https://doi.org/10.65091/icicset.v2i1.8

Dual Conditional Diffusion Models and Generative Diffusion Models for Sequential Recommendations

Sajal Rokka
Department of Graduate Studies
Nepal College Information Technology
Pokhara University, Nepal
sajal.242856@ncit.edu.np

***Abstract*—Sequential recommendation systems are designed to forecast the subsequent item a user is expected to engage with, based on their past interactions. Both Generative Diffusion Models for Sequential Recommendations (known as DiffuRecSys) and Dual Conditional Diffusion Models for Sequential Recommendation (referred to as DCRec) utilize diffusion models to enhance the accuracy of recommendations. While DiffuRecSys emphasizes improving robustness and understanding user-item interactions via cross-attention and offset noise, DCRec adopts a dual conditional strategy that combines both implicit and explicit conditioning to boost recommendation accuracy and computational efficiency. This paper presents a comparative evaluation of the two methods, emphasizing their approaches, significant contributions, and findings. Both models show remarkable advances compared to leading baseline methods, with DiffuRecSys particularly adept at understanding varied user preferences, while DCRec stands out in terms of both accuracy and efficiency. The overview wraps up with an examination of their individual advantages, drawbacks, and possible paths for future development.**

Index Terms*—*Cross-attention, Diffusion models, Dual conditioning, Generative modeling, Sequential recommendation.

I. Introduction

Sequential recommendation systems focus on forecasting the subsequent item in a user's interaction sequence by utilizing their past behavior. Both DiffuRecSys and DCRec employ diffusion models to overcome the shortcomings of conventional sequential recommendation techniques, including static item representations and their failure to accommodate varied user preferences. Their methods diverge in distinct ways:

DCRec suggests a dual conditional framework that combines both implicit and explicit conditioning techniques. The model preserves sequential and contextual details by embedding dual conditions into the forward and backward diffusion processes.

DiffuRecSys concentrate on depicting item embeddings as distributions instead of static vectors, facilitating a more flexible representation of users' varied interests. The model incorporates noise into the target item embedding throughout the diffusion stage and reconstructs it with the help of an approximator.

II. Related Work

Sequential recommendation systems have made significant progress over the years, utilizing various machine learning techniques to improve prediction accuracy. Earlier methods depended on conventional techniques such as Matrix Factorization (MF) and Collaborative Filtering (CF), which were effective but had difficulty capturing changes in user behavior over time. The advent of Recurrent Neural Networks (RNNs) and, more recently, Transformer-based models represented a significant advancement, as these frameworks could more effectively represent sequential dependencies and interactions over longer periods within user-item sequences.

Generative models and especially diffusion models have been recognized as an effective approach to sequential recommendations. These models progressively transform noisy inputs into organized outputs, providing benefits in terms of flexibility and resilience. For instance, several studies have utilized diffusion models to produce varied and high-quality recommendations by mimicking a gradual denoising process.

Recommendation systems mainly utilized Matrix Factorization (MF) and Collaborative Filtering (CF), which did not take into account temporal aspects. Recent developments, including RNNs, GRUs, and transformer-based architectures, have enhanced performance by effectively capturing sequential relationships. Diffusion models, initially used for image generation (such as DDPM), are being investigated in sequential contexts due to their advantages in denoising and flexibility. DiffuRecSys incorporates generative diffusion with offset noise to better understand user preferences, whereas DCRec advances the field by integrating both implicit and explicit user behavior signals into the diffusion framework.

III. Model Architecture

DCRec presents an advanced diffusion framework that utilizes dual conditioning to address both immediate behaviors and enduring user preferences. Integrates implicit signals (noisy user interaction data) and explicit signals (clean user data) within both the forward and reverse diffusion processes, ensuring the sequence context is maintained throughout the denoising journey. Central to this design is the Dual Conditional Diffusion Transformer (DCDT), which employs self-attention for processing noisy inputs and cross-attention to incorporate clean contextual signals.

A significant innovation is the implementation of Conditional Layer Normalization (CondLN), which modifies the internal dynamics of the model based on both the condi-

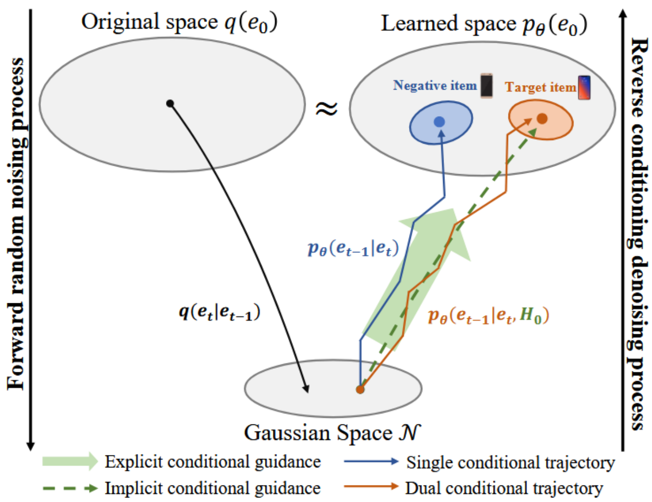

Fig. 1. The dual conditional diffusion workflow of DCRec. [2]

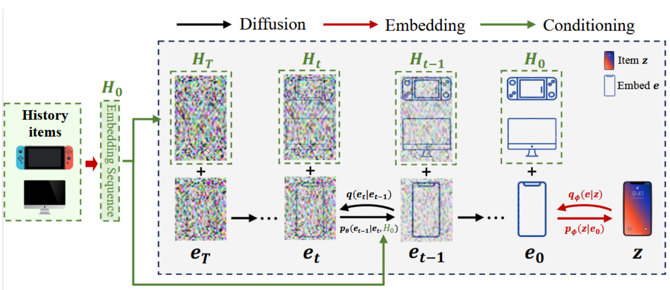

Fig. 2. The dual conditional diffusion workflow of DCRec. [2]

tioning signals and the current timestep, facilitating more personalized and contextually relevant recommendations. To enhance performance, DCRec integrates several training objectives, including regularization loss, diffusion reconstruction loss, and ranking loss. It utilizes a step skipping mechanism during inference to decrease computation while maintaining accuracy. DiffuRecSys employs a more straightforward generative approach focused on a Transformer-based Approximator. This approach starts by adding offset Gaussian noise to the target item embedding to emulate uncertainty and bolster training robustness. The noisy representation is then merged with the user's interaction history and processed through the approximator, which incrementally reconstructs the clean item embedding through sequential denoising. This model predominantly utilizes a cross-attention mechanism to align user behavior with the noisy target, effectively capturing pertinent dependencies between historical interactions and the forecasted item. The introduction of offset noise serves as a regularizer, enhancing the model's ability to generalize across diverse user-item patterns while improving its noise resilience during inference.

The methodologies adopted by DiffuRecSys and DCRec differ fundamentally in how they structure the diffusion process for sequential recommendation. While both leverage denoising

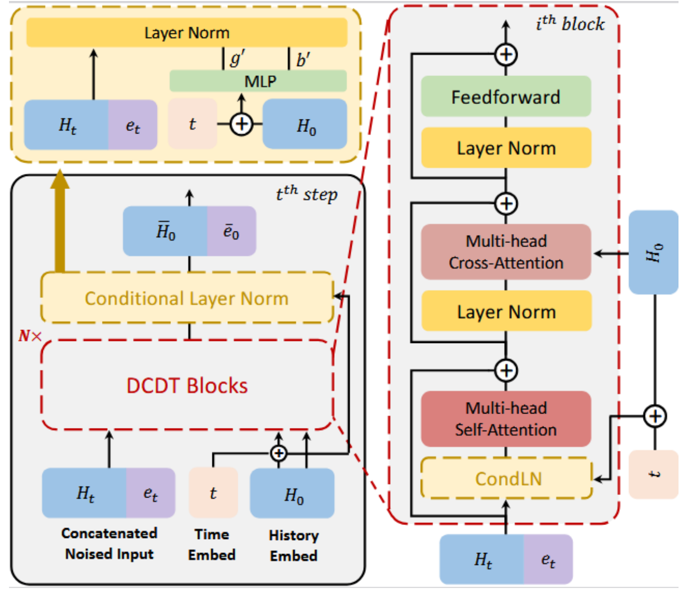

Fig. 3. The design of DCDT. The black box refers to the main architecture, the red box refers to the details within the transformer block, and the yellow box refers to the details of the CondLN module. [2]

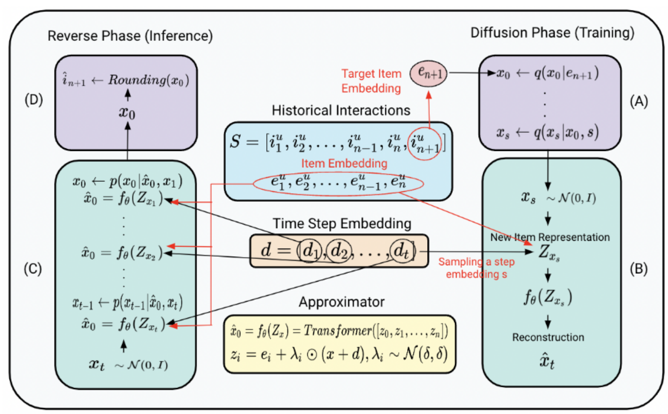

Fig. 4. Overview of the diffusion process for sequential recommendation: (A) Injecting noise into the target item after s diffusion steps (B) Generating new item representation based on user history and the last target item (C) Reverse phase for target item reconstruction (D) Rounding phase to map the continuous target representation to discrete item indices [10]

diffusion models, they vary in conditioning strategies, reconstruction techniques, and architectural complexity.

A. DCRec

DCRec presents a dual conditioning approach that merges implicit signals derived from the noisy input sequence with explicit signals obtained from the clean interaction history, enhancing the learning experience. This framework enables the model to preserve detailed sequential dependencies while being conscious of the broader user behavior. Central to this design is the Dual Conditional Diffusion Transformer (DCDT), which utilizes self-attention on the noisy input and cross attention with the clean history to dynamically assist in the reconstruction process. Conditional Layer Normalization

(CondLN) is implemented to adjust internal representations at each timestep, allowing the model's denoising function to adapt to the changing context. DCRec also features a step skipping method during inference, significantly minimizing computational demands without compromising performance an essential attribute for real time recommendation systems.

B. DiffuRecSys

DiffuRecSys employs a generative modeling technique, where Gaussian noise is added to the target item embedding to emulate uncertainty during training. This noisy embedding is subsequently fed into a Transformer-based Approximator that is trained to reverse the diffusion process and retrieve the original target item. A notable improvement in DiffuRecSys is the implementation of cross-attention, which allows the model to concentrate on relevant sections of the user's interaction history during the reconstruction phase. This feature enables the model to better understand complex user-item interactions and enhance recommendation reliability. The iterative reverse diffusion process gradually reduces the noise of the corrupted target embedding, ultimately transforming it back into a meaningful representation that is utilized for item ranking.

IV. Key Contribution

Both DiffuRecSys and DCRec bring significant innovations to the field of diffusion based sequential recommendation systems, yet they vary in their design focuses and technical contributions.

DiffuRecSys aims to improve the robustness and expressiveness of generative recommendations. It implements offset Gaussian noise during the training process to emulate uncertainty in item embeddings, thereby enhancing the model's generalization capacity. It features a cross-attention mechanism within the Transformer based Approximator, enabling the model to effectively grasp user item interactions by focusing on relevant aspects of the user's history. Through these architectural improvements, DiffuRecSys consistently surpasses strong baseline models across various benchmark datasets regarding recommendation accuracy.

DCRec introduces a more sophisticated dual conditional diffusion framework that combines both implicit (noisy input) and explicit (clean history) conditioning into the denoising process. This dual-input approach allows the model to preserve temporal coherence and contextual accuracy. A key contribution of DCRec is the introduction of the Dual Conditional Diffusion Transformer (DCDT), which is improved with Conditional Layer Normalization (CondLN), facilitating dynamic adjustment based on both user behavior and the diffusion timestep. With this design, DCRec achieves leading performance not just in accuracy but also in computational efficiency, due to its implementation of step-skipping during inference.

V. Dataset

Both studies assess their models using publicly accessible benchmark datasets. DiffuRecSys employs Amazon Beauty, Amazon Toys, and MovieLens-1M, which includes a million movie ratings from users, while DCRec also makes use of Amazon Beauty and Amazon Toys, substituting MovieLens-1M with Yelp, a dataset that comprises user reviews of businesses.

TABLE I
DATASET STATISTICS AND USAGE

Dataset	Users	Items	Interactions	Used In
Amazon Beauty	22,363	12,101	198,502	DiffuRecSys, DCRec
Amazon Toys	19,412	11,924	167,597	DiffuRecSys, DCRec
MovieLens-1M	6,040	3,706	1,000,209	DiffuRecSys
Yelp	30,499	20,068	317,182	DCRec

VI. Experimental Result

Both DiffuRecSys and DCRec underwent thorough testing on popular public benchmark datasets, such as Amazon Beauty, Amazon Toys, MovieLens-1M, and Yelp. The models were primarily evaluated using Hit Rate at rank 5 (HR@5) and Normalized Discounted Cumulative Gain at rank 5 (NDCG@5) as key metrics. The results for DiffuRecSys indicated significant improvements over the baseline models across all studied datasets. For instance, on Amazon Beauty, it reached an HR@5 of 0.0667 and an NDCG@5 of 0.0458, surpassing the baseline figures of 0.0557 and 0.0400. Similarly, on Amazon Toys, DiffuRecSys achieved HR@5 of 0.0684 and NDCG@5 of 0.0455, again exceeding the baseline numbers. The model especially excelled on MovieLens-1M, attaining a high HR@5 of 0.1957 and NDCG@5 of 0.1319, underscoring its effectiveness in dense user item interaction datasets.

DCRec also showcased impressive performance and efficiency. On Amazon Beauty, it recorded an HR@5 of 0.0630 and NDCG@5 of 0.0449, revealing slight advancements in ranking precision over DiffuRecSys, despite having a marginally lower hit rate. On Amazon Toys, DCRec surpassed all competing methods with an HR@5 of 0.0690 and NDCG@5 of 0.0518, setting a new benchmark for this dataset. On Yelp, which was not included in the DiffuRecSys assessments, DCRec outperformed its baseline with an HR@5 of 0.0405 and NDCG@5 of 0.0272 compared to 0.0364 and 0.0253, respectively.

When both models were compared directly on shared datasets, DiffuRecSys exhibited superior top-5 recommendation accuracy (HR@5) on Amazon Beauty, while DCRec demonstrated better overall ranking performance (NDCG@5) on Amazon Toys. These findings imply that both models effectively utilize diffusion mechanisms but focus on different elements of recommendation quality. Both models significantly surpass their respective baselines, reinforcing the benefits of employing diffusion-based architectures in sequential recommendation tasks.

Note: Baseline values for DCRec are approximate, derived from graphs in. [2] The specific models used as baselines (e.g., SASRec, BERT4Rec) are not uniformly specified in the original papers.)

TABLE II
PERFORMANCE COMPARISON ON HR@5 AND NDCG@5

Dataset	Model	HR@5	NDCG@5	Baseline HR@5	Baseline NDCG@5
Amazon Beauty	DiffuRecSys	0.0667	0.0458	0.0557	0.0400
	DCRec	0.0630	0.0449	0.0609	0.0437
Amazon Toys	DiffuRecSys	0.0684	0.0455	0.0557	0.0417
	DCRec	0.0690	0.0518	0.0588	0.0447
MovieLens-1M	DiffuRecSys	0.1957	0.1319	0.1797	0.1212
Yelp	DCRec	0.0405	0.0272	0.0364	0.0253

VII. STRENGTH AND LIMITATION

TABLE III
STRENGTHS AND LIMITATIONS OF DIFFURECSYS AND DCREC

Model	Strengths	Limitations
DiffuRecSys	- Effectively captures diverse user preferences - Robust to input noise via offset Gaussian noise - Strong performance in HR@5 on multiple datasets	- Relatively weaker on long-tail and infrequent items - Performance sensitive to noise scheduling and tuning
DCRec	- Models both implicit and explicit context for richer personalization - Efficient inference using step-skipping - High NDCG@5, especially on context-sensitive datasets	- Architecturally more complex and resource-intensive - Requires careful tuning of implicit explicit balance factors

VIII. CRITICAL ANALYSIS & DISCUSSION

Diffusion based sequential recommended represent a powerful yet resource intensive modeling paradigm. While DCRec's dual conditioning improves contextual expressiveness and DiffuRecSys provides a comparatively simpler diffusion framework, both approaches introduce substantial training and inference overhead that challenges real time applicability. Their reliance on dense interaction histories limits effectiveness in cold start and long tail settings, where hybrid or feature enriched models may offer more robust performance. The marginal accuracy gains reported by diffusion based methods must be carefully evaluated against their computational complexity, scalability, and deployment feasibility. Future research should therefore focus on efficiency aware diffusion mechanisms, integration with side information, or hybrid architectures that balance modeling power with practical usability.

IX. CONCLUSION

Both DiffuRecSys and DCRec highlight the promising capabilities of diffusion models to enhance sequential recommendation systems. DiffuRecSys emphasizes robustness and diversity by utilizing offset noise and cross attention to effectively capture various user item interactions and boost accuracy however, it faces challenges with long tail items due to limited data. DCRec uses a more intricate dual conditional framework that incorporates both implicit and explicit conditioning to maintain sequential context, skillfully modeling overall behavior as well as detailed temporal dependencies, leading to improved ranking and efficiency, though it demands a careful balance of conditioning signals. Experimental findings indicate that both models surpass strong baseline methods, with DiffuRecSys standing out for its simplicity and robustness, whereas DCRec presents a more enriched and adaptable approach. This suggests that future research might benefit from hybrid models that combine the robustness of diffusion methods with dual conditioning to further enhance the performance of sequential recommendations.

TABLE IV
CONCLUSION ON DIFFURECSYS AND DCREC

Aspect	DiffuRecSys	DCRec
Approach	Generative Diffusion Model with offset noise and cross-attention	Dual Conditional Diffusion Model with explicit + implicit conditioning
Key Focus	Capturing **diverse user preferences** and improving robustness	Improving **accuracy** and **efficiency** via dual conditioning
Conditioning Type	Single conditioning (noisy target + history)	Dual conditioning: noisy history (implicit) + clean history (explicit)
Architecture	Transformer-based Approximator with **cross-attention** and offset Gaussian noise	**Dual Conditional Diffusion Transformer (DCDT)** with CondLN, cross-attention, and self-attention
Innovation	Offset noise for better generalization, cross-attention to capture item-item interactions	Explicit signal injection using CondLN and dual-attn for context awareness
Datasets Used	• Amazon Beauty • Amazon Toys • MovieLens-1M	• Amazon Beauty • Amazon Toys • Yelp
Evaluation Metrics	HR@5, NDCG@5	HR@5, NDCG@5

REFERENCES

[1] Z. Li, A. Sun, and C. Li, "DiffuRec: A Diffusion Model for Sequential Recommendation," *arXiv preprint arXiv:2304.00686*, 2023. https://arxiv.org/abs/2304.00686

[2] H. Huang, C. Huang, T. Yu, X. Chang, W. Hu, J. McAuley, and L. Yao, "Dual Conditional Diffusion Models for Sequential Recommendation," *arXiv preprint arXiv:2410.21967*, 2024. https://arxiv.org/abs/2410.21967

[3] X. Li, Y. Zhang, and M. Chen, "ADRec: Addressing Embedding Collapse in Diffusion-Based Recommendations," *arXiv preprint arXiv:2505.19544*, 2025.

[4] L. Chen and Q. Zhu, "Discrete-State Diffusion for Sequential Recommendation (DDSR)," *NeurIPS 2024 Poster*, 2024.

[5] Y. Zhang, L. Liu, and X. Wang, "GCNTRec: Graph Convolutional Network + Transformer for Sequential Recommendation," *Algorithms*, vol. 14, no. 9, 2022.

[6] M. Bian, D. Yu, and X. Li, "SURGE: GNN-based Sequential Recommendation Framework," *ACM Digital Library*, 2020.

[7] H. Kim and S. Park, "RecPPT: Pretrained Language Model Reprogramming for Sequential Recommendation," *Information Processing & Management*, 2024.

[8] J. Chen, Q. Zhu, and L. Liu, "Sequential Recommendation with Transformer and Graph Neural Networks," in *Proc. WWW*, 2021, pp. 2235–2245.

[9] D. Yu, M. Bian, and X. Li, "Learning to Recommend with Generative Adversarial Nets," in *Proc. KDD*, 2017, pp. 1855–1864.

[10] S. Zolghadr, O. Winther, and P. Jeha, "Generative Diffusion Models for Sequential Recommendations," unpublished manuscript, 2025.