DDL embodies the pedagogical idea of "people-oriented". It first asks students questions, and then allows students to use high-speed retrieval methods to collect, analyze and process language materials, and summarize their own views on the problem, and finally solve the problem. DDL emphasizes that students should participate in the whole process of knowledge construction, and independently and actively carry out exploratory, discovery or verification learning. It encourages students to develop the habit of thinking, questioning and scientifically testing language materials, which is helpful to cultivate students' ability of induction, summary and reasoning. Different learners give the same or different answers to the same and different questions in different ways, which requires different learners to communicate with each other. Through the research on RL algorithm and data-driven teaching mode, it is found that RL has been successfully applied in many fields and has bright application prospects. Therefore, RL algorithm is still a problem worth studying and thinking, but due to the limited level of the author and the tight time, the content of the paper is only a stage work, and there are many shortcomings, which need further improvement and in-depth research. The innovation of this study lies in:

1. The improved fuzzy clustering method is integrated into hierarchical RL, and the large state space is automatically clustered to reduce the dimension of the state space. Based on the generated clustering subspace, subtasks are constructed, and then learning tasks are automatically layered.

2. The agent learning strategy is constantly updated according to the estimated worst outcome, which improves the robustness of the agent learning strategy and ensures the effectiveness of learning.

This article is divided into six sections from the organizational structure.

The first section is the introduction part, which leads to the language learners' learning mode of mastering certain language knowledge by using retrieval software to discover rules and draw conclusions on the basis of observing and analyzing a large number of real language materials with certain problems. The second section mainly summarizes the relevant literature, summarizes the advantages and disadvantages, and puts forward the research ideas of this paper. Section three introduces the DDL process and the data-driven model of college english in detail. The fourth section introduces RL algorithm. The fifth section is the experimental analysis and results. The sixth section is the conclusion, which summarizes the results of the full text.

2 RELATED WORK

At present, most domestic researches are based on pure text corpora for indexing and analysis to assist foreign language teaching. The combination of corpus technology and multimedia teaching methods in concept and application is relatively simple, and its advantages in foreign language teaching have not been fully reflected. Based on this situation, we integrate corpus technology and multimedia resources, and propose a DDL model based on multimedia corpus. This paper focuses on the practical and theoretical background of the model and its specific implementation.

Sun LY proposed a corpus DDL process, emphasizing the important role of corpus in language learning [18]. Shu J, Li W, Xu W elaborated on corpus data-driven foreign language learning in terms of ideas, methods and technologies [17]. Wu X defined the DDL model and discussed its specific application in teaching, such as context co occurrence, vocabulary collocation, polysemy, etc. [21]. Botvin M, Hershkovitz A, Forkosh Baruch A believe that the data-driven language teaching method allows students to query the corpus of spoken or written languages of native speakers with their own terminology indexing software, and provide them with real examples of language use, rather than some examples designed in grammar books [2]. Light, Wexler D H, Heinze J think that DDL is the process of realizing the proficient use of target language through "context co-occurrence" in classroom teaching. What he said about contextual co-occurrence is to find a large number of real corpora with search terms by searching corpora [11]. Braun's research points out that although the corpus of pure text corpus contains natural languages of various themes, it only retains the text in the transcription of multimedia corpora such as radio, television and movies, and discards the realistic scenes to which the text is attached [3]. Zhu W, Wei Z, Qin S put forward that "combining corpus index with text, audio and video organically, and establishing a multi-functional multimedia corpus can better meet the actual needs of college English learners" [25]. Lu P et al. put forward that starting from the core link of DDL language learning, they described the process of DDL data driven language learning. This method avoids as many links of knowledge as possible from materials to teachers, and from teachers to students, so that learners can directly participate in it and establish their own language meaning and use files [12]. Xiao RZ also proposed how to use one or more media to construct a learning mode that conforms to the new characteristics of knowledge dissemination, and how to transmit the information reflecting the knowledge content to the audience, so that they can achieve the most effective knowledge dissemination, which has become the mainstream of the current teaching reform [23]. On the introduction of RL algorithm, Jezek E proposed a two-layer control structure, in which the bottom layer directly controls the bottom layer behavior, while the top layer monitors the bottom layer behavior and intervenes to replace a bottom layer control strategy when the bottom layer enters the boundary state. This hierarchical structure embodies the idea of "divide and rule" and simplifies the complexity of the problem [10]. Higuchi T et al. used RL method to simulate the control problem of nonlinear system [9]. Manishina E et al. studied the learning method of intelligent robot collision avoidance behavior based on RL [13].

3 DATA-DRIVEN TEACHING MODEL

3.1 Data-Driven Learning Process

Constructivism holds that learning is not the transmission of knowledge from teachers to students, but the process of students constructing their own knowledge. This means that learning is active, and students are not passive receivers of information, but should actively choose and process external information to build their own understanding. Corpus based DDL model and constructivism theory are the perfect combination of theory and practice. DDL takes real language as the main language input. The language data provided by the corpus to students are all from real communication activities. Language materials belong to natural language. Corpus based DDL can activate students' knowledge about real world discourse and language system, and then construct an in-depth learning environment. Real language can promote language acquisition. The process is roughly as follows (Figure 1): First, students ask questions and make assumptions about them. Secondly, the real corpus of the corpus is searched through the corpus retrieval tools for the questions raised, so as to answer their own questions, or confirm or overturn the assumptions raised. Finally, find the feedback of learning results and summarize.

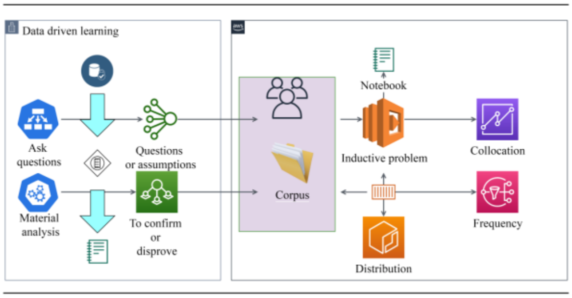

Figure 1: Flowchart of data-driven language learning.

It can be seen that data-driven college English learning advocates bottom-up and inductive learning. In this model of English learning, students first come into contact with a large number of real language data, rather than prescriptive grammar rules. After observation, they can summarize grammar rules. Emphasize the learning process of exploration and discovery, guide students to monitor the learning process themselves, and explore and discover language knowledge according to their own needs.

3.2 Data-Driven Model of College English

The application of corpus in language teaching is not only to provide rich language materials for teaching activities, but also to directly apply corpus research tools, such as retrieval tools, to learners' learning activities. This has led to significant changes in language learners' learning methods, learning content, and the form of learning materials in teaching activities. These activities provide solutions to various problems in foreign language teaching, such as autonomy in language learning, the effectiveness of language material input, and the balance between fluency and accuracy in language learning. Therefore, the introduction of corpus linguistics in the teaching of such courses can just make up for the shortcomings of the traditional classroom teaching model.

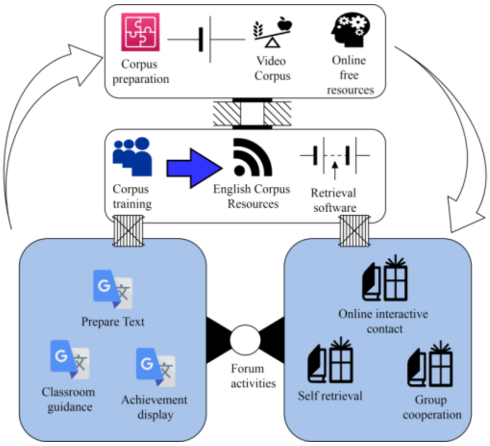

Figure 2: Data driven teaching mode of college English.

During English listening, speaking and reading training, corpus gives foreign language learners the opportunity to get in touch with a large number of real language environments. The application of DDL mode in foreign language teaching depends on the selection of language materials and indexing tools, in addition to learners' subjective ability to ask questions, summarize, summarize and analyze language materials. Therefore, the core of DDL mode based on multimedia corpus lies in the construction of multimedia corpus and the development of learning platform based on it (Figure 2). After students ask questions, they need to collect language materials and classify them. In traditional teaching, this is usually achieved by consulting dictionaries, referring to books, or making up example sentences by teachers through introspection. Such materials are extremely limited, and some of them are not real language materials.

4 INTRODUCTION TO REINFORCEMENT LEARNING ALGORITHMS

The RL can obtain the optimal decision through the continuous interactive feedback between the agent and the environment, and constantly updating the strategy with the feedback information.

During the training process, the agent obtains the evaluation of environmental feedback through continuous trial and error, so as to accumulate experience, update strategies, and maximize the cumulative rewards and punishments. It does not need correct guidance information, but only needs to obtain training samples through trial and error to finally achieve the optimal behavior strategy. The RL algorithm model is shown in Figure 3 below.

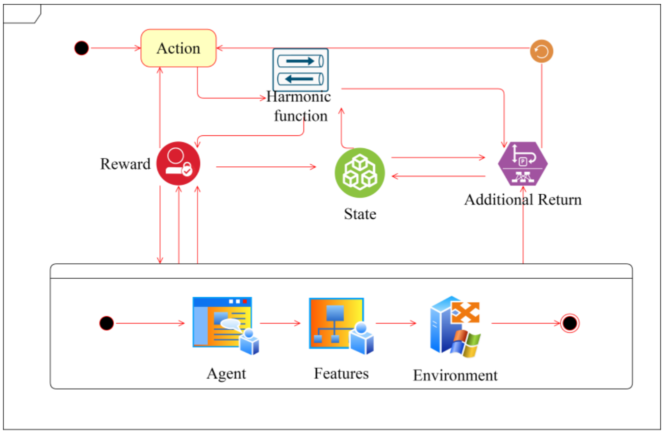

Figure 3: RL algorithm model.

An RL system has not only agents and environments, but also four basic elements; Strategy, value function, return function and environment model. RL is mainly a way of continuous interactive learning with the environment, which is different from planning. Because the general RL algorithm obtains the optimal strategy based on the value function estimation of the state action pair, the action value function can approximate the optimal action value function through continuous iterative learning. RL also has a very important feature. Under certain conditions, as long as every state can be accessed, the algorithm will eventually converge. This leads to the RL algorithm to spend a lot of time on exploration to ensure the convergence of the algorithm. Then the environment model and planning are introduced into the RL system to form a new development, which makes the RL and dynamic planning methods closely linked. Compared with the traditional RL system, the system includes a post heuristic strategy learning module. The strategy learning module is a combination of action function selection strategies to guide action selection, and does not directly affect the value function, so it will not affect the convergence of the system.

Assuming that $p(t)$ is implemented by a neural network, for each observation w, a variation Δw_t is determined. After a complete sequence processing, w is modified as follows:

$$w \to W + \sum_{t=1}^{m} \Delta w_t \tag{1}$$

Define the objective function:

$$E_t = \frac{1}{2}\left[k - p(t)\right]^2 \tag{2}$$

The RL algorithm is to continuously interact with the environment, select an action a from the current state S to the next state s', and obtain a reward value of the action a. The goal of each agent is to maximize its total return.

Once the agent finds the sub target, it can create the option set online. Our goal is to automatically find the option in the process of the agent's continuous trial and error learning, so as to speed up the learning process. There are many successful paths in the learning process. Using the aforementioned properties, we can analyze these paths to find appropriate sub goals, and then generate options:

$$Q_{\pi} = R + \delta \sum_{i=1}^{n} Q \tag{3}$$

The agent selects behavior a in the current state s, and determines the behavior at the next moment through the reward and punishment information fed back to the agent by the environment and the state it has reached.

$$E_{m} = \frac{1}{2}\left[r(t) + \beta(t+1) - p(t)^{2} \right] \tag{4}$$

Calculate the average n_i after the next learning cycle is completed, and find the sum m of the absolute value of the difference between n_i and all the values after the update:

$$m = \sum_{i=1}^{n} \left\| n_i - x_i \right\|^2 \tag{5}$$

Through the analysis of the fusion algorithm, it can be found that when the value of n shows an increasing trend, it can be considered that the average value is in an increasing state, and the action of state E can bring more gains. Therefore, other agents will learn from the agent with the largest n value. On the contrary, all learn from the agent with the lowest value.

$$u^{\Re} = -(\sqrt{2}\beta) + k \tag{6}$$

The fitness function reflects the individual's ability to adapt to the environment. By calculating the value of the fitness function, the survival chance of the individual can be effectively controlled. For different problems, the definition of fitness function is different. The fitness function we give here is:

$$f_t = \frac{1}{F_t + 1} \tag{7}$$

Where, F_t is the objective function value in the fuzzy clustering, and the smaller the value of F_t, the greater the individual's fitness value, that is, the stronger the individual's adaptability.

$$E[\sigma^2(s)] = \sum_\alpha \pi(\alpha,\theta) \cdot f(s) \tag{8}$$

In the dynamic interaction process between agent and environment, F_t-value function can be used to dynamically change the selection of action value function, so as to achieve the balance between exploration and utilization. The Boltomnm distribution with the introduction of F_t value is shown in the following formula:

$$p(a_t) = \frac{e^{Q(s,a)/t}}{\sum_{a_k} e^{Q(s,a_k)/t}} \tag{9}$$

When RL algorithm is applied to concrete practical applications, it usually needs to deal with two problems: how to choose behavior selection strategies and how to deal with the storage and generalization of behavior value functions. The Q function of the learning agent is linked with the strategic characteristics of the other agent, and the influence of the non-stationary environment on the learning process of the agent is reduced.

$$\Delta w_a = \alpha \frac{\partial\ p(t)}{\partial w_\alpha} \cdot \frac{\partial A}{\partial W_\alpha} \tag{10}$$

Where α is the learning rate.

The output of the action network does not directly affect the environment, but is regarded as the expected value of the action. The actual action can be searched and selected within a certain range around the value, and the search range $\xi(t)$ is equivalent to a probability function variable:

$$\xi(t) = \frac{K}{1 + e^{p(t)}} \tag{11}$$

5 EXPERIMENTAL ANALYSIS AND RESULTS

To understand whether the DDL model can guide students to use corpora and relevant retrieval tools for autonomous English learning, and how effective the learning is. This paper conducts a teaching experiment on this topic. The subjects of the study are 118 students in a foreign language department of a university. Through independent sample t test on their English proficiency test scores, the two classes' English proficiency is basically the same. First, one class was designated as the experimental group with 60 people, and the other class was designated as the control group with 58 people. The teachers in the experimental group and the control group are the same person, using the same teaching materials, and there are 4 class hours of classroom teaching and 2 class hours of extracurricular autonomous learning every week. The students in the experimental group were trained to understand the concept of data-driven, and English corpora suitable for college students were recommended, such as English national corpora, Chinese learners' English corpora, etc. The test results are shown in Table 1.

Group	*N*		*Mean*	*Deviation*	*T*	*P*
Experience group	*60*		*14.8*	*0.69*	*2.369*	*0.058*
Control group	*58*		*17.38*	*0.53*		
Experience group	*60*		*15.33*	*0.97*	*5.236*	*0.047*
Control group	*58*		*15.38*	*0.55*		

Table 1: The test results of the grammar learning part of the experimental group and the control group.

Looking at the above experimental results, it can be seen that compared with the traditional learning model, DDL model can significantly improve learners' English learning level. At the same time, it also helps to cultivate learners' ability to summarize from the corpus and autonomous learning ability, indicating that the implementation of DDL model is successful.

In order to verify the effectiveness of the algorithm proposed in this paper, we first set the simulation parameters as shown in Table 2. Among them, the learning rate and exploration probability decrease linearly to 0 after 1000 fitness evaluation experiments.

Learning system parameters	*RL parameters*	*Evolution process parameters*
N0	α	*0.8*
N1	γ	*0.06*
T1	δ	*0.1*
T2	Ω	*0.1*
E=100		$\Delta=1$

Table 2: Parameter settings.

In general, multi-agent RL algorithm uses good training data generated in advance for offline training. In order to meet the needs of online training of RL, it is also necessary to modify the algorithm into an online running algorithm. Realize the adaptive division of the continuous state space, obtain the discrete state representation of the continuous state problem, and then use the standard RL method for training to obtain the solution of the problem.

It can be seen from the convergence speed comparison in Figure 4 that the algorithm proposed in this paper can achieve good results no matter whether the learning environment of the agent is regular or not. In the experiment, we should pay attention to the selection of the population number. If it is too small, the performance of the algorithm will become very poor. If it is too large, it will increase the amount of computation and make the convergence time grow.

Figure 5 shows the comparison of the learning process between the method in this paper and the CMA based RL method in literature [11]. The vertical axis in the figure is the scheduled completion time of the learning task, the horizontal axis is the number of learning times, and the data in the figure is the average value of five groups of experimental results. Because the CMA based method has designed the partition of the continuous state space in advance, only a simple RL process is required. It can be seen that the method proposed in this paper has fewer partitions, but the effect is better. It realizes the adaptive partition of continuous state space, and its learning rate and learning performance are improved.

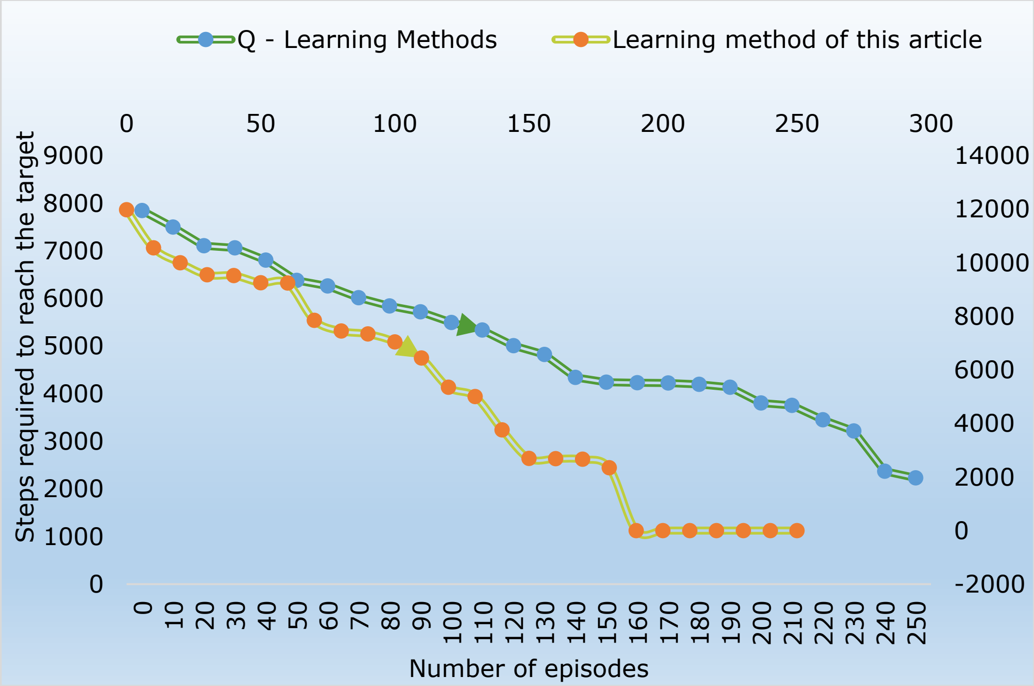

Figure 4: Comparison of the convergence speed between the algorithm in this paper and the Q-learning algorithm.

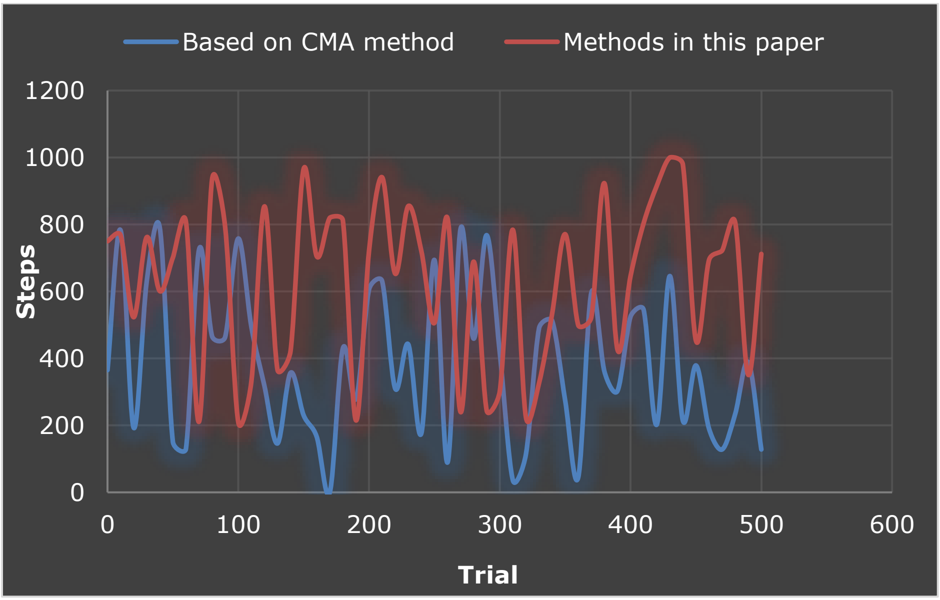

Figure 5: Comparison between the method in this paper and the RL method based on CMA.

In order to overcome the inherent randomness of RL algorithm, each algorithm in the comparison experiment is run 16 times respectively, and the result data is the average of the test results. Finally, the above algorithms of various types and parameters can converge successfully within 1500 Episodes. The following Figure 6 shows the comparison results of the two algorithms. The learning curve describes the average change of the average cumulative reward value obtained from the sub fitness evaluation experiment over time when 2600 consecutive times of RL process are in one during the operation of the sub algorithm.

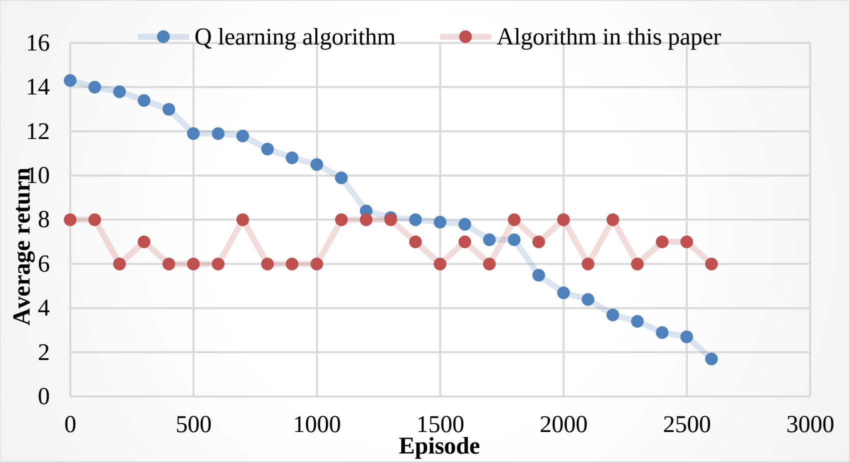

Figure 6: Learning curves of the two algorithms.

The dynamic change of the environment affects the learning algorithm differently in different states. To make full use of the hierarchical characteristics to treat the environmental changes differently, the RL method can also produce satisfactory learning results in the dynamic environment.

In order to eliminate the randomness in the results of one run, each algorithm is run independently, and the average learning performance statistics results are shown in Figure 7. Therefore, in extreme cases, you can find a solution that conforms to the given preference, and the more important the target is, the faster the learning speed is, which verifies the ability to determine the correct learning direction according to the preference information.

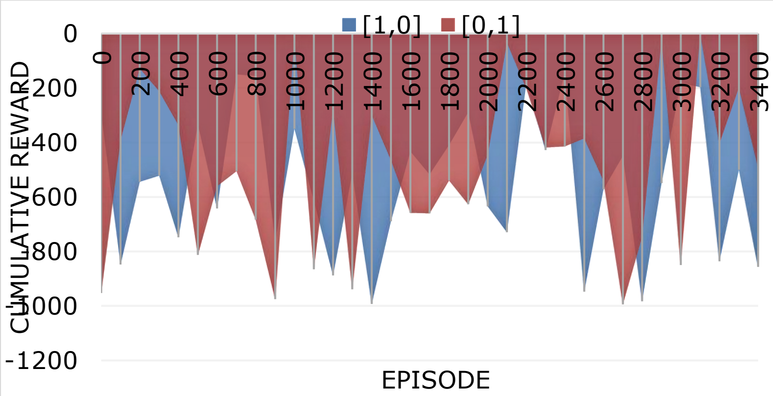

Figure 7: Learning performance on two objectives.

Figure 8 shows the learning curve on the two objectives under extreme conditions, showing the variation characteristics of the average results of the last 2000 times with times. It can be seen that when the importance vector is [0, 1], the learning curve on the target quickly converges to 0. When the importance vector is [1, 0], the learning curve on the target with fast convergence speed can also converge to 0, indicating that the strategy of final convergence is to keep still, and when the importance vector is [1, 0], the convergence speed is the fastest.

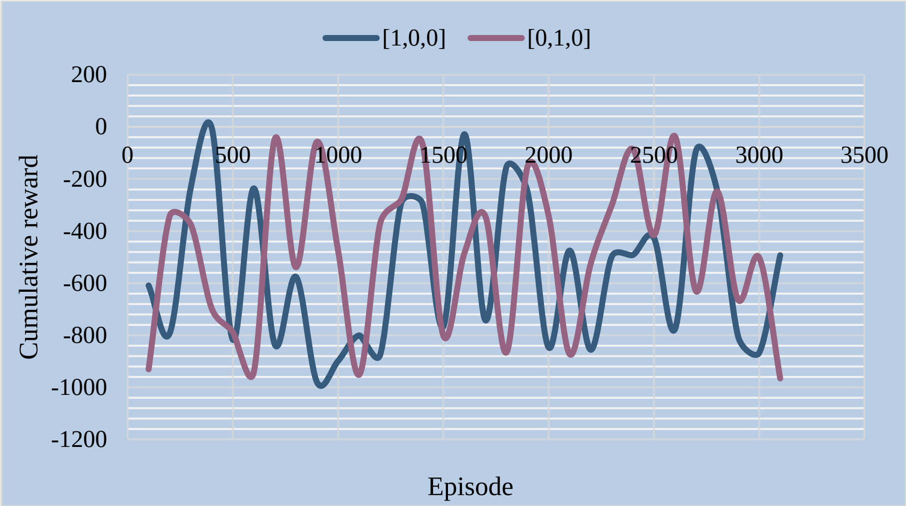

Figure 8: Learning curve on two goals.

In a word, under the data driven teaching mode proposed in this paper, teachers first complete the corpus preparation and corpus training. Then, we will combine in class micro text guidance, extracurricular online interactive exercises, group extracurricular discovery learning and in class achievements display in a step-by-step manner. You can also use the forum function of the teaching website to realize extracurricular online interaction between students and teachers and students, so as to organically combine guided teaching and extracurricular discovery learning. At the same time, the application of multi-agent deep RL algorithm in data-driven teaching mode is reviewed, and compared with other RL algorithms. The research shows that the algorithm in this paper improves the utilization rate of learning experience by 10.55%, thus greatly improving the learning performance.

6 CONCLUSIONS

The progress of the times will inevitably bring about the renewal of the teaching mode. DDL makes full use of the network and corpus resources, and changes the traditional teaching mode of one-way indoctrination of conclusion. It creates a real language environment for learners, reduces the interlanguage fossilization that easily occurs in the process teaching method, can effectively stimulate learners' interest in learning, cultivate their ability to learn independently and solve problems by themselves, and achieve the ultimate goal of assisting English teaching. Compared with the traditional teaching mode, the multimedia teaching based on RL integrates sound, image and text, and the continuous stereo information not only strengthens the stimulation of input information, but also increases the input capacity. At the same time, it also stimulates students' interest and autonomy, and activates the internal drive of students' autonomous learning of English.

It enables students to constantly take questions to the corresponding corpus to find answers. This learning mode can help students find the language rules that they do not understand or ignore. Students can use the corpus to have a deeper and more accurate understanding of the use of vocabulary in a large number of objective and real corpora. The DDL model can help learners improve their ability to summarize and summarize the language materials and enhance their autonomous learning ability. However, learners are required to have a good ability to summarize, summarize and analyze the language materials. Therefore, it is more suitable for intermediate and advanced language learners and will achieve better results. The effect of DDL model has been recognized by many experts. With the popularization of corpus and computer technology in foreign language teaching, this learning method should have broad prospects for development.

Jin Yan, https://orcid.org/0009-0000-2657-3462

REFERENCES

[1] Batory, D.; Azanza, M.: Teaching Model-Driven Engineering from a Relational Database Perspective, Software & Systems Modeling, 17(1), 2018, 118-140.
[2] Botvin, M.; Hershkovitz, A.; Forkosh-Baruch, A.: Data-Driven Decision-Making in Emergency Remote Teaching, Education and Information Technologies, 1648(4), 2020, 042029 (5pp).
[3] Braun, S.: Integrating Corpus Work into Secondary Education: From Data-Driven Learning to Needs-Driven Corpora, Cambridge University Press, 30(7), 2018, 718-724.
[4] Data-Driven School Improvement : Linking Data and Learning, Teachers College Press, 2008, 320.
[5] Dodman, S. L.; Swalwell, K.; Demulder, E.: Critical CMA data-driven decision making: A Conceptual Model of Data Use For Equity, Teaching and Teacher Education, 23(2), 2019, 89.
[6] Efopoulos, V.; Evangelidis, G.; Dagdilelis, V.: WIPE - A Model for a Web-Based Database-Driven Environment for Teaching Programming, Advanced Learning Technologies, 26(7), 2017, 3.
[7] Elliott, A.: Building Relationships to Strengthen Learning, Every Child, 21(2), 2018, 399.
[8] Hales.; Tom.: Exploring Data-Driven Language Awareness, ELT Journal: English Language Teachers Journal, 38(4), 2019, 34.
[9] Higuchi, T.; Kinoshita, K.; Delcroix, M.: Adversarial Training For Data-Driven Speech Enhancement Without Parallel Corpus, Automatic Speech Recognition And Understanding Workshop, 2(10), 2017, 68. https://doi.org/10.1109/ASRU.2017.8268914
[10] Jezek, E.; Lenci, A.: When Gl Meets the corpus: a Data-Driven Investigation of Semantic Types and Coercion Phenomena, Workshop On Generative Approaches to the Lexicon, 39(7), 2018, 345.
[11] Light, D.; Wexler, D. H.; Heinze, J.: Keeping teachers in the center: a Framework for Data-Driven Decision-Making, Education Development Center, 55(16), 2017, 463-469.
[12] Lu, P.; Huenerfauth, M.: Collecting a Motion-Capture Corpus of American Sign Language for Data-Driven Generation Research, Association for Computational Linguistics, 209(66), 2018, 292-320.
[13] Manishina, E.; Jabaian, B.; Stéphane, Hu.: Automatic Corpus Extension for Data-Driven Natural Language Generation, Language Resources and Evaluation, 20(5), 2016, 66.
[14] Mao, L.; Liu, Y.; Zhang, M.: Research on the Effectiveness of College Student English Writing Teaching Based on Data-Driven Learning, Educational Sciences: Theory and Practice, 18(5), 2018, 16.
[15] Qing, S.: A Corpus-Based Research of Data-Driven English Vocabulary Teaching and Learning, Overseas English, 3(15), 2020, 2.
[16] Rizzo, C. R.; MJM, Pérez.: Using Data-Driven Learning Methods in Telecommunication English Teaching, Computer-Assisted Foreign Language Education, 26(7), 2017, 49.

[17] Shu, J.; Li, W.; Xu, W.: Research on Data-driven Feedback Teaching Service, Advanced Technologies Enhancing Education, 29(5), 2020, 9.
[18] Sun, L. Y.: The Design and Practice of Data-Driven Teaching Evaluation Model in Colleges and Universities, Contemporary Foreign Languages Studies, 8(11), 2017, 381.
[19] Torii, T.: An Examination of a System of Data-Driven Teaching and Learning Improvement: the Albany Outcomes Assessment Model at the State University of New York at Albany in the United States, Nagoya Journal of Higher Education, 69(2), 2017, 611-625.
[20] Wang, H.: The Construction and Practice Research of Data-Driven Precision Teaching Model, Educational Technology, 7(5), 2019, 409.
[21] Wu, X.: Data-Driven Teaching–Learning-Based Optimization Framework for Expensive Engineering Problems, Structural and Multidisciplinary Optimization, 33(4), 2020, 042. https://doi.org/10.1007/s00158-021-03010-1
[22] Wu, X.: Research on the Reform of Ideological and Political Teaching Evaluation Method of College English Course Based on Online and Offline Teaching, Journal of Higher Education Research, (1), 2022, 387-90. https://doi.org/10.32629/jher.v3i1.641
[23] Xiao, R. Z.: Theory-Driven Corpus Research : Using Corpora to inform Aspect Theory, Corpus and Linguistic Theory, 29(77), 2018, 409.
[24] Xu, M.; Chen, X.; Liu, X.: Using Corpus-Aided Data-Driven Learning to Improve Chinese EFL Learners' Analytical Reading Ability, Computer and Communications Societies, 20(1), 2019, 6. https://doi.org/10.1007/978-981-13-9895-7_2
[25] Zhu, W.; Wei, Z.; Qin, S.: Corpus building for data-driven TTS systems, Speech Synthesis, 174(3), 2021, 032.