A comprehensive survey on reinforcement learning-based recommender systems: State-of-the-art, challenges, and future perspectives

Oleksandr D. Rossiiev[1], Nonna N. Shapovalova[1], Olena H. Rybalchenko[1] and Andrii M. Striuk[1,2,3]

[1]*Kryvyi Rih National University, 11 Vitalii Matusevych Str., Kryvyi Rih, 50027, Ukraine*
[2]*Kryvyi Rih State Pedagogical University, 54 Universytetskyi Ave., Kryvyi Rih, 50086, Ukraine*
[3]*Academy of Cognitive and Natural Sciences, 54 Universytetskyi Ave., Kryvyi Rih, 50086, Ukraine*

Abstract

Recommender systems play a crucial role in helping users navigate the vast amount of information available in the digital age. Traditional recommendation approaches, such as collaborative filtering and content-based methods, often face challenges in dealing with dynamic user preferences, sparse feedback, and long-term user engagement. Reinforcement learning has emerged as a promising framework to address these limitations by formulating the recommendation problem as a sequential decision-making process and learning optimal recommendation strategies through interactions with users. This survey provides a comprehensive overview of the state-of-the-art research on reinforcement learning-based recommender systems. We review the foundations of reinforcement learning in the context of recommendations, including the Markov decision process formulation, and explore various reinforcement learning algorithms and architectures used in recommender systems, such as model-free, model-based, policy gradient, and deep reinforcement learning methods. We also examine the integration of reinforcement learning with other techniques, such as collaborative filtering, content-based methods, knowledge graphs, and graph neural networks, to enhance the performance and capabilities of recommender systems. Furthermore, we identify key challenges and future research directions in this field, including offline reinforcement learning, scalability, explainability, robustness, evaluation metrics, and real-world applications.

Keywords

recommender systems, reinforcement learning, collaborative filtering, content-based filtering, knowledge graphs, graph neural networks, Markov decision process, offline reinforcement learning, explainability, robustness, evaluation metrics, real-world applications

1. Introduction

1.1. Background on recommendation systems and their importance

Recommender systems have become an indispensable tool for helping users navigate the vast amount of information available in the digital age. These systems are designed to provide personalized recommendations by predicting user preferences and suggesting relevant items, such as products, services, or content [1]. Recommender systems have been widely adopted in various domains, including e-commerce [2], streaming services [3], social media [4], and even healthcare [5], due to their ability to enhance user experience, increase user engagement, and drive business growth.

1.2. Limitations of traditional recommendation approaches

Traditional recommendation approaches, such as collaborative filtering [6, 7, 8] and content-based filtering [1], have been extensively studied and applied in various recommender systems. However,

CS&SE@SW 2024: 7th Workshop for Young Scientists in Computer Science & Software Engineering, December 27, 2024, Kryvyi Rih, Ukraine
shapovalova@knu.edu.ua (N. N. Shapovalova); rybalchenko@knu.edu.ua (O. H. Rybalchenko); andrey.n.stryuk@gmail.com (A. M. Striuk)
http://mpz.knu.edu.ua/nona-shapovalova/ (N. N. Shapovalova); http://mpz.knu.edu.ua/olena-rybalchenko/ (O. H. Rybalchenko); http://mpz.knu.edu.ua/andrij-stryuk/ (A. M. Striuk)
0000-0001-9146-1205 (N. N. Shapovalova); 0000-0001-8691-5401 (O. H. Rybalchenko); 0000-0001-9240-1976 (A. M. Striuk)

these methods often face several limitations. First, they struggle with the cold-start problem [9], where the system lacks sufficient information about new users or items to make accurate recommendations. Second, traditional methods may not scale well to large datasets, as they require extensive computational resources [10]. Finally, these approaches often fail to adapt to the dynamic nature of user preferences, which can change over time [11].

1.3. Potential of reinforcement learning in recommendation systems

Reinforcement learning (RL) has emerged as a promising approach to address the limitations of traditional recommendation methods. RL is a type of machine learning that enables an agent to learn optimal decision-making strategies through interactions with an environment [12]. By formulating the recommendation problem as a Markov decision process (MDP) and using RL algorithms to optimize the recommendation policy, recommender systems can effectively handle the sequential nature of user interactions, optimize for long-term user engagement, and adapt to changing user preferences [13, 14]. Moreover, RL-based recommender systems have the potential to deal with sparse feedback and balance exploration and exploitation, which are crucial for improving recommendation quality [15].

1.4. Research objectives and questions

The primary objective of this survey is to provide a comprehensive overview of the current state-of-the-art in reinforcement learning-based recommender systems. We aim to address the following research questions:

RQ1: What are the key advantages of using reinforcement learning in recommender systems compared to traditional approaches?
RQ2: How can reinforcement learning be formulated and applied to the recommendation problem?
RQ3: What are the main reinforcement learning algorithms and architectures used in recommender systems?
RQ4: How can reinforcement learning be integrated with other techniques, such as collaborative filtering, content-based methods, and deep learning, to improve recommendation performance?
RQ5: What are the current challenges and future research directions in reinforcement learning-based recommender systems?

1.5. Contributions and novelty of the survey

This survey makes the following contributions to the field of reinforcement learning-based recommender systems:

- We provide a comprehensive and up-to-date review of the state-of-the-art in reinforcement learning-based recommender systems, covering a wide range of approaches, algorithms, and applications.
- We propose a novel taxonomy for categorizing and analyzing reinforcement learning-based recommender systems based on their problem formulation, algorithmic approaches, and integration with other techniques.
- We identify key challenges and future research directions in the field, including offline reinforcement learning, scalability, explainability, robustness, and evaluation metrics.
- We discuss real-world applications and case studies of reinforcement learning-based recommender systems, highlighting their potential impact and practical considerations.

To the best of our knowledge, this is the first survey that extensively covers the integration of reinforcement learning with other techniques, such as knowledge graphs and graph neural networks, in the context of recommender systems. Our survey aims to bridge the gap between theory and practice by providing a comprehensive overview of the field and offering actionable insights for researchers and practitioners.

2. Methodology

2.1. Literature search strategy and inclusion criteria

To ensure a comprehensive and systematic review of the literature on reinforcement learning-based recommender systems, we adopted a rigorous search strategy and inclusion criteria. We conducted searches on Scopus, using a combination of keywords such as "reinforcement learning", "recommender systems", "collaborative filtering", "content-based filtering", "deep learning", and "graph neural networks". We also explored relevant articles from the reference lists of the retrieved papers to identify additional studies.

The inclusion criteria for the selected papers were as follows:

1. The study must be written in English and published in a peer-reviewed journal, conference proceedings, or book chapter.
2. The study must focus on the application of reinforcement learning in recommender systems or the integration of reinforcement learning with other recommendation techniques.
3. The study must provide sufficient technical details on the proposed approach, including the problem formulation, algorithmic design, and experimental evaluation.

We initially retrieved a total of 253 papers based on our search strategy. After applying the inclusion criteria and removing duplicates, we obtained a final set of 56 papers that form the basis of this survey.

2.2. Categorization and analysis framework

To systematically analyze and present the findings from the selected papers, we propose a categorization and analysis framework that consists of three main dimensions:

2.2.1. Problem formulation

We categorize the studies based on how they formulate the recommendation problem as a reinforcement learning task. This includes the definition of states, actions, rewards, and the underlying MDP framework. By examining the problem formulation, we can gain insights into the key challenges and considerations in applying reinforcement learning to recommender systems.

2.2.2. Algorithmic approaches

We classify the studies according to the reinforcement learning algorithms and architectures they employ. This includes model-free methods (e.g., Q-learning [16], SARSA [17]), model-based methods, policy gradient methods (e.g., REINFORCE [18], Actor-Critic [19]), and deep reinforcement learning (e.g., DQN [20], DDPG [21]). By analyzing the algorithmic approaches, we can identify the strengths and weaknesses of different reinforcement learning techniques in the context of recommender systems.

2.2.3. Integration with other techniques

We investigate how reinforcement learning is integrated with other recommendation techniques, such as collaborative filtering [22], content-based methods [23], knowledge graphs [24], and graph neural networks [25]. By examining the integration strategies, we can uncover the synergies and complementary advantages of combining reinforcement learning with traditional recommendation approaches and advanced deep learning architectures.

Based on this categorization and analysis framework, we provide a comprehensive and structured review of the state-of-the-art in reinforcement learning-based recommender systems. In the following sections, we present the key findings, insights, and future research directions in each of the three dimensions.

3. Traditional recommendation methods

In this section, we provide an overview of traditional recommendation methods, including collaborative filtering, content-based filtering, and hybrid methods. We also discuss the limitations and challenges of these approaches, which motivate the adoption of reinforcement learning in recommender systems.

3.1. Overview of traditional approaches

3.1.1. Collaborative filtering

Collaborative filtering (CF) is one of the most widely used recommendation techniques. It relies on the assumption that users with similar preferences in the past are likely to have similar preferences in the future [6]. CF methods can be further divided into memory-based and model-based approaches. Memory-based CF directly uses the user-item interaction data to compute similarity scores between users or items, while model-based CF learns a predictive model from the interaction data [7]. Examples of CF methods include user-based CF, item-based CF [26], and matrix factorization [27].

3.1.2. Content-based filtering

Content-based filtering (CBF) recommends items to users based on the similarity between the content of the items and the user's preferences [1]. CBF methods typically represent items using a set of features or attributes, such as genres, keywords, or user-generated tags. User profiles are constructed based on the features of the items they have interacted with in the past. The recommendation process involves matching the user profile with the item features to generate personalized suggestions. CBF methods have been applied in various domains, such as movie recommendation [28] and news article recommendation [29].

3.1.3. Hybrid methods

Hybrid recommendation methods combine multiple recommendation techniques to leverage their complementary strengths and mitigate their individual limitations [30]. Common hybridization strategies include weighted averaging, switching, cascading, and feature combination. For example, Jafarkarimi et al. [31] proposed a hybrid recommender system that integrates CF and CBF using a weighted averaging approach, while Ghazanfar and Prügel-Bennett [32] developed a switching hybrid that selects between CF and CBF based on the availability of user preference data.

3.2. Limitations and challenges of traditional methods

3.2.1. Cold-start problem

The cold-start problem refers to the difficulty of making accurate recommendations for new users or items that have little or no interaction data [33]. CF methods are particularly vulnerable to the cold-start problem, as they rely heavily on the existence of sufficient user-item interactions. CBF methods can alleviate the cold-start problem to some extent by leveraging item content features, but they still require a minimum amount of user feedback to build reliable user profiles [9].

3.2.2. Scalability issues

Traditional recommendation methods often face scalability issues when dealing with large-scale datasets [10]. Memory-based CF methods have high computational complexity, as they need to calculate similarity scores between all pairs of users or items. Model-based CF methods, such as matrix factorization, can be more efficient but still require substantial computational resources for training and updating the models. CBF methods may also suffer from scalability issues when the number of items and their associated features grow large.

3.2.3. Lack of adaptability to dynamic user preferences

User preferences are not static and can evolve over time due to various factors, such as changes in personal tastes, social influences, and contextual situations [11]. Traditional recommendation methods often struggle to adapt to these dynamic preferences, as they typically learn from historical interaction data without considering the temporal aspects. This can lead to suboptimal recommendations that fail to capture the users' current interests and needs. Incorporating temporal dynamics and sequential patterns into recommendation models is crucial for improving their adaptability and responsiveness to changing user preferences.

The limitations and challenges of traditional recommendation methods highlight the need for more advanced and flexible approaches that can handle the complexities of real-world recommendation scenarios. In the next section, we explore how reinforcement learning can be leveraged to address these issues and enhance the performance of recommender systems.

4. Reinforcement learning in recommender systems

Reinforcement learning has emerged as a promising approach to address the limitations of traditional recommendation methods. By formulating the recommendation problem as a Markov decision process and using RL algorithms to optimize the recommendation policy, recommender systems can effectively handle the sequential nature of user interactions, optimize for long-term user engagement, and adapt to changing user preferences. In this section, we provide a comprehensive overview of the formulation of recommendation as an RL problem, discuss the advantages of RL for recommendations, and review the main RL approaches and their integration with other techniques.

4.1. Formulation of recommendation as a reinforcement learning problem

4.1.1. Markov decision process (MDP) framework

The first step in applying RL to recommender systems is to formulate the recommendation problem as an MDP. An MDP is defined by a tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, $\mathcal{P}$ is the transition probability function, $\mathcal{R}$ is the reward function, and $\gamma \in [0, 1)$ is the discount factor [12]. In the context of recommender systems, the state space $\mathcal{S}$ represents the user's current context and preference, the action space $\mathcal{A}$ corresponds to the set of items that can be recommended, the transition probability function $\mathcal{P}$ models the user's behavior in response to the recommended items, and the reward function $\mathcal{R}$ measures the user's satisfaction or engagement with the recommendations [13].

4.1.2. Key components: states, actions, rewards

The design of states, actions, and rewards is crucial for the success of RL-based recommender systems. The state representation should capture the relevant information about the user's current context and historical interactions, such as the user's demographic attributes, past clicked or purchased items, and session-level features [14]. The action space can be defined as the entire item catalog or a subset of items selected based on certain criteria, such as popularity or relevance to the user's preferences [19]. The reward function should reflect the system's optimization objective, which can be user engagement metrics (e.g., click-through rate, dwell time), business metrics (e.g., revenue, conversion rate), or a combination of both [17].

4.2. Advantages of reinforcement learning for recommendations

4.2.1. Handling sequential user-system interactions

One of the key advantages of RL for recommendations is its ability to handle the sequential nature of user-system interactions. Unlike traditional methods that treat each interaction independently, RL

algorithms can learn from the entire sequence of interactions and optimize the recommendation policy based on the long-term cumulative rewards [15]. This enables the recommender system to capture the temporal dynamics of user preferences and adapt its recommendations accordingly. For example, Zheng et al. [34] proposed a deep RL framework for e-commerce recommendations that learns to optimize the long-term user engagement by considering the sequential dependencies between user actions and system recommendations.

4.2.2. Optimizing long-term user engagement

Another benefit of RL-based recommender systems is their focus on optimizing long-term user engagement rather than immediate rewards. Traditional recommendation methods often aim to maximize short-term metrics, such as click-through rate or conversion rate, which may lead to suboptimal performance in the long run. In contrast, RL algorithms can optimize for cumulative rewards over a longer horizon, taking into account the future impact of current recommendations on user satisfaction and retention [11]. This long-term optimization perspective aligns well with the business objectives of many recommendation platforms, as it helps to build a loyal user base and increase customer lifetime value.

4.2.3. Dealing with sparse feedback and exploration

RL-based recommender systems can also effectively deal with sparse user feedback and balance the exploration-exploitation trade-off. In real-world recommendation scenarios, user feedback is often implicit and sparse, as users only interact with a small fraction of the available items. RL algorithms can handle this sparsity by learning from the limited feedback and using exploration strategies to gather more information about the user's preferences [35]. By balancing exploration (recommending novel or less certain items) and exploitation (recommending items with high estimated rewards), RL-based recommenders can continuously improve their performance and adapt to changing user interests.

4.3. Reinforcement learning approaches for recommendations

4.3.1. Model-free methods

Model-free RL methods learn the optimal recommendation policy directly from the interactions with users, without explicitly modeling the environment dynamics. Q-learning and SARSA are two popular model-free RL algorithms that have been applied to recommender systems. Q-learning [36] learns the action-value function $Q(s, a)$, which represents the expected cumulative reward of taking action a in state s and following the optimal policy thereafter. Munemasa et al. [16] proposed a Q-learning-based recommender system that learns to recommend items based on the user's historical interactions and the estimated Q-values. SARSA (State-Action-Reward-State-Action) [37] is another model-free algorithm that updates the Q-values based on the actual actions taken by the system, rather than the optimal actions. Xia et al. [17] developed a SARSA-based recommender system for energy optimization in smart buildings, which learns to provide personalized recommendations based on the user's comfort preferences and energy consumption patterns.

4.3.2. Model-based methods

Model-based RL methods learn a model of the environment dynamics and use it to plan the optimal recommendation policy. These methods can be more sample-efficient than model-free approaches, as they can leverage the learned model to simulate the user's behavior and optimize the policy offline. Chen et al. [38] proposed a model-based RL framework for news recommendation, which learns a user behavior model from historical interactions and uses it to generate synthetic trajectories for policy optimization. Gunawardana and Meek [39] developed a model-based RL approach for e-commerce recommendations, which learns a customer behavior model using a variational autoencoder and optimizes the recommendation policy using the learned model.

4.3.3. Policy gradient methods

Policy gradient methods directly optimize the recommendation policy by computing the gradient of the expected cumulative reward with respect to the policy parameters. REINFORCE [40] is a well-known policy gradient algorithm that updates the policy parameters in the direction of the estimated gradient. Xin et al. [18] proposed a self-supervised RL framework for sequential recommendations, which uses REINFORCE to optimize the recommendation policy based on the user's feedback and a self-supervised learning objective. Actor-Critic methods [41] combine the advantages of value-based and policy-based methods by learning both a value function (critic) and a policy function (actor). Zhao et al. [19] developed an Actor-Critic-based recommender system for online advertising, which learns to optimize the ad selection policy based on the user's click feedback and the estimated state values.

4.3.4. Deep reinforcement learning

Deep RL methods incorporate deep neural networks into the RL framework to learn complex state representations and policies. Deep Q-Network (DQN) [42] is a popular deep RL algorithm that uses a neural network to approximate the Q-function and stabilizes the learning process using experience replay and target networks. Zheng et al. [34] applied DQN to e-commerce recommendations, where the state is represented by the user's historical interactions and the actions correspond to the recommended items. Deep Deterministic Policy Gradient (DDPG) [43] is another deep RL algorithm that combines the advantages of DQN and Actor-Critic methods, using neural networks to learn both the Q-function and the policy. Chen et al. [44] proposed a DDPG-based recommender system for diversified recommendations, which learns to balance the trade-off between accuracy and diversity in the recommendation policy.

4.4. Integration of reinforcement learning with other techniques

4.4.1. Combining RL with collaborative filtering and content-based methods

RL can be integrated with traditional recommendation methods, such as collaborative filtering (CF) and content-based filtering (CBF), to leverage their complementary strengths. Choi et al. [45] proposed a hybrid RL-CF approach for movie recommendations, which uses RL to learn the optimal recommendation policy based on the user's feedback, and CF to generate candidate items for each state. Gupta and Katarya [4] developed an RL-CBF framework for news recommendation, which uses CBF to represent the user's preferences and RL to optimize the recommendation policy based on the user's engagement with the recommended articles.

4.4.2. Incorporating knowledge graphs and graph neural networks

Knowledge graphs (KGs) and graph neural networks (GNNs) can be incorporated into RL-based recommender systems to provide rich semantic information and capture complex user-item relationships. Zhang et al. [46] proposed a KG-enhanced RL framework for news recommendation, which uses a KG to represent the semantic relationships between news articles and entities, and RL to optimize the recommendation policy based on the user's feedback. Liu et al. [47] developed a GNN-based RL approach for social recommendations, which uses a GNN to learn the user and item embeddings from the social network structure, and RL to optimize the recommendation policy based on the learned embeddings and user feedback.

4.4.3. Hybrid RL-based recommender systems

Hybrid RL-based recommender systems combine multiple RL algorithms or integrate RL with other machine learning techniques to achieve better performance and robustness. Zhao et al. [48] proposed a hybrid RL framework for e-commerce recommendations, which combines model-free and model-based RL methods to balance the trade-off between sample efficiency and generalization ability. Aboutorab

et al. [49] developed a hybrid RL-supervised learning approach for news recommendation, which uses supervised learning to pre-train the recommendation model and RL to fine-tune the model based on user feedback.

The integration of RL with other recommendation techniques and advanced machine learning methods has shown promising results in improving the performance, scalability, and interpretability of recommender systems. By leveraging the strengths of different approaches, hybrid RL-based recommenders can provide more accurate, diverse, and explainable recommendations, while adapting to the dynamic and complex nature of user preferences and item relationships.

In summary, RL has emerged as a powerful framework for building intelligent and adaptive recommender systems. By formulating the recommendation problem as an MDP and using RL algorithms to optimize the recommendation policy, RL-based recommenders can effectively handle the sequential user-system interactions, optimize for long-term user engagement, and deal with sparse feedback and exploration. The integration of RL with traditional recommendation methods, knowledge graphs, graph neural networks, and other machine learning techniques has further enhanced the capabilities and performance of RL-based recommenders. In the next section, we discuss the current challenges and future research directions in this rapidly evolving field.

5. Challenges and future research directions

Despite the significant progress and promising results of reinforcement learning (RL) in recommender systems, there are still several challenges and open research questions that need to be addressed to fully realize the potential of RL-based recommenders. In this section, we discuss the key challenges and future research directions in this field.

5.1. Offline reinforcement learning for recommendations

One of the main challenges in applying RL to real-world recommender systems is the need for online interactions with users, which can be costly and risky. Offline RL [50] aims to learn the optimal recommendation policy from historical user interaction data, without the need for online exploration. However, offline RL suffers from the distribution shift problem, where the learned policy may not generalize well to the actual user behavior. Recent works [38, 51] have proposed off-policy evaluation and correction methods to address this issue, but more research is needed to develop robust and scalable offline RL algorithms for recommendations.

5.2. Scalability and computational efficiency

Another challenge in RL-based recommenders is the scalability and computational efficiency of the algorithms, especially when dealing with large-scale user-item interactions and high-dimensional state and action spaces. Existing works have proposed various approaches to improve the scalability of RL-based recommenders, such as using deep neural networks for function approximation [34], leveraging parallelization and distributed computing [47], and adopting efficient exploration strategies [52]. However, there is still a need for more research on developing scalable and efficient RL algorithms that can handle the ever-growing scale and complexity of real-world recommendation scenarios.

5.3. Explainability and interpretability of RL-based recommendations

Explainability and interpretability are crucial factors in building trust and transparency in recommender systems. However, RL-based recommenders, especially those using deep neural networks, are often considered as black-box models that lack clear explanations for their recommendations. Recent works have proposed various approaches to improve the explainability of RL-based recommenders, such as using attention mechanisms to highlight the important features [38], generating textual explanations based on the learned policy [52], and incorporating knowledge graphs to provide semantic explanations

[46]. However, more research is needed to develop effective and user-friendly explanation methods that can help users understand and trust the recommendations generated by RL-based systems.

5.4. Robustness to adversarial attacks and biases

RL-based recommenders, like other machine learning models, are vulnerable to adversarial attacks and biases that can manipulate or degrade their performance. Adversarial attacks, such as fake user profiles or item reviews, can mislead the RL algorithms and generate suboptimal or even harmful recommendations [53]. Biases, such as popularity bias or selection bias, can also affect the fairness and diversity of the recommendations [44]. Therefore, it is important to develop robust and unbiased RL algorithms that can detect and mitigate the impact of adversarial attacks and biases. Recent works have proposed various approaches, such as adversarial training [48], counterfactual learning [54], and fairness-aware RL [44], but more research is needed to ensure the security and fairness of RL-based recommenders.

5.5. Evaluation metrics and simulation environments for RL-based recommenders

Evaluating the performance of RL-based recommenders is challenging due to the complex and dynamic nature of user-system interactions. Traditional evaluation metrics, such as accuracy and F1 score, may not fully capture the long-term user satisfaction and engagement. Therefore, it is important to develop new evaluation metrics and frameworks that can assess the effectiveness of RL-based recommenders from multiple perspectives, such as user experience, diversity, novelty, and business objectives [18]. Moreover, building realistic and standardized simulation environments for RL-based recommenders can facilitate the development and comparison of different algorithms [55]. Recent works have proposed various simulation environments, such as RecoGym [56] and VirtualTaobao [55], but more research is needed to improve their fidelity and generalizability.

5.6. Real-world applications and case studies

To fully demonstrate the potential and impact of RL-based recommenders, it is important to conduct more real-world applications and case studies in various domains, such as e-commerce, news, music, and video recommendations. Real-world applications can provide valuable insights into the practical challenges and opportunities of deploying RL-based recommenders, such as the need for online learning, the importance of user feedback and explainability, and the trade-off between exploration and exploitation [11, 17]. Case studies can also help to showcase the business value and user benefits of RL-based recommenders, such as increased user engagement, revenue, and customer satisfaction [46, 53]. More research and collaboration between academia and industry are needed to bridge the gap between theory and practice and accelerate the adoption of RL-based recommenders in real-world settings.

RL-based recommenders have shown great promise in improving the performance and adaptability of recommendation systems. However, there are still many challenges and open research questions that need to be addressed, such as offline RL, scalability, explainability, robustness, evaluation, and real-world applications.

6. Conclusion

In this survey, we have provided a comprehensive overview of the state-of-the-art research on reinforcement learning-based recommender systems. We have discussed the limitations of traditional recommendation approaches and highlighted the potential of reinforcement learning in addressing these challenges. We have reviewed the formulation of the recommendation problem as a Markov Decision Process and explored the various reinforcement learning algorithms and architectures used in recommender systems, including model-free, model-based, policy gradient, and deep reinforcement learning methods.

Moreover, we have examined the integration of reinforcement learning with other techniques, such as collaborative filtering, content-based methods, knowledge graphs, and graph neural networks, to enhance the performance and capabilities of recommender systems. We have also identified key challenges and future research directions in this field, including offline reinforcement learning, scalability, explainability, robustness, evaluation metrics, and real-world applications.

The survey highlights the significant progress and promising results achieved by reinforcement learning-based recommenders in various domains, such as e-commerce, news, music, and video recommendations. The ability of reinforcement learning to handle sequential user-system interactions, optimize long-term user engagement, and adapt to dynamic user preferences has made it a powerful framework for building intelligent and personalized recommendation systems.

However, there are still many open challenges and opportunities for future research in this rapidly evolving field. By addressing these challenges and exploring new research directions, we can unlock the full potential of reinforcement learning-based recommenders and create more engaging, diverse, and trustworthy recommendation experiences for users.

Declaration on Generative AI: During the preparation of this work, the authors used Claude 3 Opus in order to: Drafting content, Generate literature review, Abstract drafting. After using this service, the authors reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

[1] G. Adomavicius, A. Tuzhilin, Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, IEEE Trans. Knowl. Data Eng. 17 (2005) 734–749. doi:10.1109/TKDE.2005.99.

[2] J. B. Schafer, J. Konstan, J. Riedl, Recommender systems in e-commerce, in: Proceedings of the 1st ACM Conference on Electronic Commerce, EC '99, Association for Computing Machinery, New York, NY, USA, 1999, p. 158–166. doi:10.1145/336992.337035.

[3] P. Covington, J. Adams, E. Sargin, Deep Neural Networks for YouTube Recommendations, in: Proceedings of the 10th ACM Conference on Recommender Systems, RecSys '16, Association for Computing Machinery, New York, NY, USA, 2016, p. 191–198. doi:10.1145/2959100.2959190.

[4] G. Gupta, R. Katarya, A Study of Deep Reinforcement Learning Based Recommender Systems, in: ICSCCC 2021 - International Conference on Secure Cyber Computing and Communications, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 218–220. doi:10.1109/ICSCCC51823.2021.9478178.

[5] A. O. Afolabi, P. J. Toivanen, K. Haataja, J. Mykkänen, Systematic Literature Review on Empirical Results and Practical Implementations of Healthcare Recommender Systems: Lessons Learned and a Novel Proposal, Int. J. Heal. Inf. Syst. Informatics 10 (2015) 1–21. doi:10.4018/IJHISI.2015100101.

[6] Y. Shi, M. Larson, A. Hanjalic, Collaborative Filtering beyond the User-Item Matrix: A Survey of the State of the Art and Future Challenges, ACM Comput. Surv. 47 (2014) 3. doi:10.1145/2556270.

[7] J. S. Breese, D. Heckerman, C. Kadie, Empirical analysis of predictive algorithms for collaborative filtering, in: Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, UAI'98, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1998, p. 43–52. URL: https://dl.acm.org/doi/10.5555/2074094.2074100.

[8] D. Billsus, M. J. Pazzani, Learning Collaborative Information Filters, in: Proceedings of the Fifteenth International Conference on Machine Learning, ICML '98, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1998, p. 46–54. URL: https://dl.acm.org/doi/10.5555/645527.657311.

[9] P. Basile, C. Greco, A. Suglia, G. Semeraro, S. Ferilli, F. A. Lisi, Deep Learning and Hierarchical Reinforcement Learning for modeling a Conversational Recommender System, Intelligenza Artificiale 12 (2019) 125–141. doi:10.3233/IA-170031.

[10] M. Rezaei, N. Tabrizi, A survey on reinforcement learning and deep reinforcement learning for recommender systems, in: D. Conte, A. Fred, O. Gusikhin, C. Sansone (Eds.), Deep Learning

Theory and Applications, volume 1875 of *Communications in Computer and Information Science*, Springer Nature Switzerland, Cham, 2023, pp. 385–402. doi:10.1007/978-3-031-39059-3_26.

[11] J. Zhao, H. Li, L. Qu, Q. Zhang, Q. Sun, H. Huo, M. Gong, DCFGAN: An adversarial deep reinforcement learning framework with improved negative sampling for session-based recommender systems, Information Sciences 596 (2022) 222–235. doi:10.1016/j.ins.2022.02.045.

[12] R. S. Sutton, A. G. Barto, Reinforcement Learning: An Introduction, 2 ed., MIT Press, 2015. URL: https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf.

[13] M. M. Afsar, T. Crump, B. Far, Reinforcement Learning based Recommender Systems: A Survey, ACM Computing Surveys 55 (2023) 145. doi:10.1145/3543846.

[14] X. Chen, L. Yao, J. McAuley, G. Zhou, X. Wang, Deep reinforcement learning in recommender systems: A survey and new perspectives, Knowledge-Based Systems 264 (2023). doi:10.1016/j.knosys.2023.110335.

[15] L. Zou, J. Song, L. Xia, W. Liu, Z. Ding, D. Yin, Reinforcement learning to optimize long-term user engagement in recommender systems, in: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, 2019, pp. 2810–2818. doi:10.1145/3292500.3330668.

[16] I. Munemasa, Y. Tomomatsu, K. Hayashi, T. Takagi, Deep reinforcement learning for recommender systems, in: 2018 International Conference on Information and Communications Technology, ICOIACT 2018, volume 2018-January, Institute of Electrical and Electronics Engineers Inc., 2018, pp. 226–233. doi:10.1109/ICOIACT.2018.8350761.

[17] S. Xia, P. Wei, Y. Liu, A. Sonta, X. Jiang, A multi-task deep reinforcement learning-based recommender system for co-optimizing energy, comfort, and air quality in commercial buildings with humans-in-the-loop, Data-Centric Engineering 5 (2024) e26. doi:10.1017/dce.2024.27.

[18] X. Xin, A. Karatzoglou, I. Arapakis, J. M. Jose, Self-Supervised Reinforcement Learning for Recommender Systems, in: SIGIR 2020 - Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Association for Computing Machinery, Inc, 2020, pp. 931–940. doi:10.1145/3397271.3401147.

[19] Z. Zhao, X. Chen, Z. Xu, L. Cao, Tag-Aware Recommender System Based on Deep Reinforcement Learning, Mathematical Problems in Engineering 2021 (2021) 5564234. doi:10.1155/2021/5564234.

[20] A. B. A. Alwahhab, Proposed Recommender System for Solving Cold Start Issue Using k-means Clustering and Reinforcement Learning Agent, in: Proceedings - 2020 2nd Annual International Conference on Information and Sciences, AiCIS 2020, Institute of Electrical and Electronics Engineers Inc., 2020, pp. 13–21. doi:10.1109/AiCIS51645.2020.00013.

[21] A. Grishanov, A. Ianina, K. Vorontsov, Multiobjective Evaluation of Reinforcement Learning Based Recommender Systems, in: RecSys 2022 - Proceedings of the 16th ACM Conference on Recommender Systems, Association for Computing Machinery, Inc, 2022, pp. 622–627. doi:10.1145/3523227.3551485.

[22] A. Iftikhar, M. A. Ghazanfar, M. Ayub, S. Ali Alahmari, N. Qazi, J. Wall, A reinforcement learning recommender system using bi-clustering and Markov Decision Process, Expert Systems with Applications 237 (2024) 121541. doi:10.1016/j.eswa.2023.121541.

[23] R. Sun, J. Yan, F. Ren, A Knowledge Graph-based Interactive Recommender System Using Reinforcement Learning, in: Proceedings - 2022 10th International Conference on Advanced Cloud and Big Data, CBD 2022, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 73–78. doi:10.1109/CBD58033.2022.00022.

[24] S. Zhou, X. Dai, H. Chen, W. Zhang, K. Ren, R. Tang, X. He, Y. Yu, Interactive Recommender System via Knowledge Graph-enhanced Reinforcement Learning, in: SIGIR 2020 - Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Association for Computing Machinery, Inc, 2020, pp. 179–188. doi:10.1145/3397271.3401174.

[25] A. Sharifbaev, M. Mozikov, H. Zaynidinov, I. Makarov, Efficient Integration of Reinforcement Learning in Graph Neural Networks-Based Recommender Systems, IEEE Access 12 (2024) 189439–189448. doi:10.1109/ACCESS.2024.3516517.

[26] B. Sarwar, G. Karypis, J. Konstan, J. Riedl, Item-based collaborative filtering recommendation algorithms, in: Proceedings of the 10th International Conference on World Wide Web, WWW '01, Association for Computing Machinery, New York, NY, USA, 2001, p. 285–295. doi:10.1145/371920.372071.

[27] Y. Koren, R. M. Bell, C. Volinsky, Matrix Factorization Techniques for Recommender Systems, Computer 42 (2009) 30–37. doi:10.1109/MC.2009.263.

[28] P. Lops, M. de Gemmis, G. Semeraro, Content-based Recommender Systems: State of the Art and Trends, in: F. Ricci, L. Rokach, B. Shapira, P. B. Kantor (Eds.), Recommender Systems Handbook, Springer US, Boston, MA, 2011, pp. 73–105. doi:10.1007/978-0-387-85820-3_3.

[29] F. Goossen, W. IJntema, F. Frasincar, F. Hogenboom, U. Kaymak, News personalization using the CF-IDF semantic recommender, in: Proceedings of the International Conference on Web Intelligence, Mining and Semantics, WIMS '11, Association for Computing Machinery, New York, NY, USA, 2011, p. 10. doi:10.1145/1988688.1988701.

[30] R. Burke, Hybrid Recommender Systems: Survey and Experiments, User Modeling and User-Adapted Interaction 12 (2002) 331–370. doi:10.1023/A:1021240730564.

[31] H. Jafarkarimi, A. T. H. Sim, R. Saadatdoost, A Naïve Recommendation Model for Large Databases, International Journal of Information and Education Technology 2 (2012) 216–219. URL: https://www.ijiet.org/show-31-222-1.html.

[32] M. A. Ghazanfar, A. Prügel-Bennett, Building Switching Hybrid Recommender System Using Machine Learning Classifiers and Collaborative Filtering, IAENG International Journal of Computer Science 37 (2010) IJCS_37_3_09. URL: https://www.iaeng.org/IJCS/issues_v37/issue_3/IJCS_37_3_09.pdf.

[33] B. Lika, K. Kolomvatsos, S. Hadjiefthymiades, Facing the cold start problem in recommender systems, Expert Systems with Applications 41 (2014) 2065–2073. doi:10.1016/j.eswa.2013.09.005.

[34] G. Zheng, F. Zhang, Z. Zheng, Y. Xiang, N. J. Yuan, X. Xie, Z. Li, DRN: A Deep Reinforcement Learning Framework for News Recommendation, in: Proceedings of the 2018 World Wide Web Conference, WWW '18, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 2018, p. 167–176. doi:10.1145/3178876.3185994.

[35] R. Warlop, A. Lazaric, J. Mary, Fighting Boredom in Recommender Systems with Linear Reinforcement Learning, in: S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett (Eds.), Advances in Neural Information Processing Systems, volume 31, Curran Associates, Inc., 2018, pp. 1757–1768. URL: https://proceedings.neurips.cc/paper_files/paper/2018/file/210f760a89db30aa72ca258a3483cc7f-Paper.pdf.

[36] C. J. C. H. Watkins, P. Dayan, Q-learning, Machine Learning 8 (1992) 279–292. doi:10.1007/BF00992698.

[37] G. A. Rummery, M. Niranjan, On-line Q-learning using connectionist systems, CUED/F-INFENG TR 166, Cambridge University Engineering Department, Cambridge, England, 1994. URL: http://mi.eng.cam.ac.uk/reports/svr-ftp/auto-pdf/rummery_tr166.pdf.

[38] H. Chen, X. Dai, H. Cai, W. Zhang, X. Wang, R. Tang, Y. Zhang, Y. Yu, Large-scale interactive recommendation with tree-structured policy gradient, in: Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, AAAI'19/IAAI'19/EAAI'19, AAAI Press, 2019, p. 407. doi:10.1609/aaai.v33i01.33013312.

[39] A. Gunawardana, C. Meek, A unified approach to building hybrid recommender systems, in: Proceedings of the Third ACM Conference on Recommender Systems, RecSys '09, Association for Computing Machinery, New York, NY, USA, 2009, p. 117–124. doi:10.1145/1639714.1639735.

[40] R. J. Williams, Simple statistical gradient-following algorithms for connectionist reinforcement learning, Machine Learning 8 (1992) 229–256. doi:10.1007/BF00992696.

[41] V. Konda, J. Tsitsiklis, Actor-Critic Algorithms 12 (1999). URL: https://proceedings.neurips.cc/paper_files/paper/1999/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.

[42] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller,

A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, D. Hassabis, Human-level control through deep reinforcement learning, Nature 518 (2015) 529–533. doi:10.1038/nature14236.

[43] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, D. Wierstra, Continuous control with deep reinforcement learning, 2019. URL: https://arxiv.org/abs/1509.02971. arXiv:1509.02971.

[44] J. Chen, H. Dong, X. Wang, F. Feng, M. Wang, X. He, Bias and Debias in Recommender System: A Survey and Future Directions, ACM Trans. Inf. Syst. 41 (2023) 67. doi:10.1145/3564284.

[45] S. Choi, H. Ha, U. Hwang, C. Kim, J.-W. Ha, S. Yoon, Reinforcement Learning based Recommender System using Biclustering Technique, 2018. URL: https://arxiv.org/abs/1801.05532. arXiv:1801.05532.

[46] J. Zhang, B. Hao, B. Chen, C. Li, H. Chen, J. Sun, Hierarchical Reinforcement Learning for Course Recommendation in MOOCs, Proceedings of the AAAI Conference on Artificial Intelligence 33 (2019) 435–442. doi:10.1609/aaai.v33i01.3301435.

[47] H. Liu, K. Cai, P. Li, C. Qian, P. Zhao, X. Wu, REDRL: A review-enhanced Deep Reinforcement Learning model for interactive recommendation, Expert Systems with Applications 213 (2023) 118926. doi:10.1016/j.eswa.2022.118926.

[48] X. Zhao, L. Xia, L. Zou, H. Liu, D. Yin, J. Tang, Whole-Chain Recommendations, in: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, CIKM '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 1883–1891. doi:10.1145/3340531.3412044.

[49] H. Aboutorab, O. K. Hussain, M. Saberi, F. K. Hussain, D. D. Prior, Reinforcement Learning-Based News Recommendation System, IEEE Trans. Serv. Comput. 16 (2023) 4493–4502. doi:10.1109/TSC.2023.3326197.

[50] S. Levine, A. Kumar, G. Tucker, J. Fu, Offline Reinforcement Learning: Tutorial, Review, and Perspectives on Open Problems, 2020. URL: https://arxiv.org/abs/2005.01643. arXiv:2005.01643.

[51] A. Calero Valdez, M. Ziefle, K. Verbert, A. Felfernig, A. Holzinger, Recommender systems for health informatics: State-of-the-art and future perspectives, in: A. Holzinger (Ed.), Machine Learning for Health Informatics: State-of-the-Art and Future Challenges, volume 9605 of *Lecture Notes in Computer Science*, Springer International Publishing, Cham, 2016, pp. 391–414. doi:10.1007/978-3-319-50478-0_20.

[52] J. Yu, M. Gao, H. Yin, J. Li, C. Gao, Q. Wang, Generating Reliable Friends via Adversarial Training to Improve Social Recommendation, in: 2019 IEEE International Conference on Data Mining (ICDM), IEEE Computer Society, Los Alamitos, CA, USA, 2019, pp. 768–777. doi:10.1109/ICDM.2019.00087.

[53] S. Wang, Y. Cao, X. Chen, L. Yao, X. Wang, Q. Z. Sheng, Adversarial Robustness of Deep Reinforcement Learning Based Dynamic Recommender Systems, Frontiers in Big Data 5 (2022). doi:10.3389/fdata.2022.822783.

[54] T. Liu, K. Yu, L. Wang, X. Zhang, H. Zhou, X. Wu, Clickbait detection on WeChat: A deep model integrating semantic and syntactic information, Knowledge-Based Systems 245 (2022) 108605. doi:10.1016/j.knosys.2022.108605.

[55] B. Shi, M. G. Ozsoy, N. Hurley, B. Smyth, E. Z. Tragos, J. Geraci, A. Lawlor, PyrecGym: A reinforcement learning gym for recommender systems, in: RecSys 2019 - 13th ACM Conference on Recommender Systems, Association for Computing Machinery, Inc, 2019, pp. 491–495. doi:10.1145/3298689.3346981.

[56] D. Rohde, S. Bonner, T. Dunlop, F. Vasile, A. Karatzoglou, RecoGym: A Reinforcement Learning Environment for the problem of Product Recommendation in Online Advertising, 2018. URL: https://arxiv.org/abs/1808.00720. arXiv:1808.00720.