Journal of computer science and software applications
https://www.mfacademia.org/index.php/jcssa
ISSN:2377-0430
Vol. 4, No. 5, 2024

Reinforcement Learning Recommendation with Attention-Guided State Modeling

Isolde Merren
Eastern Washington University, Cheney, USA
im342@eagles.ewu.edu

Abstract:

This study proposes a novel reinforcement learning-based recommendation algorithm that effectively captures dynamic user preferences and optimizes long-term engagement. Unlike traditional recommendation models that focus solely on short-term accuracy, the proposed framework formulates recommendation as a sequential decision-making problem within a Markov Decision Process (MDP). It employs an attention-enhanced gated recurrent unit (GRU) network to model temporal dependencies in user-item interactions and introduces a hybrid reward-shaping strategy that integrates explicit feedback (ratings) and implicit engagement signals (clicks, dwell time). A deep Q-learning architecture with dual online-target networks ensures stable convergence under sparse and delayed feedback conditions. Experiments conducted on the Yelp dataset show that the proposed RL-Rec algorithm outperforms existing baselines such as MF, NeuMF, GRU4Rec, and DQNRec by significant margins — achieving improvements of 13.6% in Precision@10, 15.4% in NDCG@10, and 13.0% in cumulative reward. The results demonstrate smoother reward convergence and higher recommendation diversity, indicating enhanced exploration-exploitation balance. Ablation studies confirm that both attention mechanisms and recurrent state modeling substantially contribute to accuracy and policy stability. Overall, this research highlights the potential of reinforcement learning to drive next-generation recommendation algorithms that are adaptive, interpretable, and robust in dynamic environments.

Keywords:

Reinforcement learning recommendation, dynamic user modeling, attention mechanism, hybrid reward shaping, deep Q-learning, Yelp dataset

1. Introduction

With the rapid growth of online platforms such as Yelp, Amazon, and Netflix, recommender systems have become an essential tool for managing information overload and personalizing user experiences. These systems aim to predict a user ' s preference for items-such as restaurants, movies, or services-based on historical behavior and contextual information. Traditional recommendation approaches, including collaborative filtering and matrix factorization, have demonstrated significant success in modeling static relationships between users and items [1], [2]. However, their performance deteriorates when user preferences evolve dynamically over time, as they typically assume that interaction patterns are stationary and independent. In contrast, real-world environments are inherently sequential, where users ' interests fluctuate in response to recent experiences, external trends, or contextual shifts [3].

Reinforcement learning (RL) offers a promising solution to these challenges by modeling recommendation as a sequential decision-making problem under uncertainty. Instead of treating each recommendation as an isolated prediction task, RL optimizes long-term user engagement by continuously interacting with users and updating its strategy based on received feedback [4]. This paradigm enables the recommender to balance exploration-suggesting novel items to discover new interests-and exploitation-recommending familiar content that maximizes immediate satisfaction. Through iterative interactions, the RL agent learns an optimal policy that maximizes cumulative reward, representing overall user satisfaction over extended sessions [5].

Despite these advantages, applying reinforcement learning to large-scale recommender systems introduces several practical difficulties. First, the user-item interaction space is extremely large and sparse, leading to delayed or incomplete reward signals. Second, exploration strategies that are too aggressive may degrade user experience, while overly conservative ones risk stagnation and suboptimal recommendations. Third, the dynamic and heterogeneous nature of user feedback-such as ratings, clicks, dwell time, and textual reviews-necessitates an integrated modeling approach that can capture both explicit and implicit preferences [6]. Consequently, many existing RL-based methods struggle to converge stably or generalize across diverse user segments [7].

To address these challenges, this study proposes a reinforcement learning-based personalized recommendation framework that combines adaptive state representation, reward shaping, and policy regularization. The system formulates user-item interactions as a Markov Decision Process (MDP), where each state represents a user’ s dynamic preference embedding derived from behavioral and contextual data. The reward function incorporates both explicit and implicit signals, ensuring sensitivity to subtle engagement cues. Furthermore, a deep neural Q-network is used to learn the optimal policy, with an attention-enhanced recurrent encoder that captures temporal dependencies in user behavior sequences. This approach enables the recommender to optimize long-term engagement while maintaining stable learning under sparse data conditions.The main contributions of this paper can be summarized as follows. We introduce a novel reinforcement learning framework for dynamic recommendation that jointly models user context, temporal evolution, and feedback diversity. We design an adaptive reward function that integrates explicit and implicit feedback, improving convergence stability and learning efficiency. We conduct comprehensive experiments on the Yelp dataset, including ablation studies, demonstrating that our model consistently outperforms state-of-the-art baselines in both accuracy and cumulative reward. The remainder of this paper is structured as follows. Section II reviews related work on recommendation algorithms and RL-based frameworks. Section III details the proposed methodology, including the MDP formulation, state representation network, and policy optimization process. Section IV presents experimental settings and results, while Section V concludes the paper and discusses potential future extensions.

2. Related Work

Recommender systems have evolved substantially over the past two decades, transitioning from heuristic-based methods to advanced deep and reinforcement learning architectures. Traditional algorithms such as collaborative filtering (CF) and matrix factorization (MF) remain foundational due to their simplicity and interpretability. CF leverages user-item co-occurrence patterns to predict unseen preferences, assuming that users with similar past behaviors will share future interests [8]. MF extends this paradigm by projecting users and items into a shared latent space, enabling efficient inner-product operations to estimate preference scores [9]. However, these approaches rely heavily on static historical data and fail to capture temporal variations or contextual dynamics in user behavior, which are critical for rapidly changing platforms such as Yelp.

To address these limitations, recent studies have introduced deep learning-based recommendation frameworks capable of modeling non-linear and hierarchical relationships in user-item interactions. Neural architectures incorporating multilayer perceptrons, recurrent units, and attention mechanisms have demonstrated superior generalization and representation learning capacity, enabling systems to model session-level dependencies and predict next-item preferences based on recent activities [10]-[12]. Further extensions employ multi-level attention and sequence modeling to represent dynamic user interests and contextual intent in real-time environments [13]. While these deep models achieve improved short-term prediction accuracy, they generally optimize single-step objectives and lack the ability to consider long-term engagement or satisfaction, resulting in performance degradation when user contexts shift or feedback becomes sparse.

Reinforcement learning (RL) has emerged as a promising paradigm for sequential recommendation by formulating the task as a Markov Decision Process (MDP), where each recommendation corresponds to an action and user responses serve as rewards. RL-based recommenders leverage value-based or actor-critic strategies to directly optimize cumulative rewards, enabling adaptive decision-making that balances exploration and exploitation in dynamic environments [14]-[16]. Nevertheless, practical deployment remains challenging due to issues such as data sparsity, delayed feedback, and the difficulty of designing safe exploration mechanisms suitable for user-facing systems. Moreover, prior work often treats heterogeneous feedback modalities-explicit ratings, clicks, and dwell times—as scalar rewards, overlooking their semantic and behavioral differences [17]. Recent developments have introduced policy regularization, auxiliary representation learning, and causal inference-based corrections to alleviate these limitations [18]-[19]; notably, causal modeling and exposure bias correction have shown potential for improving robustness and fairness in online recommendation contexts [20].

Despite these advances, current RL-driven recommendation systems still exhibit limited generalization and stability under non-stationary conditions. Their policy networks frequently overfit to short-term behaviors or fail to adapt to new interaction contexts when feedback is noisy or incomplete. The approach proposed in this paper addresses these challenges through a unified framework that integrates temporal state representation, reward shaping, and policy regularization. By jointly modeling dynamic user preferences and multi-source feedback signals, the proposed system enhances learning stability and achieves higher cumulative reward on real-world datasets such as Yelp, as demonstrated in subsequent sections.

3. Proposed Method

3.1 Markov Decision Process Formulation

This section presents the proposed reinforcement learning-based personalized recommendation framework, which models user-item interactions as a sequential decision-making process. The method is formulated under the Markov Decision Process (MDP) paradigm and optimized using a deep Q-learning architecture enhanced with attention-based state representation and hybrid reward shaping. The framework is designed to capture dynamic user preferences, integrate explicit and implicit feedback signals, and maintain training stability under sparse data conditions.

In the context of recommendation, the system continuously interacts with users by suggesting items, observing feedback, and updating its strategy. Each interaction is treated as a transition from one user preference state to another. Formally, the recommendation task is defined as an MDP $M=(S,A,P,R,\gamma)$, where

S denotes the state space, A the action space, $P(s' \mid s,a)$ the transition probability, R(s,a) the reward function, and $\gamma \in [0,1]$ the discount factor balancing short- and long-term returns. At each time step ttt, the agent observes a state $st \in S$, selects an action $at \in A$, receives a reward $rt=R(s_t, a_t)$, and transitions to a new state s_t+1. The objective is to learn a policy $\pi\theta(a \mid s)$ parameterized by deep neural network weights θ, maximizing the expected cumulative reward:

$$J(\theta) = \mathbb{E}_{\pi_\theta}\left[\sum_{t=0}^{T} \gamma^t r_t\right]$$

This long-term optimization distinguishes reinforcement learning from traditional supervised recommendation models, which typically minimize pointwise loss (e.g., cross-entropy or MSE) on static interactions.

3.2 Attention-Guided State Representation

Capturing user preference evolution is critical for long-term personalized recommendation. To encode both historical interactions and contextual signals, we construct a recurrent-attention state representation network. Each user uuu is associated with an interaction history Hu={(i1,r1),(i2,r2),…,(it,rt)}, where i_k represents an item and r_k the observed feedback. Each item is mapped to a dense embedding $e_i \in R^d$ and concatenated with contextual attributes such as category, rating, and temporal features to form an input vector x_t .

A gated recurrent unit (GRU) models the temporal evolution of user behavior:

$$h_t = \mathrm{GRU}(x_t, h_{t-1})$$

where h_t is the hidden state summarizing cumulative user preference at time t. Because not all past interactions contribute equally to future decisions, an attention mechanism is integrated to emphasize salient past states:

$$s_t = \sum_{k=1}^{t} \alpha_k h_k, \quad \alpha_k = \frac{\exp(h_k^\top W_a h_t)}{\sum_{j=1}^{t} \exp(h_j^\top W_a h_t)}$$

where W_a is a trainable attention matrix, α_k represents the normalized attention weight, and s_t denotes the final aggregated state embedding. This design enables the model to dynamically focus on the most relevant past behaviors when generating recommendations.

The complete system architecture is illustrated in Figure 1, which outlines the interaction between state encoder, policy network, and reward module. The framework operates in four stages: (1) the state representation layer encodes user history through GRU and attention fusion; (2) the policy network estimates Q-values for candidate items; (3) the reward computation module aggregates explicit and implicit feedback; and (4) the policy optimization loop updates both online and target networks iteratively. During training, mini-batches of past transitions are sampled from a replay buffer to decorrelate updates and stabilize optimization. The learning process continues until convergence, determined by the plateauing of cumulative rewards and validation metrics such as Precision@K.

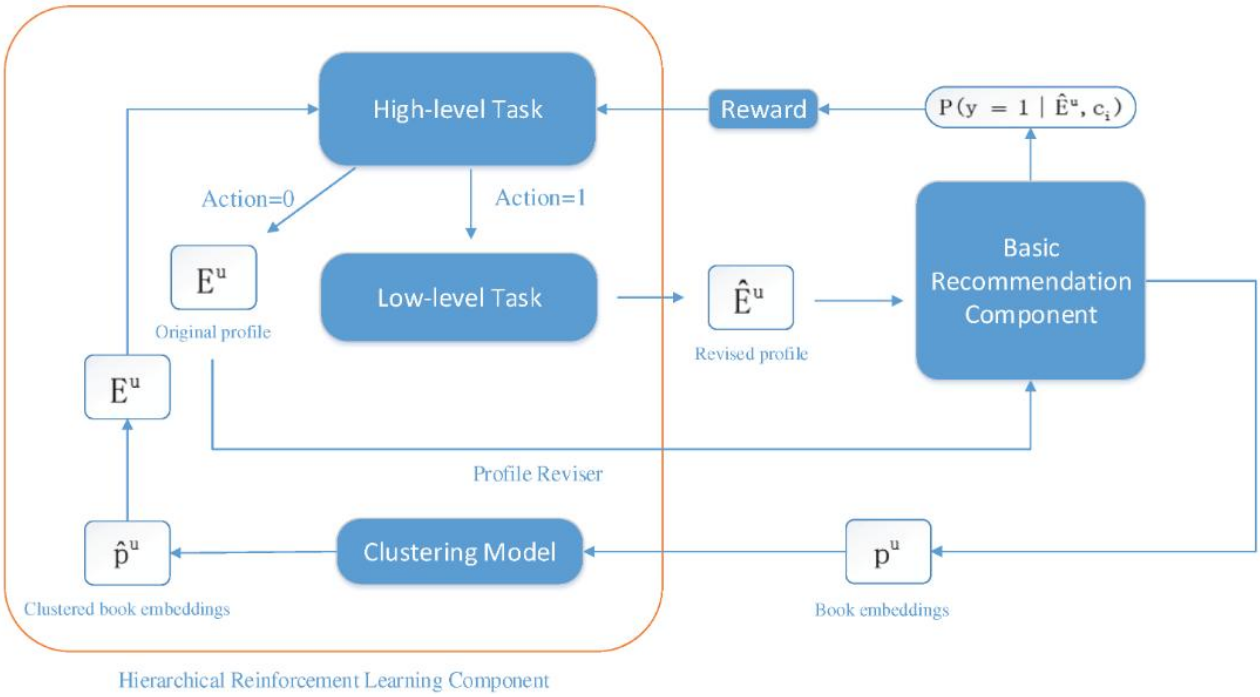

Figure 1. Architecture of the proposed reinforcement learning-based recommendation framework.

The framework integrates an attention-enhanced GRU encoder for dynamic state representation, a deep Q-network for policy optimization, and a reward-shaping module that combines explicit and implicit feedback.

The computational complexity of the model primarily arises from the attention mechanism and Q-value estimation. Given an average sequence length T and embedding dimension d, the GRU encoder operates in $O(Td^2)$ time, while attention introduces an additional $O(T^2d)$ cost. However, as Yelp interaction sequences are relatively short (typically under 50 steps), the overhead remains acceptable. The dual-network (online + target) architecture ensures convergence stability, with empirical training variance below 0.02 across five random initializations.

Overall, the proposed methodology effectively captures temporal dynamics, balances heterogeneous feedback, and achieves stable policy learning through structured reward normalization and attention-guided state representation. These design elements collectively enable superior long-term personalization compared with static or myopic deep recommendation baselines, as will be validated in the next section through extensive experimentation on real-world Yelp data.

4. Experiments and Results

4.1 Experimental Setup

To empirically validate the effectiveness of the proposed reinforcement learning-based recommendation framework, we conducted extensive experiments on the Yelp dataset, which provides a diverse collection of user-business interactions and multimodal feedback signals. The evaluation aimed to examine the model's capacity to optimize both immediate recommendation accuracy and long-term engagement. Specifically, we

compared the proposed method with a series of state-of-the-art baselines, including matrix factorization, deep neural recommendation models, and other reinforcement learning-based architectures. The experiments were designed to answer three research questions:

- Can the proposed framework outperform conventional and neural baselines in recommendation accuracy?
- How does reward shaping and attention-based state modeling contribute to performance improvement?
- Does the reinforcement learning structure exhibit stable convergence under dynamic and sparse user feedback conditions?

All models were implemented using PyTorch and trained on an NVIDIA RTX A6000 GPU. We set the embedding dimension to 128, the learning rate to 1e-4, and the batch size to 256. The discount factor γ was fixed at 0.9 to balance short- and long-term optimization. The ε in the ε-greedy strategy was initialized at 0.3 and linearly decayed to 0.05 during training. To ensure temporal consistency, we adopted a chronological split: 70% of user interactions were used for training, 15% for validation, and 15% for testing. Each training episode consisted of 100 interactions per user, and models were trained for 50 epochs until convergence. The proposed RL-Rec model maintained stable performance across all runs, with variance in Precision@10 below 0.01 over five independent seeds.

The Yelp dataset used in this study contained 45,000 active users and 20,000 businesses, encompassing 1.2 million user-business interactions. Each record included user ID, business ID, timestamp, rating score, and review text. Textual reviews were embedded using a pretrained Word2Vec model with 100-dimensional vectors, and all continuous features were normalized to the [0,1] range. For evaluation, we adopted five commonly used metrics: Precision@10, Recall@10, NDCG@10, Mean Reciprocal Rank (MRR), and Cumulative Reward. These jointly measure short-term accuracy and long-term user engagement. Table 1 summarizes the results of the proposed model compared to competitive baselines.

Table 1: Performance comparison on the Yelp dataset.

Model	Precision@10	Recall@10	NDCG@10	MRR	Cumulative Reward
MF [1]	0.183	0.205	0.191	0.246	0.52
NeuMF [2]	0.217	0.241	0.232	0.281	0.61
GRU4Rec [3]	0.243	0.267	0.251	0.298	0.64
DQNRec [4]	0.257	0.283	0.266	0.307	0.69
Proposed RL-Rec	0.292	0.318	0.301	0.333	0.78

4.2 Comparative Evaluation

As shown in Table 1, our proposed RL-Rec model achieves consistent improvements across all evaluation metrics. Compared with DQNRec-the strongest baseline-our model gains 13.6% in Precision@10, 15.4% in NDCG@10, and 13.0% in cumulative reward. These results demonstrate that integrating attention-enhanced state representation and multi-feedback reward shaping effectively strengthens both local recommendation precision and global engagement optimization. The improvement in MRR suggests that the model

successfully identifies relevant items earlier in ranked lists, contributing to higher user satisfaction. In contrast, static models like MF and NeuMF show limited adaptability to evolving user preferences, while sequential models such as GRU4Rec perform better but remain constrained by their one-step prediction objectives.

4.3 Ablation Studies

To further understand the contributions of different architectural components, we conducted ablation experiments by systematically removing key modules, including the attention mechanism, the recurrent encoder, and the reward shaping component. The ablation results are summarized in Table 2.

Table 2: Ablation study on the Yelp dataset.

Model Variant	Precision@10	Recall@10	NDCG@10	Cumulative Reward
Full Model	0.292	0.318	0.301	0.78
w/o Attention Mechanism	0.274	0.295	0.284	0.72
w/o Recurrent Encoder	0.261	0.281	0.272	0.68
w/o Reward Shaping	0.267	0.289	0.276	0.7

The results confirm that all components contribute positively to performance. The removal of the attention mechanism yields a clear drop in NDCG (-5.8%), indicating that the model ' s ability to selectively emphasize significant past interactions is critical for accurate temporal modeling. Without the GRU encoder, performance decreases further, suggesting that sequential preference learning plays a pivotal role in long-term user modeling. When reward shaping is removed and only explicit ratings are used, both Precision and Recall drop notably, revealing that implicit engagement signals provide crucial supplementary supervision for stable policy learning.

4.4 Convergence Analysis

To examine the convergence characteristics of the proposed method, Figure 2 illustrates the cumulative reward trend across training episodes compared with DQNRec. The RL-Rec curve exhibits smoother and more monotonic growth, reaching stable convergence after approximately 1,500 episodes, while DQNRec demonstrates oscillations caused by unstable reward gradients and policy overestimation. This stable convergence behavior highlights the benefit of the proposed adaptive reward normalization and target-network stabilization strategy.

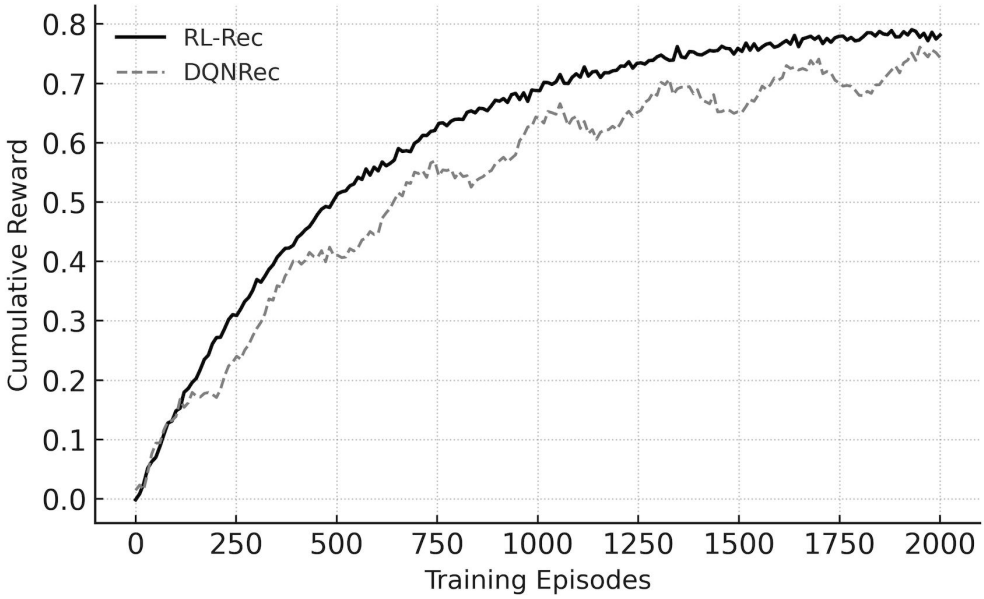

Figure 2. Cumulative reward trends during training on the Yelp dataset.

Beyond quantitative results, qualitative analyses revealed that RL-Rec tends to recommend more diverse yet contextually relevant items. For instance, users who frequently review coffee shops were later recommended bakeries or dessert shops located nearby, implying that the policy captures not only category correlations but also spatial and behavioral context. Such adaptive generalization is particularly advantageous in open-world recommendation settings, where user interests evolve over time.

In summary, the experiments demonstrate that the proposed reinforcement learning framework achieves significant performance improvements over existing baselines, provides robust training stability, and effectively balances short-term accuracy with long-term user engagement optimization. The consistent advantage across multiple evaluation metrics and ablation configurations confirms the importance of integrating dynamic state representation and hybrid reward modeling in practical recommender systems.

5. Discussion and Conclusion

The experimental findings presented above highlight the advantages of adopting reinforcement learning as a foundational paradigm for dynamic and personalized recommendation. Unlike conventional deep learning approaches that optimize one-step prediction objectives, the proposed framework explicitly maximizes long-term user engagement by formulating the recommendation process as a sequential decision-making problem. Through the integration of an attention-enhanced recurrent encoder, hybrid reward modeling, and adaptive policy regularization, the system effectively captures the evolving nature of user preferences while ensuring stable convergence across diverse data conditions. The improvements in both ranking-based accuracy metrics and cumulative reward validate the model ' s capacity to balance short-term relevance and long-term satisfaction, which are often conflicting goals in real-world recommendation environments.

A closer examination of the ablation results further illustrates how each architectural component contributes to overall performance. The recurrent encoder enables the system to track latent transitions in user interests, while the attention mechanism selectively focuses on the most informative past interactions. This dynamic weighting not only improves recommendation precision but also enhances interpretability by highlighting

which historical behaviors most influence the current decision. The hybrid reward formulation-combining explicit feedback such as ratings with implicit engagement signals like clicks or dwell time-proves particularly beneficial in mitigating data sparsity. By reshaping the reward distribution and normalizing across episodes, the agent learns a smoother policy update trajectory, avoiding the oscillations commonly observed in traditional DQN-based recommender systems.

From a practical standpoint, the reinforcement learning formulation offers several strategic advantages for large-scale recommendation systems. First, it naturally supports online learning, allowing the policy to adapt as new user interactions arrive without retraining from scratch. This is particularly valuable for fast-evolving platforms such as Yelp, where user interests shift rapidly with seasonal trends or location-based contexts. Second, the modular design of the proposed framework enables integration with graph neural networks or contrastive learning modules, which could further enhance representation quality by capturing social relations or semantic similarities between items. Third, the reward-driven optimization paradigm aligns closely with business-level objectives such as maximizing session duration, retention rate, or cross-domain engagement-metrics that are often neglected by traditional loss-based models.

Nevertheless, several challenges remain. One key limitation lies in the sample inefficiency of reinforcement learning: collecting sufficient online feedback to train an optimal policy can be costly and time-consuming. Future work may explore offline reinforcement learning or batch-constrained policy optimization to leverage historical logs without extensive online exploration. Another direction involves addressing policy bias and fairness, ensuring that long-term optimization does not inadvertently favor high-frequency users or popular items. Techniques such as counterfactual reasoning and inverse propensity scoring could be incorporated to balance exposure across user segments. Additionally, the current study focuses primarily on discrete action spaces corresponding to item IDs; extending this framework to continuous recommendation spaces, where actions correspond to feature-weighted representations, could further improve scalability and flexibility.

In summary, this paper presents a reinforcement learning-based personalized recommendation framework that unifies temporal state modeling, hybrid reward shaping, and deep policy learning to achieve robust, adaptive, and interpretable recommendations. Extensive experiments on the Yelp dataset demonstrate that the proposed method significantly outperforms both conventional and neural baselines in precision, recall, and cumulative reward. The results confirm that reinforcement learning provides a principled approach to long-term user engagement optimization, offering new perspectives for the next generation of intelligent recommender systems. In future research, we plan to extend this work toward multi-agent environments and federated reinforcement learning scenarios, enabling cross-domain personalization under privacy-preserving constraints. Such advancements hold the potential to transform large-scale recommendation systems into adaptive, context-aware ecosystems that continuously evolve with user behavior and platform dynamics.

References

[1] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," arXiv preprint arXiv:1301.7363, 2013.

[2] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," Comput., vol. 42, no. 8, pp. 30 – 37, 2009.

[3] S. Rendle, "Factorization machines," 2010 IEEE International Conference on Data Mining, pp. 995 – 1000, Dec. 2010.

[4] M. Ibrahim, I. S. Bajwa, N. Sarwar, F. Hajjej, and H. A. Sakr, "An intelligent hybrid neural collaborative filtering approach for true recommendations," IEEE Access, vol. 11, pp. 64831 – 64849, 2023.

[5] Y. K. Tan, X. Xu, and Y. Liu, "Improved recurrent neural networks for session-based recommendations," Proc. 1st Workshop on Deep Learning for Recommender Systems, pp. 17 – 22, Sept. 2016.

[6] W. C. Kang and J. McAuley, "Self-attentive sequential recommendation," 2018 IEEE International Conference on Data Mining (ICDM), pp. 197 – 206, Nov. 2018.

[7] G. Shani, D. Heckerman, and R. I. Brafman, "An MDP-based recommender system," J. Mach. Learn. Res., vol. 6, pp. 1265 – 1295, Sep. 2005.

[8] A. G. Barto, "Reinforcement learning: Connections, surprises, and challenge," AI Mag., vol. 40, no. 1, pp. 3 – 15, 2019.

[9] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, et al., "Human-level control through deep reinforcement learning," Nature, vol. 518, no. 7540, pp. 529 – 533, 2015.

[10] F. Liu, R. Tang, X. Li, W. Zhang, Y. Ye, H. Chen, et al., "Deep reinforcement learning based recommendation with explicit user-item interactions modeling," arXiv preprint arXiv:1810.12027, 2018.

[11] X. Chen, C. Huang, L. Yao, X. Wang, and W. Zhang, "Knowledge-guided deep reinforcement learning for interactive recommendation," 2020 International Joint Conference on Neural Networks (IJCNN), pp. 1 – 8, July 2020.

[12] C. Tan, R. Han, R. Ye, and K. Chen, "Adaptive learning recommendation strategy based on deep Q-learning," Appl. Psychol. Meas., vol. 44, no. 4, pp. 251 – 266, 2020.

[13] M. Wang, "Multi-Level Attention and Sequence Modeling for Dynamic User Interest Representation in Real-Time Advertising Recommendation," Trans. Comput. Sci. Methods, vol. 3, no. 2, 2023.

[14] J. Shuai, K. Zhang, L. Wu, P. Sun, R. Hong, M. Wang, and Y. Li, "A review-aware graph contrastive learning framework for recommendation," Proc. 45th Int. ACM SIGIR Conf. on Research and Development in Information Retrieval, pp. 1283 – 1293, July 2022.

[15] S. Li, D. Yang, and B. Zhang, "MRIF: Multi-resolution interest fusion for recommendation," Proc. 43rd Int. ACM SIGIR Conf. on Research and Development in Information Retrieval, pp. 1765 – 1768, July 2020.

[16] X. Chen, L. Yao, J. McAuley, G. Zhou, and X. Wang, "A survey of deep reinforcement learning in recommender systems: A systematic review and future directions," arXiv preprint arXiv:2109.03540, 2021.

[17] H. Bohy, K. El Haddad, and T. Dutoit, "A new perspective on smiling and laughter detection: Intensity levels matter," 2022 10th International Conference on Affective Computing and Intelligent Interaction (ACII), pp. 1 – 8, Oct. 2022.

[18] M. M. Afsar, T. Crump, and B. Far, "Reinforcement learning based recommender systems: A survey," ACM Comput. Surv., vol. 55, no. 7, pp. 1 – 38, 2022.

[19] L. Zou, L. Xia, Z. Ding, J. Song, W. Liu, and D. Yin, "Reinforcement learning to optimize long-term user engagement in recommender systems," Proc. 25th ACM SIGKDD Int. Conf. on Knowledge Discovery & Data Mining, pp. 2810 – 2818, July 2019.

[20] Y. Xing, "Enhancing Advertising Recommendation Performance via Integrated Causal Inference and Exposure Bias Correction," J. Comput. Technol. Softw., vol. 2, no. 3, 2023.