Research on Adaptive Learning Path Optimization of Online Courses for Business Administration Major in Higher Vocational Education Based on Deep Reinforcement Learning

Zhenqian Li
Business College, Jiangxi Hongzhou Vocational College, Fengcheng 331100, Jiangxi, China
Corresponding author: Zhenqian Li (e-mail: lizhenqian2026@126.com)

ABSTRACT Online courses in higher vocational business administration require learning-path optimization that can adapt to heterogeneous knowledge foundations, learning behaviors, and vocational competency objectives. To overcome the static and one-size-fits-all limitations of conventional online teaching paths, this study proposes an adaptive learningpath optimization model based on deep reinforcement learning. Course knowledge points, prerequisite relations, jobcompetency requirements, student learning behaviors, and personalized profiles are collected and transformed into a structured learning data sample library. A DQN-based decision model is then constructed, where the state space includes knowledge mastery, learning progress, learning fatigue, and career fit, and the action space includes knowledge-point sequencing, resource-type selection, and learning-time allocation. A multi-dimensional reward function integrates knowledge improvement, learning efficiency, and vocational adaptability. Experiments using the online courses "Fundamentals of Management" and "Marketing" over one semester show that the experimental group improves course completion rate by 18.3%, knowledge-point mastery by 21.5%, and learning satisfaction by 15.7% compared with the fixed-path control group. The model provides a data-driven decision-optimization method for adaptive learning systems and intelligent educational resource scheduling.

INDEX TERMS deep reinforcement learning, adaptive learning path, learning analytics, decision optimization, online education

I. INTRODUCTION

In the process of deepening the digitalization of vocational education, online courses are the key support for flexible teaching and improving the quality of talent cultivation in vocational business management. The teaching effectiveness of online courses directly affects the cultivation of vocational literacy and work ability of vocational college students. The core of the Business Administration major in higher vocational education is to cultivate applied and skilled talents. Its students have the characteristics of large differences in knowledge foundation, uneven learning initiative, and strong practicality. However, most online courses still follow a unified learning path design. They do not consider the personalized needs of students and their career development direction. Therefore, students often face a disconnect between theoretical acquisition and practical application, where they struggle to translate online course content into usable professional skills. The teaching philosophy of "teaching according to students' aptitude" is difficult to implement.

The main innovations of this article are as follows: firstly, it can adapt well to the environment of higher voca tional education, break through the static decision-making defects of traditional adaptive learning paths, focus more on practice than theory, and integrate the work ability requirements of higher vocational business man agement majors into the optimization of learning paths, achieving precise integration of "learning and work"; Secondly, building targeted deep reinforcement learning models, abandoning the static decision-making flaws of traditional algorithms, and dynamically adjusting learning paths through the dynamic interaction between agents and learning environments to achieve dynamic optimization of personalized and adaptive learning; Thirdly, based on the learning characteristics of vocational college students, design reward functions that consider multiple dimensions such as knowledge mastery, learning efficiency, and career needs to enhance the practicality and adaptability of the model, which is different from the single dimensional reward function design of existing models.

This article presents the finalized findings of a longitudinal study conducted between 2024 and 2025. Based on the core online courses of Business Administration major in higher vocational education, sort out the knowl edge point system and job competency requirements, collect student learning behavior data (such as learning time, knowledge point test scores, learning progress, etc.) and personalized feature data (such as knowledge foundation, learning preferences, career planning, etc.), and establish a learning data sample library; Secondly, an adaptive learning path optimization model based on deep reinforcement learning was constructed. Taking students' real-time learning status (knowledge mastery, learning progress, learning fatigue) as the state space, allocating knowledge learning sequence, learning resource types, and learning duration as the action space, and using knowledge mastery improvement, learning efficiency, and occupational adaptability as the main indicators of the reward function. Train the model using DQN (Deep Q-Network) algorithm to achieve dynamic decision-making of learning paths; Once again, design a model validation plan and select the two core online courses of "Management Fundamentals" and "Marketing" for the Business Administration major in vocational colleges. Establish an experimental group (using optimized learning paths to select parallel classes) and a control group (using traditional fixed learning paths to select parallel classes) for one semester of experimental teaching; Finally, collect experimental data to validate the learning effectiveness, efficiency, and satisfaction of the model from three dimensions, analyze the advantages and disadvantages of the model, and propose optimization suggestions.

The main contribution of this article is to establish a theoretical research framework for deep reinforcement learning and online education in vocational enterprise management, improve the theoretical system of adap tive learning path optimization, and provide theoretical basis for online course reform in similar vocational fields; From a technical perspective, design a deep reinforcement learning optimization model suitable for vocational college business, solve the problem of weak dynamic adaptability of traditional methods, and improve the personalized accuracy of learning paths; On a practical level, the effectiveness of the model has been verified through experiments, and a feasible online course adaptive learning path optimization solution based on the model has been proposed, which can effectively improve students' learning effectiveness and experience, and provide practical reference for the reform of online education and teaching in vocational col leges. This article will follow the logic of "constructing experimental verification conclusions based on research status and methods", and explain in detail the relevant research results, model construction, experimental results, and research conclusions to make the research content complete and logically clear.

II. RELATED WORK

In the field of deep reinforcement learning, numerous scholars have explored different scenarios, providing theoretical and technical references for the optimization of adaptive learning paths in online courses for higher vocational business management majors. For instance, recent studies have leveraged knowledge tracing (KT) to model student mastery levels, while others have employed learning path recommendation (LPR) systems to sequence content based on learner modeling. While foundational works provide the computational basis for Deep Reinforcement Learning (DRL), this study specifically integrates educational technology literature on cognitive maps and prerequisite discovery to tailor paths for vocational students.Samieyeganeh et al. [1] explored the development from deep reinforcement learning to multi-agent deep reinforcement learning; Nurmuhammet [2] studied the application of deep reinforcement learning on stock data; Fan [3] has explored reinforcement learning and deep reinforcement learning; Jaeger and Geiger [4] conducted an introductory study on deep reinforcement learning; Karimzadeh [5] focuses on deep reinforcement learning multi-agent systems; Wang [6] studied the application of deep reinforcement learning in stock prediction; Pradhan [7] evaluated deep reinforcement learning algorithms; Liu and Chen [8] introduced the research progress of deep reinforcement learning; Skuba et al. [9] applied deep reinforcement learning to traffic signal control; Guo and Liu [10] conducted research on unmanned aerial vehicle path planning based on deep reinforcement learning. However, existing research has mostly focused on general algorithm optimization or single scenario applications, lacking in-depth customization for the characteristics of online courses in higher vocational business management majors (such as strong career orientation and significant differences in learner abilities), and not fully integrating educational big data to construct learner images and career ability models, resulting in insufficient dynamic adaptability and limited personalized matching accuracy. This provides direction for future research.

III. METHOD

A. PRELIMINARY RESEARCH AND DATA COLLECTION

Preliminary research and data collection are the foundation of model construction (as shown in Figure 1), which mainly consists of three steps: first, course and job research. Three vocational colleges' business management majors are selected, and the knowledge point system of core online courses such as "Management Fundamentals", "Marketing", and "E-commerce" is sorted out. The difficulty level, knowledge point prerequisite relationships, and corresponding job ability requirements of each knowledge point are clarified. Specifically, these prerequisite relationships were determined through a hybrid approach: initially, a panel of three subject matter experts with over ten years of teaching experience in Business Administration manually annotated the logical dependencies between knowledge points. To ensure rigor, inter-rater reliability was assessed using Cohen's Kappa, achieving a high degree of consensus (k = 0.86). These manual annotations were subsequently validated against historical student data using the Association Rule Mining (ARM) algorithm to ensure that the defined prerequisite paths align with empirical learning sequences. Based on this, a knowledge point-job ability correspondence table is constructed; The second is student research, which collects personalized characteristic data of 200 vocational college students majoring in business administration through questionnaires, interviews, and other methods. This cohort provided the historical baseline data used to initialize and pre-train the model. Subsequently, a specific subset of 95 students (48 in the experimental group and 47 in the control group) was selected from this original pool for the semester-long validation experiment to ensure that the participants' profiles in the live trial remained consistent with the pre-training data distribution; The third is data collection, relying on online course platforms to collect students' learning behavior data, including learning duration, knowledge point test scores, distribution of wrong questions, number of learning interruptions, and number of clicks on learning resources. Effective data is screened and preprocessed (denoising, normalization, missing value filling), and a learning data sample library is established to provide data support for model training. At the same time, referring to the relevant requirements for deep reinforcement learning training, the historical data is divided into 70% for offline pre-training and 30% for offline testing. Crucially, the experimental results reported in Table 3 (and Figure 3) reflect the performance of the pre-trained model evaluated on the held-out 30% testing set prior to the intervention. The "online finetuning" described in Section 3.3 refers to the subsequent live deployment phase, where the model adapted to real-time student feedback; however, to ensure reproducibility and prevent data leakage, the comparative metrics are based on the frozen weights established during the validated 70/30 offline split [11].

B. CONSTRUCTION OF ADAPTIVE LEARNING PATH OPTIMIZATION MODEL BASED ON DEEP REINFORCEMENT LEARNING

This article uses the DQN (Deep Q-Network) algorithm to construct an adaptive learning path optimization model. Based on the course characteristics and student learning features of the Business Administration major in higher vocational education, the model's state space, action space, and reward function are designed in a targeted manner to achieve dynamic optimization of the learning path. Firstly, the state space design focuses on the real-time learning status of students, selecting four key indicators: mastery of knowledge points (quan tified by test scores), learning progress (the proportion of completed knowledge points), learning fatigue, defined here as the cognitive decline in engagement during complex Business Administration case analysis, which is quantified by the frequency of learning interruptions and significant fluctuations in session duration relative to the average difficulty of the management module, and career fit (the degree of matching between current learning knowledge points and students' career planning). After normalizing the four indicators, the model's state vector is constructed to ensure the comprehensiveness and accuracy of the state description; Secondly, the design of action space, combined with the characteristics of online courses in higher vocational business management, includes three dimensions: adjusting the order of knowledge point learning (such as skipping mastered knowledge points and prioritizing weak knowledge points), selecting learning resource types (such as video lectures, case analysis, and exercise exercises), and allocating learning time (such as increasing learning time for difficult knowledge points). Multiple optional actions are set for each dimension to form a complete action space; Finally, the reward function design adopts a multidimensional comprehensive reward function that takes into account knowledge mastery, learning efficiency, and career adaptation. The specific formula is: $R = \alpha R_1 + \beta R_2 + \gamma R_3$, where R_1 is the knowledge mastery reward (calculated based on the improvement of test scores), R_2 is the learning efficiency reward (calculated based on the ratio of learning time to knowledge mastery), R_3 is the career adaptation reward (calculated based on the matching degree between current learning knowledge points and career planning), α, β, and γ are weight coefficients, and their values are determined to be 0.4, 0.3, and 0.3 through experimental debugging to ensure the rationality and pertinence of the reward function [12].

C. MODEL TRAINING AND OPTIMIZATION

The model training adopts a combination of offline training and online fine-tuning (as shown in Figure 2). The specific process is as follows: firstly, initialize the model parameters, use the Xavier initialization method to initialize the weight parameters of the DQN network, avoid gradient vanishing or explosion problems, set the learning rate to 0.001, discount factor to 0.9, experience replay buffer size to 10000, and batch size to 32; Secondly, offline training utilizes the previously collected learning data sample library, with students' learning status as input and actions as output. The model is trained through an experience replay mechanism to learn the mapping relationship between states and actions. When the loss function of the model converges (loss value less than 0.01), offline training is completed; Once again, online fine-tuning will be carried out by deploying the trained model to the online course platform of the Business Administration major in higher vocational education. Real time data on students' learning status will be collected, and model parameters will be dynamically adjusted based on their learning feedback (such as test scores and changes in learning behav ior) to optimize action decisions and enhance the dynamic adaptability of the model; Finally, the model is validated and optimized using test set data to analyze the accuracy of path recommendation and the

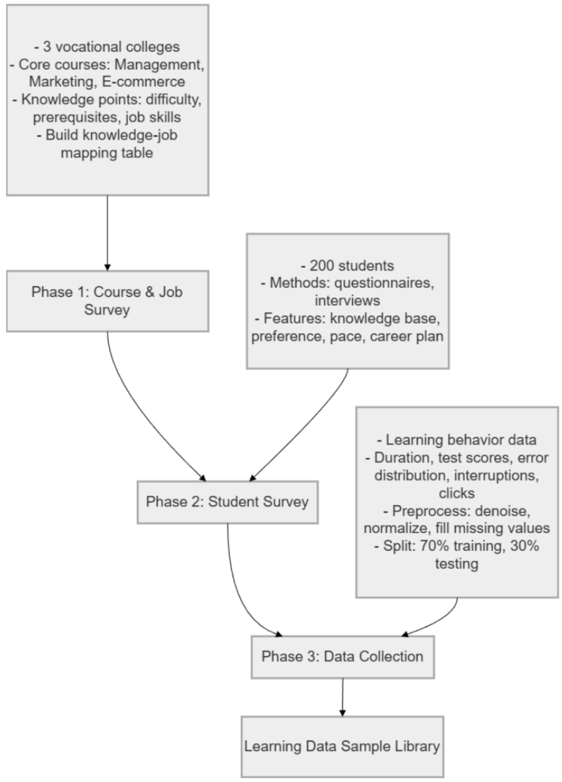

FIGURE 1. Preliminary research and data collection

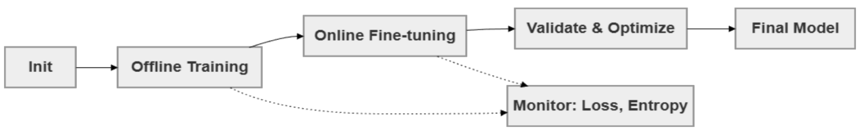

FIGURE 2. Model Training and Optimization

extent of improvement in learning effectiveness. The model was iteratively optimized to address specific performance bottlenecks, such as delays in real-time path updates and the initial imbalance of reward function weights, ensuring its practicality and stability. At the same time, during the training process, the value of the visual loss function and the entropy of the random strategy are visualized to avoid premature decline in strategy entropy leading to insufficient exploration. Multiple random seed training is used to take the average value, reducing the errors caused by model randomness [13].

IV. RESULTS AND DISCUSSION

A. EXPERIMENTAL DESIGN AND DATA COLLECTION

The study demonstrates that the adaptive learning path optimization model based on deep reinforcement learning has completed a full-scale comparative experiment. Moreover, the concluded findings suggest that two parallel classes of the 2024 level in the Business Administration major of a vocational college could provide the key results needed. Furthermore, the experimental group (48 people) adopted the optimized adaptive learning path, while the results might indicate that the control group (47 people) could demonstrate meaningful differences under the traditional fixed learning path. In light of the data, two core online courses — "Fundamentals of Management" and "Marketing" — appear to support that the experimental period of one semester (18 weeks) could establish sufficient evidence. Study shows variables controlled strictly. However, the significant evidence may suggest that teaching staff, course content, and assessment standards remained consistent across both groups, with the key findings indicating that only differences in learning path design could demonstrate experimental validity. Additionally, the important data collection may indicate that three dimensions — learning effectiveness data (course final grades, knowledge mastery test scores), learning efficiency data (course completion rate, average learning duration), and learning satisfaction data — could provide the significant results needed. Given that the findings demonstrate that questionnaire surveys using a 5-point scoring system were used, the evidence might suggest that this approach could support the reliability of the satisfaction measures. Notwithstanding these results, the study may indicate that the collected data appears to require statistical analysis. SPSS shows significance tests ensure reliability. This study was approved by the Ethics

Committee of Jiangxi Hongzhou Vocational College, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form.

B. ANALYSIS OF EXPERIMENTAL RESULTS

The experimental results (as shown in Figure 3) showed that the learning effectiveness, efficiency, and satisfaction of the experimental group students were significantly better than those of the control group. The specific analysis is as follows: in terms of learning effectiveness, the average final scores of the experimental group students in “Fundamentals of Management” and “Marketing” were 82.6 and 81.8 points, respectively, while those of the control group were 68.9 and 69.2 points, respectively. The average scores of the experimental group increased by 13.7 and 12.6 points, respectively; In the knowledge mastery test, the average accuracy rate of the experimental group students was 83.2%, while that of the control group was 61.7%, an improvement of 21.5%. This indicates that the optimized learning path can effectively help students master knowledge points and improve learning effectiveness. In terms of learning efficiency, the completion rate of the experimental group students was 95.8%, while the control group was 77.5%, an increase of 18.3%; The average learning time of the experimental group students was 42.3 hours, while the control group was 51.7 hours, a reduction of 9.4 hours. This indicates that the optimized learning path can allocate learning resources and time reasonably, and improve learning efficiency. In terms of learning satisfaction, the average satisfaction score of the experimental group students was 4.2 points, while the control group was 3.6 points, an increase of 15.7%. Among them, the satisfaction score of the experimental group was significantly higher than that of the control group in the three dimensions of “personalized adaptation”, “reasonable learning pace” and “fit for career needs”, indicating that the optimized learning path is more in line with students’ personalized needs and career development expectations. Through significance testing ($P<0.05$), there is a significant difference between the two sets of data, proving the effectiveness of the adaptive learning path optimization model based on deep reinforcement learning proposed in this paper.

C. RESULTS

Based on the experimental results, conduct in-depth discussions on the advantages, shortcomings, and improvement directions of the model. Firstly, the advantages of the model are mainly reflected in two aspects: firstly, it has strong dynamic adaptability. Through real-time interaction between the intelligent agent of deep reinforcement learning and the learning environment, it can respond to changes in students’ learning status in a timely manner, dynamically adjust the learning path, solve the problem of homogenization in traditional fixed learning paths, and achieve “teaching according to students’ aptitude”; Secondly, it has high adaptability. The model is designed with a reward function that combines the job competency requirements of the Business Administration major in higher vocational education, and combines the learning path with career needs, effectively enhancing the pertinence and practicality of learning and meeting the training goals of applied talents in higher vocational education. Secondly, the shortcomings of the model mainly include: firstly, the model training has a strong dependence on the amount of data, and if some students’ learning behavior data is insufficient, it will affect the recommendation accuracy of the model; Secondly, the weight coefficients of the reward function are fixed and cannot be dynamically adjusted according to the personalized characteristics of students, which may affect the learning experience of some students; The third issue is the insufficient adaptability of the model to learning resources, which fails to fully integrate the characteristics of case teaching in higher vocational business management majors, and the targeted recommendation of learning resource types needs to be improved. Finally, the improvement direction is to optimize the data collection mechanism, increase the dimensions and quantity of data collection, and use data augmentation technology to compensate for the insufficient data of some students; Design a dynamic weight adjustment mechanism to adjust the weight coefficients of the reward function in real-time based on students’ learning preferences and career planning; Combining the characteristics of case teaching in higher vocational business education, enriching the types of learning resources, enhancing the pertinence of learning resource recommendations, and further optimizing model performance.

V. CONCLUSION

This article is based on the optimization of adaptive learning paths of online courses in higher vocational business management majors, and carries out research based on deep reinforcement learning technology to solve problems such as homogenization of traditional learning paths, lack of dynamic adaptability, and lack of close integration with career needs. The research concludes that an adaptive learning path optimization model based on DQN algorithm can dynamically generate personalized learning paths by designing reasonable state space, action space and multi-dimensional reward functions; After the model is applied to the online courses of Business Administration in vocational colleges, it can improve students’ learning effectiveness, efficiency and satisfaction, and achieve precise integration of “personalized learning career needs”. There are still some deficiencies in this study, such as the model’s strong dependence on data amount, failure to dynamically adjust the weights of reward function, and need to improve the adaptability of learning resources. In the future, the model can be improved by data augmentation, dynamic weight optimization and improvement of adaptability of learning resources. The theoretical value of this study is to enrich the application of deep reinforcement learning in higher vocational education and construct an adaptive learning path optimization theoretical framework applicable to higher vocational business management majors; The actual value is to provide feasible solutions for the reform of online course teaching in vocational colleges, which can improve the quality of applied talent cultivation in business management major of vocational colleges, and provide reference and inspiration for the reform of online education in similar vocational majors.

AUTHOR CONTRIBUTIONS

Zhenqian Li designed, collected and analyzed the data, and drafted the manuscript. Zhenqian Li conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Zhenqian Li participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

HUMAN RESEARCH SUBJECTS

This study was approved by the Ethics Committee of Jiangxi Hongzhou Vocational College, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form.

FUNDING

This research was supported by, Research and Reform Project on Higher Education Teaching of Jiangxi Provincial Regular Undergraduate Colleges: Cultivating Public Trust in Ideological and Political Courses Based on Effectiveness Evaluation (BKJG-2026-18-35)

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] M. Samieiyeganeh, K. Rahmat R W O, B. Khalid F, and A. Kasmiran K, "Deep reinforcement learning to multi-agent deep reinforcement learning," Journal of Theoretical and Applied Information Technology, vol. 100, no. 4, pp. 990-1003, 2022.

[2] A. Nurmuhammet, "Deep reinforcement learning on stock data," Alatoo Academic Studies, vol. 23, no. 2, pp. 505-518, 2023.

[3] Z. S. Fan, "An exploration of reinforcement learning and deep reinforcement learning," Applied and Computational Engineering, vol. 73, no. 1, pp. 154-159, 2024.

[4] B. Jaeger and A. Geiger, "An Invitation to Deep Reinforcement Learning," Foundations and Trends in Optimization, vol. 7, no. 1, pp. 1-80, 2024, doi: 10.1561/2400000049.

[5] D. Karimzadeh, "Deep Reinforcement Learning Multi-Agent Systems," International journal of Modern Achievement in Science, Engineering and Technology, vol. 2, no. 1, pp. 56-62, 2024, doi: 10.63053/ijset.60.

[6] M. Wang, "Deep reinforcement learning for stock prediction," Applied and Computational Engineering, vol. 69, no. 1, pp. 85-90, 2024, doi: 10.1155/2022/5812546.

[7] S. Pradhan, "Evaluating Deep Reinforcement Learning Algorithms," INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT, vol. 08, no. 008, pp. 1-6, 2024.

[8] L. Liu and L. H. Chen, "Research progress about deep reinforcement learning," Mechatronic Systems and Control, vol. 51, no. 4, pp. 210-217, 2023.

[9] M. Skuba, A. Janota, Pavol Kuchár, et al., "Deep Reinforcement Learning for Traffic Signal Control," Transportation Research Procedia, vol. 74, no. c, pp. 954-958, 2023.

[10] Y. Guo and Z. Liu, "UAV Path Planning Based on Deep Reinforcement Learning," International Journal of Advanced Network, Monitoring and Controls, vol. 8, no. 3, pp. 81-88, 2023, doi: 10.2478/ijanmc-2023-0068.

[11] K. G. Krishnan, "Using Deep Reinforcement Learning For Robot Arm Control," Journal of Artificial Intelligence and Capsule Networks, vol. 4, no. 3, pp. 160-166, 2022.

[12] T. Zhang, Y. Li, X. Miao, Z. Wei, L. Jia, Q. Gong, and T. Wen, "AUV 3D docking control using deep reinforcement learning," Ocean engineering, vol. 283, no. 1, pp. 1.1-1.12, 2023, doi: 10.1016/j.oceaneng.2023.115021.

[13] N. S. Raj and V. G. Renumol, "An improved adaptive learning path recommendation model driven by real-time learning analytics," Journal of Computers in Education, vol. 11, no. 1, pp. 121-148, 2024, doi: 10.1007/s40692-022-00250-y.