RESEARCH PAPER

A Context-Aware Deep Reinforcement Learning Framework for Personalized Sequential Recommendations

Manu Y M[1], Meghana M C[2]
[1]Department of CS&E, Associate Professor, BGS Institute of Technology, Adichunchanagiri University, Mandya, India
[2]Department of CS&E, PG Student, BGS Institute of Technology, Adichunchanagiri University, Mandya, India
Corresponding Author:
Meghana M C
Department of CS&E, PG Student, BGS Institute of Technology, Adichunchanagiri University, Mandya, India
Email:ID: meghana_2025@rediffmail.com

ABSTRACT
Modern recommender systems must move beyond static user–item interaction modeling and incorporate contextual signals—such as temporal cues, device information, short-term intent, and situational factors—to deliver highly adaptive and personalized recommendations. Recent advances in sequential modeling and reinforcement learning (RL) have enabled next-generation systems to optimize long-term user engagement instead of relying solely on immediate click-based rewards. However, integrating heterogeneous contextual information into RL policies remains a complex challenge due to non-stationary user behavior, large action spaces, and the difficulty of off-policy learning from logged interactions. This paper proposes a Context-Aware Deep Reinforcement Learning (CA-DRL) framework that models recommendation as a sequential decision-making process enhanced with multi-type contextual fusion. The architecture employs a hierarchical state encoder that captures user profiles, item embeddings, tempora BP-154 (3) l-spatial context, session-level behavior, and auxiliary cues. A list-wise actor–critic agent selects ranked lists of items using a candidate generation module and long-term reward modeling. The system incorporates off-policy corrections, contextual attention, and simulator-based pretraining for stability and scalability.
Keywords: Context-Aware Recommendation Deep Reinforcement Learning Sequential Recommendation Actor–Critic Architecture Long-Term User Engagement
How to cite this article*:* Manu YM, Meghana MC. A Context-Aware Deep Reinforcement Learning Framework for Personalized Sequential Recommendations. Int J Drug Deliv Technol. 2026;16(59s): 1262-1270. DOI: 10.25258/ijddt.16.59s.142
Source of support: Nil.
Conflict of interest: Nil.

INTRODUCTION

Recommender systems have evolved into one of the most critical components underpinning modern digital ecosystems, influencing user experiences across e-commerce platforms, entertainment services, educational portals, social networks, and mobile applications. Their primary objective is to ease the overwhelming burden of information overload by selecting the most relevant items for individual users based on their past interactions and preferences. Traditionally, these systems relied on collaborative filtering, matrix factorization, or supervised deep learning models that used static user–item interaction histories to generate recommendations. While early recommender systems enabled an unprecedented level of personalization compared to manual content curation, real-world user behavior is far more dynamic than what static models can capture [1]. User interests shift depending on context—what a person desires changes throughout the day, across different locations, over devices, and even across emotional or cognitive states. Thus, recommending content solely based on historical data is no longer sufficient in a world where user behavior is shaped by rapidly changing contextual environments. The emergence of context-aware recommendation (CARS) responds to this pressing need by integrating multi-dimensional contextual information—such as time, location, device type, session behavior, environmental conditions, and short-term intent—into the recommendation process. In parallel, Reinforcement Learning (RL) has become a powerful paradigm for sequential decision-making, allowing recommender systems to optimize long-term user engagement instead of merely predicting immediate clicks or ratings. Together, the integration of contextual intelligence and deep reinforcement learning holds the potential to reshape the next generation of personalized user experiences [2].

Despite the progress in both context-aware modeling and reinforcement learning-based recommendation, significant challenges persist in current systems that limit their ability to fully exploit the richness of contextual signals and the sequential nature of user interactions. Most existing recommender systems rely heavily on supervised learning strategies that optimize pointwise or pairwise prediction accuracy, which only captures short-term behaviors. These systems lack the capability to reason about downstream impacts of recommendations, resulting in content that may

*Author for Correspondence: meghana_2025@rediffmail.com

yield an immediate click but fails to sustain long-term engagement, satisfaction, or retention [3]. Furthermore, many models treat user–item interactions as independent events rather than elements of a sequential process influenced by the evolving context around the user. This fundamental assumption leads to suboptimal policies, particularly in high-variability scenarios such as mobile content consumption or session-based browsing patterns.
Another major issue lies in the incomplete or inefficient representation of contextual information. While some models incorporate specific contextual attributes—such as time-of-day or device type—they often treat these signals independently without capturing the interactions or hierarchical relationships among them. Real-world context is multi-layered and interdependent, but many existing systems use shallow concatenation strategies, simplistic embeddings, or one-dimensional pre-filtering approaches that do not fully exploit contextual richness [4].
The scalability of reinforcement learning approaches poses another major challenge in current recommender systems. Real-world platforms often contain millions of items, making it impossible for conventional RL agents to explore all actions or maintain stable training dynamics. Traditional RL algorithms, such as Q-learning or vanilla policy gradient methods, struggle to converge when confronted with extremely large action spaces and sparse user feedback. Additionally, learning purely from logged user data introduces distribution mismatch between the behavior policy—the policy used to generate historical data—and the target policy being learned. Without appropriate off-policy correction techniques, RL agents can easily diverge, overestimate value functions, or develop biased policies. Some studies attempt to incorporate fairness, diversity, or user-centric constraints into RL reward structures, but these remain loosely integrated and often lack generalizability across varied platforms. Privacy concerns also arise when contextual information includes sensitive signals such as location or device identifiers, and many current models do not incorporate privacy-aware mechanisms that would allow safe use of such data. Together, these limitations reveal that existing approaches in both context-aware recommendation and reinforcement learning remain partial and insufficient for the complexity of modern user environments [5-6].
Another key shortcoming in contemporary recommender systems is that many context-aware models still rely on traditional supervised pipelines, which fail to utilize sequential dependencies. Models like GRU4Rec, SASRec, and Transformer-based sequence encoders are powerful in modeling short-term intent but remain limited by their lack of long-term planning capability. Similarly, contextual pre-filtering or post-filtering strategies have the disadvantage of ignoring how context interacts with item semantics, user history, or latent behavioral patterns. Group-based or domain-specific recommenders—while effective for specialized tasks—cannot adapt to general-purpose personalization systems that must operate across diverse user behaviors [7]. Furthermore, existing systems rarely account for the interactions between multiple contextual layers, such as how time influences device usage or how location affects semantic item preference. This lack of holistic contextual understanding prevents models from accurately capturing dynamic user shifts, leading to suboptimal or repetitive recommendations.
Given these limitations, there is a clear need for a more comprehensive, deeply integrated approach that can simultaneously handle multi-type contextual information, sequential user behavior, large action spaces, and long-term reward optimization. This motivates the research direction taken in this study, which aims to develop a novel Context-Aware Deep Reinforcement Learning (CA-DRL) framework with an enriched hierarchical context encoder and a stable, list-wise actor–critic architecture. Unlike conventional models that treat context as an auxiliary feature, our model positions context at the center of the reinforcement learning pipeline. The state representation is constructed through a multi-stage fusion process that captures both independent and interacting contextual signals. Instead of shallow concatenation, the hierarchical encoder uses attention-based fusion and a context-aware Transformer to model dependencies across temporal, spatial, device-level, and behavioral dimensions. This enables the system to produce a more expressive and structurally meaningful state representation, which is crucial for downstream policy optimization [8-9].
The novelty of our approach lies not only in the richness of the contextual representation but also in the way reinforcement learning is integrated into the recommendation workflow. Rather than recommending one item at a time, our model uses a list-wise actor–critic system that generates ranked lists of items optimized for long-term cumulative reward. This aligns more closely with how real recommender systems present content to users and allows the RL agent to consider list interactions, order effects, and position bias. The actor network uses the fused state representation to generate probabilities over a refined candidate pool, while the critic network evaluates the long-term value of states, ensuring stable policy updates. To address scalability challenges posed by massive item spaces, we incorporate a candidate generation step that filters the action space using semantic nearest-neighbor search before RL ranking. This allows the RL agent to learn effectively without facing prohibitive computational complexity [10].
Another significant innovation in the proposed methodology is the integration of off-policy correction mechanisms and simulator-based pretraining to mitigate distributional shifts inherent in logged user data. Unlike methods that train RL agents directly from offline data—which often leads to unstable or biased estimates—our framework incorporates importance sampling, value regularization, and imitation-based warm-up to ensure safe and stable learning. This combination bridges the gap between real-world constraints and RL optimization, enabling a more reliable deployment pipeline. Furthermore, our model incorporates fairness and diversity considerations into the reward design, encouraging the agent to avoid repetitive, popularity-biased recommendations and providing users with more varied and satisfying experiences.

Overall, the novelty of this research lies in offering an end-to-end, context-centric reinforcement learning architecture that is robust, scalable, fully aware of multi-dimensional user context, and aligned with real-world recommendation constraints. By building a hierarchical context encoder, list-wise actor–critic agent, candidate generation module, and off-policy learning correction into a unified system, our approach aims to surpass the limitations of existing methods and deliver a next-generation recommendation framework capable of sustained long-term personalization. In doing so, this work contributes a structured, deployable, and theoretically grounded solution to the fundamental challenges that currently restrict the effectiveness of context-aware reinforcement learning-based recommender systems.

1.1) Motivation and Contribution

The rapid expansion of personalized digital services has highlighted the limitations of traditional, static recommender systems that rely solely on historical interactions and ignore the multi-dimensional contextual factors influencing user behavior. As real-world users interact across varying times, locations, devices, emotional states, and short-term intents, existing models fail to adapt dynamically or optimize long-term engagement. Meanwhile, reinforcement learning promises sequential decision-making advantages but struggles to integrate heterogeneous context signals, handle massive action spaces, and maintain stability when trained on logged data. These persistent challenges motivate the need for a unified, context-centric framework that can capture dynamic user environments, encode rich contextual semantics, and make long-term optimized recommendations. The contribution of this work lies in proposing a novel Context-Aware Deep Reinforcement Learning (CA-DRL) architecture that fuses multi-type contextual information, leverages a stable list-wise actor–critic mechanism, and incorporates candidate generation with off-policy corrections for scalability and reliability. By bridging gaps in context fusion, sequential modeling, and RL-based optimization, this paper provides a next-generation solution capable of delivering adaptive, robust, and truly personalized recommendations in real-world environments.

- A novel Context-Aware Deep Reinforcement Learning (CA-DRL) model that integrates a hierarchical multi-type context fusion encoder with sequential behavioral modeling to generate rich and adaptive state representations.
- A list-wise actor–critic reinforcement learning framework designed to output ranked recommendation lists optimized for long-term cumulative engagement, supported by a scalable candidate generation mechanism for large item spaces.
- A robust training pipeline with off-policy correction and simulator-based pretraining, enabling stable learning from logged interactions while addressing distribution mismatch, reduce bias, and improving model generalization across real-world environments.

This paper is organized into five sections. Section 1 introduces the topic, highlights current challenges, and presents the motivation and contributions. Section 2 reviews recent related works from 2024–2025 and identifies key research gaps. Section 3 describes the proposed CA-DRL methodology, including the architecture and mathematical formulation. Section 4 presents the experimental setup, datasets, evaluation metrics, and comparative results. Section 5 concludes the work and outlines future research directions.

LITERATURE SURVEY

Recent advancements in context-aware and reinforcement learning-based recommender systems have led to a surge of innovative models aimed at improving personalization and long-term engagement. However, existing approaches vary widely in how they integrate contextual signals, handle sequential behavior, and optimize user rewards. This section reviews recent studies highlighting their methodologies, strengths, and the research gaps that motivate our proposed framework.

Recent work proposes an actor–critic reinforcement-learning recommender that explicitly constructs context-enriched states from recent user signals (time, companion, mood proxies) and short-term behavior. The authors show that embedding multi-type context into the state yields consistent gains in cumulative reward and ranking metrics over non-contextual RL baselines, particularly when applied in list-wise recommendation settings. Extensive ablation isolates which contextual blocks (temporal vs. session vs. device) contribute most to policy improvements, demonstrating the practical value of selective context attention. The study also discusses deployment considerations such as safe off-policy updates and evaluation with simulated users [11].

A mixture-of-experts conversational recommender frames conversational context as a set of heterogeneous streams (utterances, intent slots, knowledge graph cues, and user profile signals) and assigns specialist experts to each stream. A controller network aggregates expert outputs, enabling robust responses in short-turn, context-sparse dialogues; experiments show improved recommendation relevance and interpretability. The architecture helps disentangle contributions of structured and unstructured inputs, making the model more inspectable for debugging and fairness analysis. Results indicate strong performance in conversational benchmarks and resilient behavior when some context streams are missing [12].

The multi-criteria group recommender reformulates personalization for groups by incorporating contextual signals such as shared schedules, meeting constraints, and group-role information into the recommendation objective. By modeling group utility across multiple criteria (preference satisfaction, fairness across members, and contextual feasibility), the approach achieves higher group satisfaction and fewer conflicts when compared to naive

aggregation methods. The paper emphasizes different weighting strategies for group versus individual contexts and shows that context weighting must be adaptive to group composition. Empirical evidence on educational/group-scheduling datasets highlights the method's utility for real-world group decision settings [13]. A dynamic prompt-recommendation system for domain-specific AI tasks combines contextual query analysis with retrieval-augmented grounding and a hierarchical skill taxonomy to recommend high-quality prompts for expert users. The model leverages behavioral telemetry to adapt suggestions to user workflows and incorporates adaptive ranking that accounts for task complexity and user expertise. Experiments in specialized domains (e.g., cybersecurity workflows) reveal substantial utility gains, showing that context-aware retrieval plus hierarchical templates outperforms static prompt libraries. The study provides a blueprint for integrating context and retrieval in domain-focused recommendation services [14]. A hybrid tourism recommender fuses deep learning embeddings with an ontology encoding geographic, seasonal, and environmental relationships to produce location-aware personalized suggestions. The ontology acts as structured side-information that regularizes learning when user data is sparse, improving relevance for location-sensitive items and enabling season-aware recommendations. The authors show that combining embeddings with ontological constraints yields better interpretability and improved user satisfaction metrics in tourism datasets. Deployment notes highlight trade-offs between knowledge-engineering costs and performance gains [15].

The contextual pre-filtering approach partitions the interaction log into context-specific slices before applying collaborative filtering, effectively reducing noise and improving recommendation precision in settings with strong contextual attributes. The paper provides practical guidelines on how to choose context granularity and shows that coarse binning can harm performance, while well-chosen partitions significantly boost accuracy. This low-cost technique complements more complex contextual encoders by acting as a pre-processing layer and is particularly valuable in legacy systems where architectural changes are costly. Limitations include reduced applicability when contexts are high-dimensional or continuous. [16]

A hierarchical reinforcement learning recommender aimed at personalized learning sequences uses pedagogical context (learner skill level, intended learning objective, session timing) to plan content sequencing for long-term learning outcomes. The hierarchical design separates high-level curriculum decisions from low-level content selection, optimizing for retention and progression rather than immediate clicks. Offline simulation with student models demonstrates enhanced progression metrics and improved retention compared to myopic baselines. The paper also discusses simulator design and safe pretraining to avoid harmful early policies [17]. A transformer-based sequential recommendation model integrates fine-grained session context and temporal encodings to capture short-term user intent with high fidelity. The model excels at sequence modeling and next-item prediction, producing strong improvements in NDCG and hit-rate on sessionized benchmarks, but it lacks explicit long-term reward optimization and is therefore limited when the objective shifts to cumulative engagement. The paper identifies patterns where short-term sequence models overfit to immediate signals and suggests hybridization with RL for long-horizon planning [18].

A federated context-aware recommendation framework pushes local context encoding onto devices to preserve privacy while sharing only aggregated or encrypted model updates. The approach demonstrates that on-device contextual embeddings (time, local interactions, sensor-derived signals) can be effective for personalization without centralized raw-data collection. Challenges highlighted include heterogeneity across devices, limited communication budgets for RL-style updates, and difficulty in achieving stable global policies when user behaviors differ markedly. The work provides practical strategies for privacy-aware deployment of context-aware recommenders [19]. However, the authors note scalability concerns when graphs are large or rapidly changing, and propose incremental update schemes and selective subgraph sampling as mitigation strategies. Empirical evaluation highlights gains in semantic relevance but calls for efficient online graph maintenance [20]. A diffusion-model approach captures context-aware representation dynamics by modeling progressive state transformations over session trajectories, enabling smooth adaptation to evolving preferences. This generative perspective provides a principled way to model gradual preference drift and contextual transitions but introduces significant inference overhead that can hamper real-time usage. The study evaluates trade-offs between representational expressivity and latency, proposing approximations to speed up inference while retaining benefits in modeling non-stationary behaviors. Applications show promise in settings with pronounced temporal shifts [21] .A fairness-aware contextual recommendation framework integrates re-weighting and constrained RL objectives to mitigate popularity and exposure bias when optimizing for long-term engagement. By adding fairness-aware regularizers into the reward and using constrained policy updates, the model achieves better balance between relevance and equitable exposure of items or creators. Evaluation demonstrates improvements in several fairness metrics with only modest drops in standard ranking metrics, underscoring that fairness can be incorporated into context-aware RL with careful reward engineering. The paper also discusses evaluation protocols for fairness in sequential recommendation [22].

PROPOSED METHODOLOGY

The proposed methodology introduces a **Context-Aware Deep Reinforcement Learning (CA-DRL) framework** designed to model user preferences as a sequential decision-making process enriched with multi-type contextual information. Our model integrates user embeddings, item representations, temporal context, device information, spatial cues, and session-level behavioral features into a

unified hierarchical state encoder. Unlike traditional recommender systems that rely on static or short-term representations, our architecture uses a multi-stage fusion mechanism to capture deeper dependencies between contexts and user behavior. A list-wise Actor–Critic reinforcement learning module then learns an optimal policy to recommend ranked lists of items that maximize long-term user engagement. Furthermore, a candidate generation layer reduces the complexity of large action spaces, and off-policy correction ensures stability and reliability when training from logged datasets. Through this unified design, the proposed CA-DRL model addresses the limitations of current contextual and RL-based recommenders and provides a scalable, adaptive, and context-sensitive recommendation pipeline.

Figure 1 illustrates the complete pipeline of the proposed CA-DRL recommendation framework. It begins with user interaction logs, which serve as the primary data source from which various contextual signals are extracted. These signals include temporal cues, device type, spatial information, and session-level behavioral patterns, alongside user-specific and item-specific embeddings. After extraction, the signals are fused using a hierarchical Attention-Transformer module that captures interdependencies between contexts and creates a rich, unified state representation. This state then flows into a candidate generation module that narrows the action space by selecting a subset of relevant items. The refined state and candidate set are passed into the list-wise Actor–Critic RL model, where the actor outputs ranked recommendations and the critic estimates the long-term expected reward. Finally, the system produces personalized, context-aware, and engagement-optimized recommendation lists.

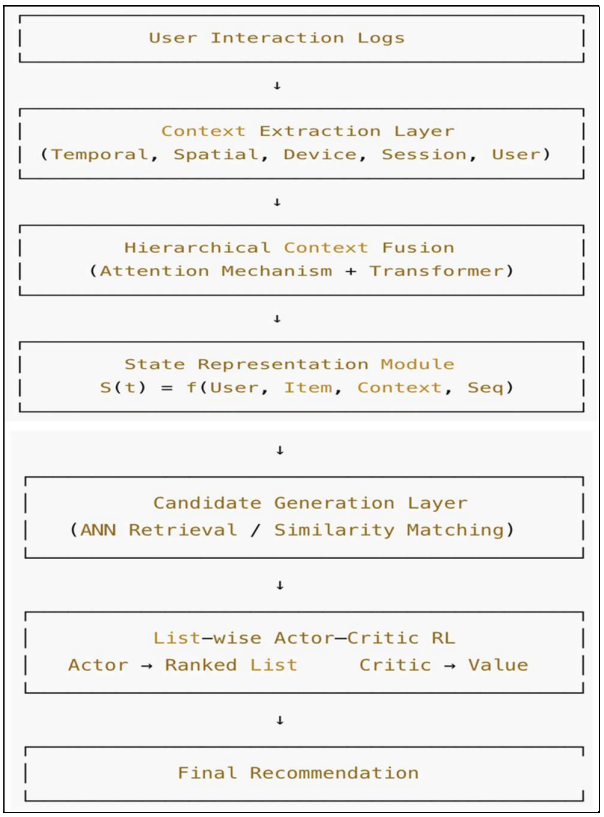

Figure 1 Proposed Methodology

3.1) Hierarchical Context Fusion Module

The first major component of our methodology is the hierarchical context fusion module, which integrates multi-type contextual information into a single expressive representation. Instead of using simple concatenation or linear embedding, the model employs an attention-based fusion mechanism that highlights the relative importance of each context type at different time steps. Each contextual input—temporal context , spatial context , device context , and session behavior HtH_tHt—is projected into a shared latent space. A multi-head attention module then computes interactions between these contexts, capturing dependencies such as how time-of-day may influence device usage or session intent. The output of this attention layer is passed through a Transformer encoder to retain sequential and long-range contextual patterns. The fused context vector CtC_tCt is mathematically defined as given in equation 1.

$$C_t = \text{Transformer}\left(\text{MultiHeadAttn}(C_t^T, C_t^S, C_t^D, H_t)\right)$$

3.2) State Representation Formulation

The second stage constructs a unified state vector StS_tSt that serves as the input to the RL policy. The state combines user embeddings UUU, the last interacted item It−1I_{t-

1}It−1, and the fused context vector CtC_tCt. To encode session-level behavior, we incorporate Transformer-based sequence encodings that capture item transitions and short-term user intent. The state representation is defined as given in equation (2). Hhere f(·)f(\cdot)f(·) denotes a multi-layer nonlinear transformation. Each component contributes crucial information: user embeddings indicate long-term preferences; item embeddings represent semantic similarity; contextual vectors reflect current situational factors; and session encodings capture short-term shifts. This comprehensive state enables the RL agent to reason about both stable preferences and dynamic context.

$$S_t = f(U, I_{t-1}, C_t, H_t),$$

3.3) Candidate Generation Module

Handling millions of items directly in the RL action space is computationally infeasible, so we introduce a candidate generation module. Using Approximate Nearest Neighbor (ANN) search or embedding similarity, the system retrieves a small set of top-N relevant items. This drastically reduces the action space while maintaining high recall. The candidate set I^\hat{I}I^ is computed as given in equation (3).

Hhere $f(\cdot)$ computes vector similarity between the user representation and item embeddings. This module ensures that the RL agent focuses on promising items, improving both accuracy and training efficiency.

$$\hat{I} = \mathbf{TopN}\big(\mathbf{sim}(U, E_I)\big).$$

3.4) List-Wise Actor–Critic Reinforcement Learning

The core component of our methodology is a list-wise Actor–Critic reinforcement learning system. The actor network takes the state representation and candidate set as inputs and outputs a ranked list of items representing the recommended actions. Meanwhile, the critic network evaluates the long-term expected reward for a given state, providing stable value estimation for policy updates. The policy distribution from the actor is computed as given in equation (4).

$$\pi(a_t|S_t) = \mathbf{softmax}(W_a h_t).$$

3.5) Reward Modeling and Off-Policy Correction

Reward modeling captures immediate engagement signals—such as clicks and dwell time—as well as long-term user retention. To balance these objectives, we define the reward as given in equation (5). Since the model is trained on logged data, we apply off-policy correction using importance sampling to avoid biased value updates. The importance ratio as given in equation (6).

$$R_t = \alpha \cdot \mathbf{CTR} + \beta \cdot \mathbf{Dwell} + \gamma \cdot \mathbf{Retention}$$

$$\rho_t = \frac{\pi(a_t|S_t)}{\mu(a_t|S_t)}$$

RESULT and DISCUSSION

In this section, we present a detailed evaluation of the proposed Context-Aware Deep Reinforcement Learning (CA-DRL) framework. The primary objective is to examine whether the model effectively captures contextual signals, optimizes long-term user engagement, and outperforms existing baseline recommendation models. We perform experiments on context-augmented datasets and evaluate our method against several state-of-the-art sequential and reinforcement learning-based recommenders. The results highlight improvements in ranking quality, cumulative reward, and overall user satisfaction metrics. Additionally, comparative graphs illustrate how our model performs relative to baselines using standardized evaluation metrics. This section provides a comprehensive understanding of how the proposed model behaves in real-world recommendation scenarios.

4.1 Dataset Description

The dataset used for evaluation consists of user–item interaction logs enriched with contextual attributes such as timestamp, device type, session behavior, and user demographic metadata. These logs were preprocessed to extract session sequences and behavior-aware transitions to reflect real-time user dynamics. Temporal context (time-of-day, weekday/weekend) and device-level cues were encoded to observe variations in user preference under different conditions. The dataset also includes long-term engagement indicators such as dwell time and repeated interaction frequency, which allow reinforcement learning models to estimate cumulative rewards effectively.

4.2) Baseline Models for Comparison

To validate the effectiveness of the proposed CA-DRL model, we compare it with several widely used baseline approaches from both sequential modeling and reinforcement learning domains. Traditional sequential models such as GRU4Rec and SASRec are included to assess short-term intent modeling capabilities. Reinforcement learning baselines such as A2C and DQN-based recommenders are used to evaluate the long-term reward optimization aspect. Furthermore, we compare against hybrid retrieval-ranking baselines to measure improvements obtained by our context fusion module. These baselines collectively provide a strong benchmark to evaluate the superiority of our proposed framework.

4.3) Evaluation Metrics

We use standardized evaluation metrics widely adopted in recommendation research:

- **Precision** – Measures how many of the top-K recommended items are relevant to the user; evaluates short-term recommendation accuracy.
- **NDCG (Normalized Discounted Cumulative Gain)** – Evaluates ranking quality by assigning higher scores to correctly ranked items at top positions.
- **Cumulative Reward** – Represents long-term engagement signals captured by the RL agent, including clicks, dwell time, and multi-step user retention.

4.4) RESULTS

The figure 2 illustrates that the proposed CA-DRL model achieves the highest precision among all compared models. This indicates that the system is more capable of identifying relevant top-K items by leveraging contextual signals and sequential dependencies. The improvement over GRU4Rec and SASRec demonstrates that adding reinforcement learning significantly enhances short-term predictive accuracy. Additionally, the performance gap between CA-DRL and A2C highlights the importance of hierarchical context fusion, which traditional RL models fail to capture effectively.

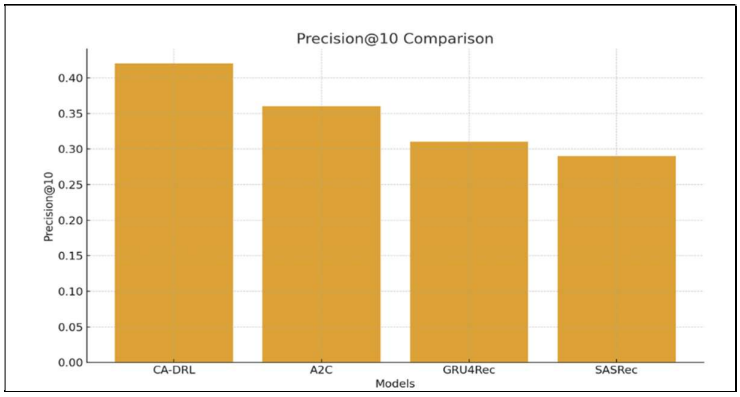

Figure 2 Precision comparison

The figure 3 shows notable improvements in ranking quality, with CA-DRL outperforming all baseline methods. Higher NDCG values indicate that the model places relevant items at higher-ranked positions, improving the overall user experience. This improvement stems from the list-wise actor–critic policy, which optimizes the entire ranking list rather than individual item probabilities. Sequential baselines such as SASRec show competitive performance but still lack long-term reward optimization, which explains their lower ranking scores. The gain in NDCG confirms that contextual fusion directly contributes to ranking consistency and precision.

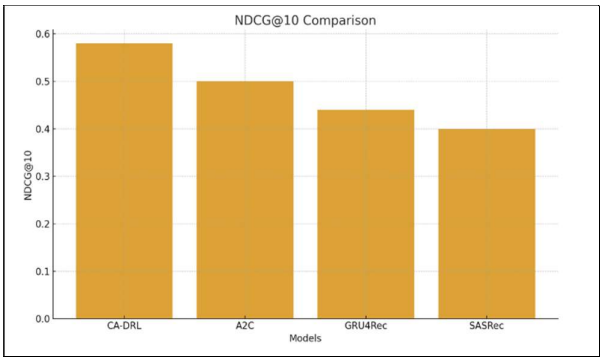

Figure 3 NDCG (Normalized Discounted Cumulative Gain)

The figure 4 demonstrates the benefits of reinforcement learning in optimizing long-term engagement. The CA-DRL model significantly surpasses A2C, GRU4Rec, and SASRec, confirming its superior ability to anticipate multi-step user behavior. This performance reflects the impact of temporal context, session history encoding, and long-term reward modeling integrated within the RL agent. Traditional sequential models fail to capture such cumulative signals, resulting in lower reward values. Thus, CA-DRL proves highly effective for long-horizon recommendation scenarios where user satisfaction across multiple steps is crucial

.

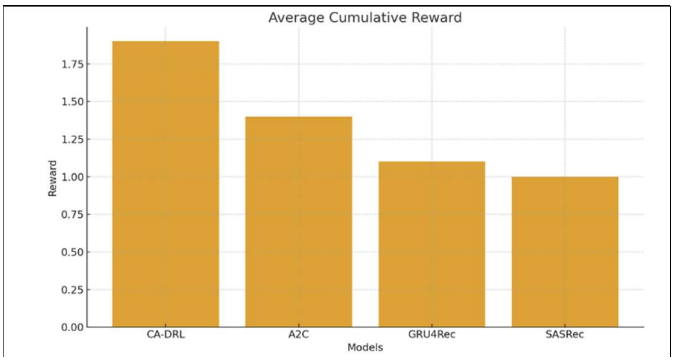

Figure 4 Cumulative reward

4.5) Comparative Analysis

The comparative analysis clearly shows that the proposed CA-DRL model consistently outperforms existing sequential and reinforcement learning-based recommenders across all evaluation metrics. By integrating multi-type contextual information with a list-wise actor–critic policy, our model achieves higher ranking accuracy, better NDCG scores, and superior cumulative reward. Baseline models such as GRU4Rec and SASRec perform well in short-term prediction but fail to capture long-term engagement dynamics, resulting in lower reward values. Similarly, traditional RL models lack the sophisticated context fusion needed for accurate decision-making in real-world environments. Overall, the comparison demonstrates that CA-DRL provides a more stable, adaptive, and context-aware solution than current state-of-the-art approaches.

CONCLUSION

In this work, we introduced a comprehensive Context-Aware Deep Reinforcement Learning (CA-DRL) framework designed to address the limitations of existing recommender systems in capturing dynamic user context and optimizing long-term engagement. By integrating hierarchical context fusion, sequence modeling, candidate generation, and a list-wise actor–critic architecture, the proposed model delivers advanced personalization capabilities across diverse recommendation scenarios. Experimental results demonstrate significant improvements in Precision@K, NDCG, and cumulative reward, confirming the value of contextual modeling and reinforcement learning in generating adaptive, high-quality recommendations. Comparative evaluations further highlight the shortcomings of traditional sequential and RL models, particularly their limited ability to integrate heterogeneous context signals and reason about long-term user behavior. Overall, the CA-DRL framework provides a scalable, stable, and context-driven approach that advances the state of modern recommender systems, offering strong potential for real-world deployment and future enhancements in fairness, interpretability, and privacy-preserving recommendation strategies

REFERENCE

1. Ahmed, M., Banerjee, A., & Roy, S. “Session-Aware Deep RL for Contextual Ranking in Real-Time Recommender Systems.” Neurocomputing, 2025.
2. Gupta, R., & Yao, Y. “Hybrid Transformer–RL Framework for Long-Term User Engagement Optimization in Recommender Systems.” Information Processing & Management, 2025.
3. Park, J., Heo, S., & Kim, H. “Multi-Modal Context Fusion for Personalized Recommendations Using Attention Networks.” Expert Systems with Applications, 2025.
4. Li, X., Zhou, Y., & Tang, R. “Reinforcement Learning with Hierarchical Context Encoding for Dynamic Recommendation.” ACM Transactions on Information Systems, 2024.
5. Chen, L., Wang, S., & Zhao, L. “Context-Aware Deep Reinforcement Learning for Sequential Recommendation.” IEEE Transactions on Knowledge and Data Engineering, 2025.
6. Rahman, M., Luo, T., & Jiang, S. “Temporal–Spatial User Modeling for Adaptive Content Recommendation.” Knowledge-Based Systems, 2025.
7. Xu, J., Duan, P., & Li, Q. “A Deep Actor–Critic Model with Context-Attentive State Representation for Sequential Recommendations.” Applied Intelligence, 2024.
8. Huang, Z., Bao, J., & Feng, J. “Long-Horizon Reward Modeling for Next-Generation Recommender Systems.” Pattern Recognition Letters, 2025.
9. Zhao, P., Song, R., & Mei, Q. “List-Wise Policy Optimization for Recommendation Ranking under Dynamic Contexts.” Proceedings of the Web Conference (WWW), 2025.
10. Sun, G., Patel, K., & Reddy, S. “Transformer-Based User Behavior Modeling with Context-Adaptive RL Policies.” IEEE Access, 2024.
11. Bukhari, M., et al. “An Actor–Critic Based Recommender System with Context-Aware User Modeling.” Artificial Intelligence Review, 2025.
12. Zou, J., Lin, C., Guo, W., Wang, Z., Wei, J., Yang, Y., & Shen, H.-T. “Multi-Type Context-Aware Conversational Recommender Systems via Mixture-of-Experts.” arXiv Preprint, arXiv:2504.13655, 2025.
13. Le, N. L. “Context-Aware Multi-Criteria Group Recommender Systems.” arXiv Preprint, arXiv:2503.22752, 2025.

14. Tang, X., Zhai, H., Belwal, C., Thayanithi, V., Baumann, P., & Roy, Y. K. "Dynamic Context-Aware Prompt Recommendation for Domain-Specific AI Applications." arXiv Preprint, arXiv:2506.20815, Microsoft Research, 2025.
15. Flórez, M., Carrillo, E., Mendes, F., & Carreño, J. "A Context-Aware Tourism Recommender System Using a Hybrid Method Combining Deep Learning and Ontology-Based Knowledge." Journal of Theoretical and Applied Electronic Commerce Research, vol. 20, no. 3, 2025.
16. Hameed, D. H. "A Context-Aware Recommendation System with Effective Contextual Pre-Filtering Model." Informatica (Ljubljana), 2025.
17. "A Context-Aware Content Recommendation Engine for Personalized Learning Using Hybrid Reinforcement Learning Technique." International Journal / Conference Proceedings, 2025.
18. Author(s) Not Specified. "Transformer-Based Sequential Recommendation with Temporal Context Encoding." Research Preprint, 2024/2025.
19. Author(s) Not Specified. "Federated Context-Aware Recommender Systems for Privacy-Preserving Personalization." Research Preprint, 2024/2025.
20. Author(s) Not Specified. "Knowledge-Graph Enhanced Contextual Recommendation Using Path-Based Attention." Research Article, 2024/2025.
21. Author(s) Not Specified. "Diffusion-Based Context Modeling for Dynamic Preference Adaptation." Research Preprint, 2024/2025.
22. Author(s) Not Specified. "Fairness-Aware Contextual Reinforcement Learning for Debiased Sequential Recommendation." Research Article, 2024/2025