Proceedings of the 4th International Conference on Mathematical Physics and Computational Simulation
DOI: 10.54254/2753-8818/2026.33408

The Application and Practice of Reinforcement Learning in Recommendation Systems

Yiyang Wang

Swjtu-Leeds Joint School, Southwest Jiaotong University, Chengdu, China
wangyiyang@my.swjtu.edu.cn

Abstract. The application scenarios of recommendation systems are becoming increasingly complex, and traditional algorithms struggle to balance users' short-term click behaviors with their long-term value demands. Reinforcement learning, with its dynamic policy iteration capability for the interaction between the subject and the environment, provides a core technical path for optimizing the long-term value of recommendation systems. This paper focuses on three typical recommendation scenarios: e-commerce, games, and knowledge retrieval. It systematically reviews the algorithm paradigms of reinforcement learning and their applicability in different scenarios, compares the application logic differences between product recommendation and content recommendation, and summarizes the verification results of basic reinforcement learning recommendation models based on public datasets. The research shows that value-based reinforcement learning algorithms are suitable for product recommendation scenarios, while policy-based algorithms are more suitable for content recommendation scenarios. The related research results provide theoretical references and practical basis for the application of reinforcement learning in industrial-level recommendation scenarios.

Keywords: reinforcement learning, recommendation system, long-term value, scenario adaptability

1. Introduction

Recommendation systems serve as the core carrier for connecting user needs with information resources and have been widely applied in e-commerce and content platforms. Their technological iterations directly affect user experience and the commercial value of the platform. Most traditional recommendation algorithms focus on short-term clicks and conversions, which can lead to the problem of information cocoons and make it difficult to achieve the coordinated optimization of users' long-term value and the long-term interests of the platform. Reinforcement learning has core characteristics such as dynamic decision-making and long-term cumulative reward optimization. It can optimize recommendation strategies through the continuous interaction and iteration of intelligent agents and the environment, becoming a key technology to solve the limitations of traditional algorithms and attracting close attention from both the academic and industrial communities.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.33408

This paper analyzes the technical application differences in three typical scenarios of e-commerce product recommendation, game content recommendation, and knowledge retrieval, summarizes the existing research conclusions and practical experiences of reinforcement learning recommendation models, and provides theoretical references and technical support for the large-scale application of reinforcement learning in the recommendation system field.

2. Application scenarios

2.1. E-commerce scenarios

The core objective of e-commerce recommendation is to enhance the long-term consumption value of users. It mainly involves three aspects: improving conversion efficiency, increasing repeat purchase rates, and expanding product varieties. Reinforcement learning technology has been widely applied in core modules such as homepage recommendations, "You may also like" sections, and personalized searches. Current research focuses on three dimensions: modeling of user behavior sequences, design of multi-objective reward functions, and joint optimization offline and online. This has formed a mature technology application system.

In the field of modeling user behavior sequences, Wang Jian et al. proposed an e-commerce recommendation model based on deep reinforcement learning. By introducing an attention mechanism to model the long-term user behavior sequence, this model can accurately capture users' potential consumption demands. This model has been verified in the e-commerce product recommendation scenario and effectively enhanced users' willingness to repurchase and average order value [1]. In the aspect of multi-objective reward optimization, Liu Jian et al. constructed a multi-dimensional reward function that integrates GMV (total purchase amount), repurchase rate, and user satisfaction to solve the recommendation bias problem caused by a single reward orientation. This enables the model to achieve multi-objective collaborative optimization in large-scale traffic environments [2]. In addition, Huang Zhigang et al. proposed a deep hierarchical reinforcement learning framework, which uses time abstraction methods to decompose the complex commodity recommendation problem into multiple sub-problems. This framework is suitable for the characteristics of discrete candidate commodity spaces in e-commerce and can ensure the stability of model convergence while improving the long-term recommendation effect [3].

2.2. Game scenarios

The game scene features core characteristics such as continuous decision-making, long-term confrontation, and multi-agent interaction. Reinforcement learning is mainly applied in areas such as intelligent control of Non-player Characters (NPCs), adaptive adjustment of level difficulty, and optimization of player experience. It dynamically adjusts the recommended strategies to meet the complex interaction requirements of the game scene.

Zhao X. et al. focused on the continuous decision-making feature of the game scene and proposed a multi-task recommendation framework based on deep reinforcement learning. They restructured game content recommendation into a multi-objective Markov decision process and optimized the content recommendation effect for different player groups through task-specific strategies [4]. In terms of player experience optimization, Largue et al. deeply integrated reinforcement learning with the environmental feedback mechanism, achieved dynamic adaptive adjustment of game difficulty, and optimized the recommended content and level design based on real-time player operation feedback, effectively prolonging the player's life cycle [5]. Chen Tao. et al. from the perspective of

DOI: 10.54254/2753-8818/2026.33408

interpretability, proposed a behavior explanation framework for reinforcement learning recommendation models, providing theoretical support for the deployment of the model and its compliant application in the game recommendation scene [6].

2.3. Knowledge retrieval scenarios

The core requirement of knowledge retrieval scenarios is to precisely match users' query intentions, improve the relevance of retrieval results and the efficiency of knowledge acquisition, and learning has become a key technology for improving the quality of knowledge retrieval through dynamic optimization of sorting strategies and adaptation to user feedback information.

Li Jie et al. proposed a deep reinforcement learning retrieval model based on multi-attention mechanisms, which integrates semantic representation and reinforcement learning decision-making mechanisms. This alleviates the defect of traditional retrieval models' over-reliance on keyword matching and effectively improves retrieval accuracy under complex query intentions [7]. In the aspect of user feedback optimization, Zhang Yu et al. constructed a dynamic retrieval adjustment mechanism based on reinforcement learning, continuously optimizing the document sorting strategy by analyzing user clicks, stay time, secondary queries, etc., feedback data, and gradually improving retrieval quality over the long term [8]. The research conducted by Cao Hongye and others has further expanded the integration path of semantic representation and reinforcement learning, integrating knowledge graph information into the decision-making process of reinforcement learning, providing technical support for search scenarios in specific vertical fields such as intelligent question answering and literature retrieval [9].

3. Dataset

3.1. MovieLens-1M

This dataset contains approximately 1 million user and movie interaction records, covering core fields such as user basic information, movie classification labels, viewing history, and rating data. The data size is moderate and the structure is clear. It is a classic dataset for verifying discrete action space recommendation models. Zhao X et al. used this dataset to verify the recommendation effects of basic reinforcement learning algorithms such as DQN in their research. By comparing with traditional collaborative filtering algorithms, they found that the long-term user satisfaction of the reinforcement learning model increased by approximately 12%, providing standardized reference for the performance verification of basic algorithms [4].

3.2. Amazon review data

This dataset covers interaction data of various product categories, including user reviews, purchase records, ratings, product attributes, and other multi-dimensional information. It is large in scale and covers various scenarios, suitable for algorithm verification in e-commerce recommendation scenarios. Wang Jian et al. used this dataset in their research on e-commerce scenarios to train a multi-objective reinforcement learning recommendation model, to verify the improvement effect of multi-dimensional reward functions on recommendation performance, and to provide data support for the practical application of e-commerce recommendation models [1].

DOI: 10.54254/2753-8818/2026.33408

3.3. Karu recommendation

This dataset focuses on the short video recommendation scenario, including interaction data such as viewing duration, likes, comments, and forwards, as well as video content features and user information. It is highly aligned with the business logic of short video platforms. Zhang Yu et al. used this dataset to verify the applicability of reinforcement learning in content recommendation scenarios and confirmed that the policy-based reinforcement learning algorithm can effectively improve user retention and interaction rates in short video recommendations [8].

4. Current limitations and future development prospects

4.1. Current limitations

Currently, there are three core limitations in applying reinforcement learning in recommendation systems: Firstly, the design of the reward function is extremely challenging. Most existing research relies on manually designed fixed reward functions, which are difficult to adapt to the dynamic changes of user needs and platform business goals, and may lead to deviations in recommendation strategies. Secondly, the sample efficiency and convergence issues are prominent. Reinforcement learning models require a large number of interaction samples to complete training, and there are defects such as high training costs and slow convergence in large-scale industrial scenarios. Thirdly, the interpretability of the model is insufficient. The "black box" nature of reinforcement learning makes it difficult to clearly present the decision logic of the recommendation results, and its application in scenarios such as e-commerce and healthcare, which require high interpretability, is limited [6,9].

4.2. Future outlook

Regarding future research on reinforcement learning recommendation systems, it can be carried out along the following three main directions: Firstly, design adaptive reward functions, combining large language models with dynamic reward functions generated based on user profiles to achieve real-time optimization of reward strategies; Secondly, develop lightweight reinforcement learning algorithms, reducing model training and deployment costs through techniques such as model compression and knowledge distillation, adapting to large-scale industrial scenarios; Thirdly, study interpretable reinforcement learning recommendation models, introducing technologies such as attention mechanisms and behavior visualization to clarify the decision logic of recommendation results, expanding the application scope of reinforcement learning in highly sensitive scenarios [7,10].

5. Conclusion

This article conducts a systematic review of the application of reinforcement learning in recommendation systems, clearly identifying reinforcement learning as the core technical path for addressing the insufficient long-term value optimization of traditional recommendation algorithms. By analyzing the algorithm application patterns in three typical scenarios - e-commerce, games, and knowledge retrieval, it is confirmed that the commodity recommendation scenario is suitable for value-based reinforcement learning algorithms, while the content recommendation scenario is more suitable for strategy-based algorithms. By summarizing the existing research results based on public

DOI: 10.54254/2753-8818/2026.33408

datasets, it is demonstrated that reinforcement learning can effectively alleviate the information cocoon problem and enhance user long-term satisfaction and platform revenue.

Meanwhile, this paper analyzes the limitations of current reinforcement learning recommendation systems, such as the difficulty in designing the reward function, low sample efficiency, and insufficient interpretability. It proposes three future research directions: adaptive reward function, lightweight algorithm, and interpretable model. The review study systematically summarizes the application framework and scenario adaptation rules of reinforcement learning in recommendation systems, providing theoretical references and technical support for the practical application of reinforcement learning in industrial-level recommendation scenarios, and promoting the development of recommendation system technology towards long-term value optimization and intelligent adaptation.

References

[1] Wang, J., & Li, L. (2023). Research progress on the application of deep reinforcement learning in recommendation systems. Journal of Software, 34(2), 567–592.
[2] Liu, J., & Zhang, W. (2022). A review on long-term user value optimization in recommendation systems. Journal of Computer Science, 45(6), 1234–1256.
[3] Huang, Z., Liu, Q., Zhang, L., et al. (2023). Research and development of deep hierarchical reinforcement learning. Journal of Software, 34(2), 733–760.
[4] Zhao, X., et al. (2022). Deep reinforcement learning for online recommendation in e-commerce. In Proceedings of the IEEE Conference.
[5] Raghu, R. B., Lakshya, K., & Sreekanth, V. (2022). Fine-grained session recommendations in e-commerce using deep reinforcement learning. arXiv preprint arXiv: 2210.15451.
[6] Chen, T., & Zhao, M. (2024). A review on explainable reinforcement learning. Acta Automatica Sinica, 50(1), 1–20.
[7] Li, J., & Wang, H. (2024). Deep reinforcement learning recommendation system algorithm based on multi-level attention mechanisms. Electronics, 13(23), 4625.
[8] Zhang, Y., & Chen, L. (2023). Research on the application of reinforcement learning in content recommendation systems. Computer Application Research, 40(7), 2015–2020.
[9] Cao, H., Liu, X., Dong, S., et al. (2024). A review on explainability for reinforcement learning. Journal of Computer Science, 47(8), 1854–1878.
[10] Alibaba Technology Team. (2021, May 10). Reinforcement learning practices in Taobao's personalized recommendation system. Retrieved from https: //developer.aliyun.com/article/781234.