

Adaptive Diffusion-based Augmentation for Recommendation

Na Li¹, Fanghui Sun¹, Yan Zou¹, Yangfu Zhu², Xiatian Zhu³, Ying Ma^{1*}

¹Harbin Institute of Technology, Heilongjiang, China

²Capital Normal University, Beijing, China

³University of Surrey, Guildford, United Kingdom

{24s003043, 23b903066}@stu.hit.edu.cn, {Sunfanghui, y.ma}@hit.edu.cn,
zhuyangfu@cnu.edu.cn, eddy.zhuxt@gmail.com

Abstract

Recommendation systems often rely on implicit feedback, where only positive user-item interactions can be observed. Negative sampling is therefore crucial to provide proper negative training signals. However, existing methods tend to mislabel potentially positive but unobserved items as negatives and lack precise control over negative sample selection. We aim to address these by generating controllable negative samples, rather than sampling from the existing item pool. In this context, we propose **Adaptive Diffusion-based Augmentation for Recommendation (ADAR)**, a novel and model-agnostic module that leverages diffusion to synthesize informative negatives. Inspired by the progressive corruption process in diffusion, ADAR simulates a continuous transition from positive to negative, allowing for fine-grained control over sample hardness. To mine suitable negative samples, we theoretically identify the transition point at which a positive sample turns negative and derive a score-aware function to adaptively determine the optimal sampling timestep. By identifying this transition point, ADAR generates challenging negative samples that effectively refine the model’s decision boundary. Experiments confirm that ADAR is broadly compatible and boosts the performance of existing recommendation models substantially, including collaborative filtering and sequential recommendation, without architectural modifications.

1 Introduction

In modern recommendation systems, implicit feedback (Ding et al. 2020) has become the predominant signal for learning user preferences, due to its ubiquitous availability in real-world applications. Such feedback typically includes user behaviors like clicks, views, or purchases, which are treated as positive samples. However, explicit negative signals are usually absent, making it challenging to identify items that users truly dislike. Yet the learning process critically depends on the ability to distinguish preferred items from non-preferred ones. In this context, the construction of high-quality negative samples is essential for guiding recommendation models (Rendle and Freudenthaler 2014).

To construct negative samples from implicit feedback, most existing models adopt sampling methods that pick negative samples from unobserved user-item interactions. A

common method is to perform static random sampling over these unobserved interactions due to its simplicity and efficiency (Rendle et al. 2012). However, this approach cannot provide enough negative signal, limiting model’s ability to learn accurate user preferences. To address this, recent studies in collaborative filtering have explored hard negative sampling strategies (Zhang et al. 2013), which leverage unobserved items that are incorrectly ranked highly, thereby providing more challenging supervision signals. Building on this, more sophisticated sampling strategies such as MixGCF (Huang et al. 2021) and AHNS (Lai et al. 2024) have been proposed to find hard negative samples by mixing or scoring. In sequential recommendation, sampling is further challenged by temporal and contextual dynamics of user behavior. GNNO (Fan et al. 2023) enhances sampling by identifying hard ones based on global transitions, while recent context-aware method (Seol and Choi 2025) adapts sampling to varying distributions across time.

However, these strategies are suboptimal and suffer from several limitations. First, they tend to introduce false-negative signals, as some unobserved items may actually align with user interests but remain unseen due to limited exposure. Incorrectly treating potentially relevant items as negatives may introduce misleading signals during training, thereby hindering the learning of user preferences. Second, existing methods lack fine-grained control in the negative sampling process. Although dynamic strategies (Zhang et al. 2013; Zhao et al. 2023) improve upon static heuristics by ranking candidates based on their scores, they lack flexibility in obtaining negative samples with controllable hardness. Specifically, the current approaches are unable to precisely regulate the true gap between positive and negative samples, resulting in suboptimal ranking performance.

Instead of relying on negative samples drawn from the existing item pool, we aim to model a continuous transition from positive to negative samples to address these limitations. To achieve this, we adopt a generative approach based on diffusion processes, which naturally simulate the gradual transformation. Diffusion models (Ho, Jain, and Abbeel 2020; Dhariwal and Nichol 2021a), originally designed for high-fidelity image synthesis (Zhou et al. 2025), gradually transform structured input into noise and learn to reverse this process. We observe that, during the forward noising process, a positive sample progressively loses its informa-

*Corresponding author.

tive characteristics and eventually becomes indistinguishable from a negative sample. The continuous degradation offers a natural lens for identifying the transition point at which a sample crosses from positive to negative.

In this paper, we propose **Adaptive Diffusion-based Augmentation for Recommendation (ADAR)**, a novel and model-agnostic augmentation module that employs diffusion to adaptively synthesize informative negative examples. Our core idea is to model the degradation of positive samples through a diffusion process. During this process, positive samples are progressively corrupted, and a critical transition occurs when their identity shifts from positive to negative. So we can treat the outputs at each timestep as a candidate pool of negative samples. Then we provide a theoretical characterization of the transition and introduce a score-aware function to adaptively determine the optimal sampling timestep. Leveraging this well-defined transition point, ADAR is able to adaptively generate informative and challenging negative samples. These generated samples can then be used as a plug-and-play augmentation to enhance existing various recommendation models.

Our main contributions are summarized as follows:

- We propose ADAR, a novel and model-agnostic augmentation module that utilizes diffusion to generate high-quality negative samples, addressing the limits of false-negative signals and controllability in existing methods.
- We theoretically define a transition point within the diffusion process that marks when a positive sample becomes negative, thereby enabling the adaptively choose of informative negative samples.
- ADAR is compatible with a wide range of models, including collaborative filtering and sequential recommendation, demonstrating strong generalizability without any model-specific modifications.

2 Related Work

2.1 Negative Sampling

In recommendation, negative sampling plays a crucial role in constructing training pairs that guide the model to distinguish between preferred and non-preferred items. The effectiveness of learning heavily depends on the quality and hardness of these negative samples. Simple uniform sampling over unobserved interactions (Rendle et al. 2012) often yields trivial negatives and weak supervision. For better discriminative user/item representations, hard negative sampling selects challenging negatives ranked highly by the model (Zhang et al. 2013). MixGCF (Huang et al. 2021) and DropMix (Ma, Chen, and Li 2023) obtain negative samples through a dimensional mixing mechanism. Going a step further, DINS (Wu et al. 2023) and TriSampler (Yang et al. 2024) expand the sampling area to spatial and triangular areas. AHNS (Lai et al. 2024) can select existing negative samples of different hardness.

In sequential recommendation, negative sampling must handle temporal and contextual dynamics. GNNO (Fan et al. 2023) mines hard negatives based on global item transitions, while context-aware methods (Seol and Choi 2025) adapt

sampling to context-dependent item distributions, improving recommendation quality under dynamic user behaviors.

However, these methods lack a principled way to model the continuous transition between positive and negative samples, leading to suboptimal model performance.

2.2 Diffusion Model

Diffusion model (Ho, Jain, and Abbeel 2020; Dhariwal and Nichol 2021b; Ho et al. 2022) is a generative modeling method that progressively adds noise to data and learns to reverse this process, which can effectively capture data distribution for tasks such as images (Yi et al. 2024), text (Arriola et al. 2025), or molecular generation (Wang et al. 2024). In recommendation domain, diffusion models are commonly used to model user preferences. For instance, DiffuRec (Li, Sun, and Li 2023) employs denoising diffusion to model sequential recommendation with item information. DiQDiff (Mao et al. 2025) guides diffusion model through semantic codebook. CCDRec (Yang et al. 2025) proposes a conditioned diffusion for multi-model recommendation with curriculum learning. Moreover, potential of diffusion models for negative sampling has gained attention across different domains. DMNS (Nguyen and Fang 2024) formulates multi-level sampling via conditional diffusion for link prediction, and MMKGC (Niu and Zhang 2025) extends this to knowledge graph by incorporating multi-model semantics.

However, existing diffusion-based samplers typically rely on fixed timesteps and are often tailored to specific domains. In contrast, our method introduces a transition-point-driven sampling, which can adaptively select negative samples and serve as a plug-and-play module to enhance models.

3 Preliminary

3.1 Problem Formulation

We denote the set of historical user-item interactions by $\mathcal{D}^+ = \{(u, i) | u \in \mathcal{U}, i \in \mathcal{I}\}$, where \mathcal{U} and \mathcal{I} are the set of users and the set of items, respectively. We consider all unobserved interactions $\mathcal{D}^- = \{(u, j) | u \in \mathcal{U}, j \notin \mathcal{D}_u^+\}$ as candidates for negative samples. Standard recommendation models optimize a pairwise loss by training on triplets (u, i, j) , where i is a positive item and j is a negative item sampled from \mathcal{D}^- . Our goal is to enhance model performance by introducing extra negative samples generated through diffusion-based perturbation of positive sample i . These samples can simulate hard negatives, thus encouraging model to learn more discriminative representations.

3.2 Diffusion Model

Diffusion model (Ho, Jain, and Abbeel 2020; Dhariwal and Nichol 2021a) runs a forward process to add Gaussian noise to the input, and then performs a reverse process to denoise and reconstruct the original input in an iterative manner. Given an input $\mathbf{x}_0 \sim q(\mathbf{x}_0)$ and total number of diffusion steps T to get subsequent state $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T$, the forward process of the diffusion model can be defined as

$$q(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t} \mathbf{x}_0, (1 - \bar{\alpha}_t) \mathbf{I}), \quad (1)$$

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{(1 - \bar{\alpha}_t)} \epsilon_t, \quad (2)$$

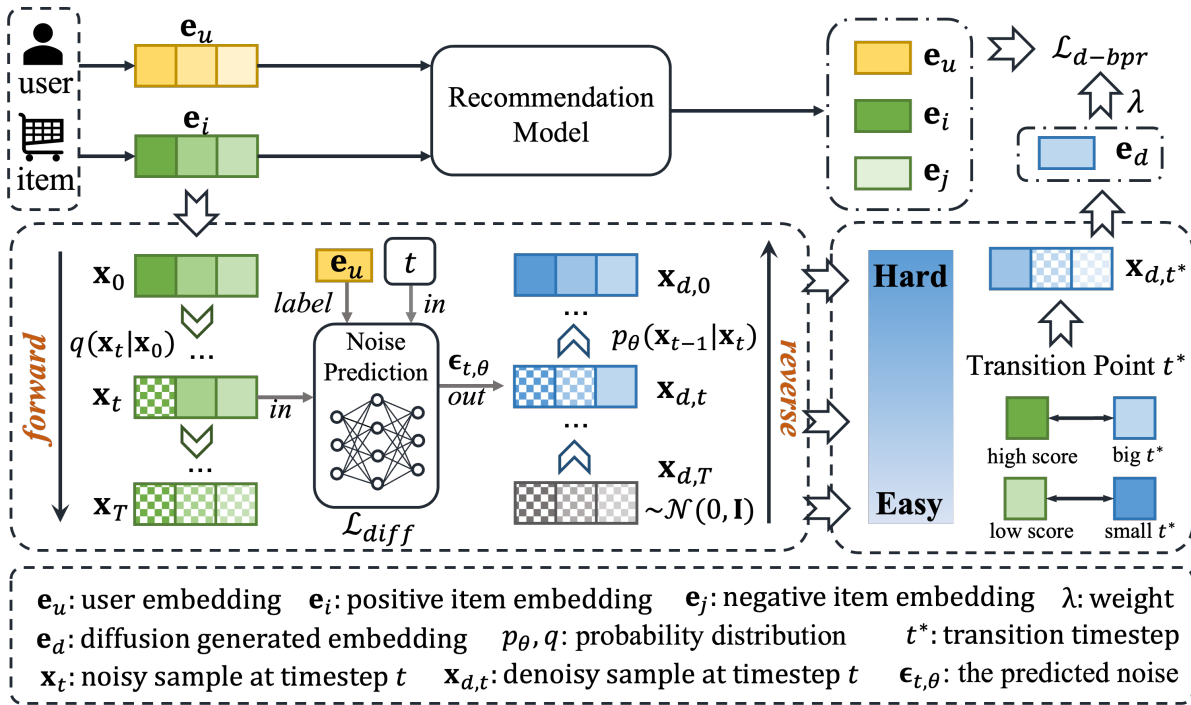


Figure 1: Overview of the proposed ADAR: ADAR leverages the generative dynamics of diffusion models to produce meaningful negatives with controllable hardness. First, we apply a diffusion model to generate negative candidates by gradually corrupting positive samples. Then, we determine the transition point t^* through a score-aware function to adaptively choose the augmented negative. ADAR can serve as a plug-and-play module to enhance diverse recommendation models.

where cumulative signal retention $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$, $\alpha_t = 1 - \beta_t$, and $\beta_1, \beta_2, \dots, \beta_T$ controls the amount of added noise. $t \in [0, T]$ represents the index of timesteps. ϵ_t represents the noise sampled from $\mathcal{N}(0, \mathbf{I})$. Typically, β_t increases monotonically over time according to a predefined schedule (e.g., linear, cosine or sigmoid), resulting in a gradual increase in noise level.

And the reverse process uses a neural network with parameter θ to predict the noise added in the forward process. The reverse transformations can be formulated as:

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \Sigma_\theta(\mathbf{x}_t, t)). \quad (3)$$

In order to model $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$, DDPM (Ho, Jain, and Abbeel 2020) usually fixes $\Sigma_\theta(\mathbf{x}_t, t)$ in advance, and use a reparameterization trick to derive $\mu_\theta(\mathbf{x}_t, t)$:

$$\mu_\theta(\mathbf{x}_t, t) = \frac{1}{\sqrt{\alpha_t}} \mathbf{x}_t - \frac{1 - \alpha_t}{\sqrt{\alpha_t} \sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(\mathbf{x}_t, t), \quad (4)$$

where $\epsilon_\theta(\mathbf{x}_t, t)$ is the predicted target, which should be similar to Gaussian noise. In actual operation, we start from $\mathbf{x}_T \sim \mathcal{N}(0, \mathbf{I})$, and then restore \mathbf{x}_0 step by step.

4 Method

We propose Adaptive Diffusion-based Augmentation for Recommendation (ADAR), a model-agnostic module for generating high-quality negative samples. ADAR simulates a gradual degradation of positive samples through a diffusion process and we detail the formulation of this process. ADAR’s main architecture is shown in Figure 1.

4.1 Diffusion-based Sample Generation

In this section, we describe how ADAR generates candidate negative samples. ADAR learns the progressive noising and denoising process, enabling controllable negative sample generation for recommendation.

Let i in $(u, i) \in \mathcal{D}^+$ be a positive sample. The corresponding embedding $e_i \in \mathbb{R}^d$ is first obtained from an encoder $\phi(i)$, which may be derived from any different model according to the specific recommendation tasks. The user representation e_u is similarly obtained from $\phi(u)$.

Forward Process. We apply the forward diffusion process according to Eq.2, producing a sequence of progressively noised embeddings. Specifically, we set \mathbf{x}_0 equal to e_i , perform the forward diffusion process, and obtain noisy sample set $\mathcal{X} = \{\mathbf{x}_t\}_{t=0}^T$. This process gradually removes positive information from the original interaction, pushing \mathbf{x}_t toward a standard Gaussian distribution.

Noise Prediction Model. To generate candidate negative samples, we need to approximate the noise added at each time step, which can be formulated as

$$\epsilon_{t,\theta} = \tau(\mathbf{x}_t, \mathbf{e}_t, \mathbf{e}_u; \theta), \quad (5)$$

where $\mathbf{e}_t = \text{PE}(t)$ is the time embedding, and τ is the noise prediction model with learnable parameter θ . Meanwhile, in order to make full use of the user embedding e_u , we use the conditional diffusion model (Dhariwal and Nichol 2021b; Ho et al. 2022) to input the user embedding as label information to guide the generation.

We adopt sine and cosine coding to construct a continuous, periodic time embedding, following the approach in (Ho, Jain, and Abbeel 2020). The timestep embedding is formulated as:

$$\text{PE}(t) = \begin{cases} \sin(t \cdot 10000^{-\frac{2i}{d_t}}), & \text{if } i \text{ is even} \\ \cos(t \cdot 10000^{-\frac{2i}{d_t}}), & \text{if } i \text{ is odd} \end{cases}, \quad (6)$$

where $i \in [0, d - 1]$ is the dimension index. This encoding method facilitates the model’s learning of relative timestep relationships.

To better adapt to recommendation tasks, we implement τ using feature-wise linear modulation (FiLM) (Perez et al. 2018), which dynamically modulates intermediate feature representations based on timestep and label embeddings. This approach enables efficient noise prediction while ensuring a lightweight adaptation to structured recommendation data. Specifically, τ is formulated as:

$$\epsilon_{t,\theta} = \gamma(\mathbf{e}_t, \mathbf{e}_u; \theta_\gamma) \odot \mathbf{x}_t + \eta(\mathbf{e}_t, \mathbf{e}_u; \theta_\eta), \quad (7)$$

where γ and η are interpreted as scale and offset, respectively, and \odot represents element-wise multiplication. In our model, these components are implemented as a simple multilayer perceptron with multiple fully connected layers.

Diffusion Loss. Following the standard formulation of diffusion models (Dhariwal and Nichol 2021a), we employ a mean squared error (MSE) loss to minimize the discrepancy between the predicted noise in the reverse process and the Gaussian noise injected during forward diffusion. However, in recommendation settings, recovering the noise alone does not ensure that the final generated samples retain the semantic essence of the original positive instances (Li, Sun, and Li 2023). To address this, we incorporate an additional loss that enforces consistency between the final generated output and its corresponding positive sample. Specifically, for a given timestep t , the diffusion loss function is defined as:

$$\mathcal{L}_{diff} = \|\epsilon - \epsilon_{t,\theta}\|^2 + \|\mathbf{x}_0 - \mathbf{x}_{d,0}\|^2. \quad (8)$$

Reverse Process. After the noise prediction model τ is trained, we employ it to generate candidate samples via the reverse diffusion process, where each timestep output serves as a potential negative sample. Specifically, we start from a Gaussian noise sampled from the standard normal distribution, representing a maximally corrupted version in the latent space:

$$\mathbf{x}_{d,T} \sim \mathcal{N}(0, \mathbf{I}), \quad (9)$$

and progressively apply the reverse update rule:

$$\mathbf{x}_{d,t-1} = \frac{1}{\sqrt{\alpha_t}} \mathbf{x}_{d,t} - \frac{1 - \alpha_t}{\sqrt{\alpha_t} \sqrt{1 - \alpha_t}} \epsilon_{t,\theta}, \quad (10)$$

for $t = T, T - 1, \dots, 1$. Each denoising step incrementally restores information, resulting in the denoised sample set $\mathcal{X}_d = \{\mathbf{x}_{d,t}\}_{t=0}^T$. These outputs are novel, plausible candidates resembling corrupted versions of positive items.

4.2 Transition Point Detection

However, not all generated samples in \mathcal{X}_d are equally informative for training. Samples that are too similar to the original positives may reinforce redundancy, while those that are

too noisy may introduce harmful signals. To address this, we propose an adaptive mechanism to identify the transition point, the moment when a generated sample shifts from being positive to negative in terms of user preference. We select the generated sample at the transition point as the augmented negative sample, which lies near the decision boundary, and thus offer maximum utility for training.

Theorem 1. *Let μ^+ be the positive item score. Then the cumulative signal retention $\bar{\alpha}_{t^*}$ value at the transition point t^* is negatively correlated with μ^+ .*

Proof. We begin by modeling the user preference score as a function $f(\mathbf{u}, \mathbf{x})$, where \mathbf{u} and \mathbf{x} denote the user and item representations, respectively (with $\mathbf{u} = \mathbf{e}_u$). As we agreed before, \mathbf{x}_0 denote a positive sample, and \mathbf{x}_t its noised version at diffusion step t . Under the assumption that the scoring function $f(\mathbf{u}, \mathbf{x})$ is approximately linear in \mathbf{x} (e.g., inner product models), the expected preference score of the noised item \mathbf{x}_t becomes:

$$\begin{aligned} \mathbb{E}[f(\mathbf{u}, \mathbf{x}_t)] &= \mathbb{E}[f(\mathbf{u}, \sqrt{\alpha_t} \mathbf{x}_0 + \sqrt{(1 - \alpha_t)} \epsilon_t)] \\ &= \sqrt{\alpha_t} \cdot f(\mathbf{u}, \mathbf{x}_0) \\ &= \sqrt{\alpha_t} \cdot \mu^+, \end{aligned}$$

where $\mu^+ = f(\mathbf{u}, \mathbf{x}_0)$ is the score of the original positive item, which is assumed to be greater than zero to reflect user preference. We define the transition point t^* as the first diffusion step at which the expected user score for \mathbf{x}_t falls below a predefined negative threshold μ^- . That is,

$$\mathbb{E}[f(\mathbf{u}, \mathbf{x}_{t^*})] = \mu^- \quad \Rightarrow \quad \bar{\alpha}_{t^*} = \left(\frac{\mu^-}{\mu^+} \right)^2.$$

Taking derivative w.r.t. μ^+ :

$$\frac{d\bar{\alpha}_{t^*}}{d\mu^+} = -2 \left(\frac{\mu^-}{\mu^+} \right)^2 \cdot \frac{1}{\mu^+} < 0,$$

demonstrating that $\bar{\alpha}_{t^*}$ is strictly decreasing in μ^+ . \square

Since $\bar{\alpha}_t$ value is a strictly decreasing function of t , it follows that t^* must increase with μ^+ . That is, positive samples with higher scores retain their information identity longer during diffusion, and undergo transition at later steps.

While this formulation provides theoretical guidance, practical datasets lack explicit negative interactions, making μ^- unobservable. To circumvent this, we adopt a parameterized heuristic based on the observed positive score $p_s = f(\mathbf{u}, \mathbf{x}_0)$ to approximate t^* . Specifically, we define a score-aware function:

$$t^* = \text{sigmoid}(\omega \cdot \exp(k \cdot p_s)) \cdot T, \quad (11)$$

where $w > 0$ and $k > 0$ are tunable hyperparameters that control the shape and scale of the transition function. In practice, we can set $\omega = 1$ and $k = 1$, which consistently lead to performance improvements across most datasets, as verified by extensive experiments.

This score-aware function aligns with our theoretical result, capturing the negative correlation between the positive score and the cumulative signal retention $\bar{\alpha}_{t^*}$ value. By

adaptively mapping higher p_s to deeper diffusion step t^* , it dynamically adjusts the distance of generated samples according to the observed positive score.

After determining the transition point t^* , we then select the sample \mathbf{x}_{d,t^*} at timestep t^* from the denoised sample set \mathcal{X}_d as the final augmented negative.

4.3 Optimization Objectives

To optimize the recommendation model, we adopt BPR (Rendle et al. 2012), a widely established objective for pairwise learning from implicit feedback. Given that ADAR operates as a model-agnostic augmentation module, it can be seamlessly integrated into existing recommendation models without modifying their core structure.

In our formulation, let \mathbf{e}_j denote the item embedding selected via the baseline negative sampling method, and \mathbf{e}_d denote the embedding of the sample generated through ADAR. To fully exploit the information of generated negatives, we integrate them into the learning objectives as a supplement. Specifically, the total loss is given by

$$\mathcal{L}_{d-bpr} = - \sum_{u,i,j,d} \ln \sigma(\mathbf{e}_u^\top \mathbf{e}_i - (\mathbf{e}_u^\top \mathbf{e}_j + \lambda \mathbf{e}_u^\top \mathbf{e}_d)), \quad (12)$$

where the weighting factor $\lambda \in [0, 1]$ is a hyperparameter that regulates the contribution of the generated samples. ADAR introduces negative samples with controlled hardness, which encourages the model to construct more discriminative representations. We adopt an alternating training strategy between the diffusion model and the encoder. For a comprehensive understanding of the training procedure of ADAR, the training pseudo-code is provided in Algorithm 1.

Algorithm 1: ADAR

```

1: Input: Set of Implicit Feedback  $\mathcal{D}^+ = \{(u, i) | u \in \mathcal{U}, i \in \mathcal{I}\}$ , predefined hyperparameters  $T$  and  $\lambda$ .
2: Output: Encoder model, Diffusion model.
3: Initialization
4: while not converged do
5:   for each mini-batch  $\mathcal{B}$  sampled from  $\mathcal{D}^+$  do
6:     Get embeddings of user  $u$  and positive item  $i$ 
7:     // Diffusion-based sample generation
8:      $\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{(1 - \bar{\alpha}_t)} \epsilon_t, \epsilon_t \sim \mathcal{N}(0, \mathbf{I})$ 
9:      $\mathbf{e}_t = \text{PE}(t)$ 
10:    Optimization Diffusion by minimizing  $\mathcal{L}_{diff}$ 
11:    Iteratively compute the reverse diffusion step following  $\mathbf{x}_{d,t-1} = \frac{1}{\sqrt{\bar{\alpha}_t}} \mathbf{x}_{d,t} - \frac{1 - \bar{\alpha}_t}{\sqrt{\bar{\alpha}_t} \sqrt{1 - \bar{\alpha}_t}} \epsilon_{t,\theta}$ 
12:    Obtain candidate samples set  $\mathcal{X}_d = \{\mathbf{x}_{t=0}^T\}$ 
13:    // Transition point detection
14:     $t^* = \text{sigmoid}(\omega \cdot \exp(k \cdot p_s)) \cdot T$ 
15:    Obtain embedding of transition point  $\mathbf{x}_{d,t^*}$ 
16:     $\mathbf{e}_d = \mathbf{x}_{d,t^*}$ 
17:    // Train Encoder model
18:    Get triples  $(u, i, j)$  from existing method
19:    Optimization Encoder by minimizing  $\mathcal{L}_{d-bpr}$ 
20:  end for
21: end while

```

Dataset	#Users	#Items	#Interactions	Avg. Length	Density
Beauty	22,363	12,101	0.2m	8.9	0.05%
Toys	19,412	11,924	0.17m	8.6	0.07%
Sport	35,598	18,357	0.3m	8.3	0.05%
Yelp	30,431	20,033	0.3m	8.3	0.05%

Table 1: Statistics of the datasets.

5 Experiments

5.1 Experimental Settings

Datasets. We conduct experiments on four benchmark datasets: Beauty, Toys, Sport, and Yelp, which are commonly used in both collaborative filtering (CF) and sequential recommendation (SR) tasks. All datasets are derived from the Amazon Review dataset and the Yelp Review dataset. The detailed statistics of the four datasets are summarized in Table 1. For CF task, we follow standard preprocessing procedures (Lai et al. 2024) by converting user behavior logs into user-item interaction matrices. Each user’s interactions are randomly split into 80% for training and 20% for testing. For SR task, user interactions are first sorted chronologically into sequences, and we filter out users with fewer than five interactions to ensure modeling stability.

Evaluation Metrics. To assess recommendation performance, we adopt two widely used ranking metrics: Recall@10, 20 and NDCG@10, 20. Recall measures the proportion of relevant items successfully retrieved in the top- k recommendations, while NDCG accounts for the position of the hit items, giving higher weight to correctly ranked items appearing earlier in the list.

Integrated Methods. To evaluate the effectiveness of our proposed ADAR, we conduct experiments under two major recommendation paradigms: collaborative filtering (CF) and sequential recommendation (SR).

For CF, we adopt the following models: NGCF (Wang et al. 2019), LightGCN (He et al. 2020), MixGCF (Huang et al. 2021) and AHNS (Lai et al. 2024). For SR, we consider three representative sequential models with datasets augmented by DR4SR (Yin et al. 2024): GRU4Rec (Hidasi et al. 2015), SASRec (Kang and McAuley 2018) and CL4SRec (Xie et al. 2022). See Appendix for a more detailed description of the baselines.

For each model, we incorporate our ADAR without modifying their core architectures, ensuring a fair and direct comparison with the original baselines.

Implementation Details. The embedding dimension is fixed to 64, and the embedding parameters are initialized with the Xavier initialization (Glorot and Bengio 2010). We optimize all parameters with Adam optimizer (Kingma and Ba 2014) and use the default learning rate of 0.001. For CF, we use the default mini-batch size of 2,048. For SR, we use the default mini-batch size of 256, with the maximum sequence length N set to 50. For ADAR, we fix $\omega = 1$ and $k = 1$, and search λ in the range $[0, 1]$. We set the maximum diffusion step T to 50 for CF and 20 for SR.

Task	Methods	Amazon-Beauty			Amzon-Toys			Amazon-Sport			Yelp		
		R@10	R@20	N@10	R@10	R@20	N@10	R@10	R@20	N@10	R@10	R@20	N@10
CF	NGCF	6.37	9.50	4.11	5.94	8.50	3.92	3.92	6.21	2.50	4.27	6.92	2.73
	+ADAR	8.13	11.76	5.21	7.82	11.39	5.08	5.29	8.11	3.35	5.06	8.19	3.24
	RelImp	27.6%	23.8%	26.8%	31.6%	34.0%	29.6%	34.9%	30.6%	34.0%	18.5%	18.3%	18.7%
	LightGCN	8.69	12.40	5.77	8.22	11.50	5.55	5.96	8.62	3.82	5.27	8.33	3.39
	+ADAR	8.99	12.88	6.05	8.46	11.93	5.84	6.32	9.08	4.11	5.55	8.81	3.58
	RelImp	3.45%	3.55%	4.85%	2.92%	3.74%	5.23%	6.04%	5.34%	7.59%	5.31%	5.76%	5.60%
	MixGCF	9.82	13.67	6.63	9.05	12.82	6.19	6.25	8.98	4.21	5.80	9.01	3.72
	+ADAR	10.01	14.10	6.90	9.25	13.09	6.39	6.31	9.13	4.32	5.95	9.31	3.86
	RelImp	1.93%	3.15%	4.07%	2.21%	2.11%	3.23%	0.96%	1.67%	2.61%	2.59%	3.33%	3.76%
	AHNS	9.66	13.50	6.51	8.89	12.55	6.06	6.69	9.70	4.39	5.53	8.88	3.60
+ADAR	10.05	14.04	6.89	9.15	12.91	6.21	6.84	9.85	4.52	5.82	9.19	3.80	
RelImp	4.04%	4.00%	5.84%	2.92%	2.87%	2.48%	2.24%	1.55%	2.96%	5.24%	3.49%	5.56%	
SR	GRU4Rec	2.79	4.87	1.37	2.24	3.71	1.08	1.81	3.17	0.90	2.30	3.88	1.14
	+ADAR	4.34	6.96	2.17	3.78	5.84	1.89	2.55	4.09	1.39	3.21	5.17	1.59
	RelImp	55.6%	42.9%	58.4%	68.8%	57.4%	73.4%	40.9%	29.0%	54.4%	39.6%	33.2%	39.5%
	SASRec	5.99	8.82	3.20	7.27	9.95	4.13	2.97	4.72	1.58	2.99	4.99	1.50
	+ADAR	6.32	9.69	3.27	7.72	10.52	4.50	3.62	5.71	1.89	3.30	5.58	1.66
	RelImp	5.51%	9.87%	2.19%	6.19%	5.73%	8.96%	21.9%	21.0%	19.6%	10.4%	11.8%	10.7%
	CL4SRec	7.24	10.26	4.20	7.80	10.59	4.54	4.33	6.19	2.37	3.27	5.58	1.60
	+ADAR	7.44	10.65	4.31	8.12	11.13	4.73	4.46	6.35	2.50	3.69	5.86	1.82
RelImp	2.76%	3.80%	2.62%	4.10%	5.10%	4.19%	3.00%	2.58%	5.49%	12.8%	5.02%	13.8%	

Table 2: Performance on four datasets. RelImp: Relative Improvement. We have conducted the paired t-test to verify that the difference between each base model and our proposed method is statistically significant for $p < 0.05$.

5.2 Result Analysis

In this section, we compare the performance of each target model with our ADAR to verify the efficacy of the proposed module. The evaluation results for recommendations are summarized in Table 2 (more details in Appendix), and we can draw the following conclusions:

- **Effectiveness.** ADAR consistently improves performance across multiple datasets and baseline models. For example, AHNS enhanced by ADAR leads to a relative improvement of 3.13% in Recall@10 and 4.58% in NDCG@10 on Amazon-Beauty. Similar trends are observed across remaining methods, confirming its effectiveness in improving ranking quality.
- **Generality.** ADAR exhibits strong compatibility with diverse model architectures, including CF and SR models. ADAR can be compatible with models such as GRU4Rec (RNN-based), SASRec (attention-based), and CL4SRec (contrastive-based), achieving performance improvements without modifying the underlying model architecture.
- **Complementarity.** Beyond general applicability, ADAR also complements strong existing training objectives. For instance, ADAR improves CL4SRec’s NDCG@10 by 13.8% on Yelp, indicating that it introduces additional beneficial training signals beyond those captured by contrastive learning alone.

Method	Version	R@10	R@20	N@10	N@20
NGCF	base	6.37	9.50	4.11	5.10
	random	7.81	11.59	5.13	6.29
	fixed	7.79	11.50	5.06	6.22
	mixed	7.91	11.63	5.11	6.30
	ADAR	8.13	11.76	5.21	6.34
GRU4Rec	base	2.79	4.87	1.37	1.90
	random	4.00	6.31	1.99	2.45
	fixed	3.37	5.59	1.66	2.22
	mixed	3.51	5.76	1.77	2.33
	ADAR	4.34	6.96	2.17	2.82

Table 3: Performance comparison of different t^* selection strategies on Amazon-beauty.

5.3 Variant Study

In this section, we investigate the impact of different strategies for selecting t^* . Recall that t^* denotes the diffusion step of the final augmented negative, at which the generative sample translate from positive to negative. We compare five variants: (1) base, where no augmentation is applied; (2) random, where t^* is randomly sampled; (3) fixed, where t^* is fixed at a predefined midpoint (we set it as $T/2$); (4) mixed, where model mixes representations from multiple fixed t (e.g., $T/2, T/4, T/8$ and $T/10$), motivated by DMNS (Nguyen and Fang 2024) on link prediction; (5) ADAR, the proposed method, which adaptively identifies t^* based on

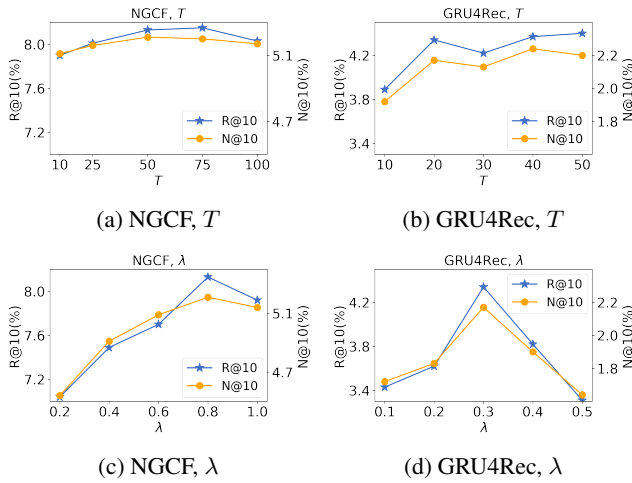


Figure 2: Performance comparison of ADAR *w.r.t.* three different hyperparameter.

theoretical criteria.

The experiments are conducted on Amazon-Beauty, evaluating two representative recommendation models: NGCF and GRU4Rec. We report the results in Table 3, and make the following observations: First, ADAR consistently outperforms all variants on both models, confirming that adaptive selection of t^* yields more informative negatives than other variants. Second, while effective in link prediction, the mixed variant is less suited to recommendation, performing worse than ADAR and even worse than the random variant. This indicates that adaptive, theory-driven transition point is crucial for choosing high-quality negatives.

5.4 Hyperparameter

To further investigate robustness and stability of ADAR, we conduct a hyperparameter analysis on the diffusion step T and weighting factor λ . The analysis of the hyperparameters ω and k can be found in Appendix. In this analysis, we fix the other hyperparameters to their default values to isolate the effect of each individual variable. The experiments are conducted on Amazon-beauty using two representative backbone models: NGCF for collaborative filtering and GRU4Rec for sequential recommendation. As shown in Figure 2(a) and 2(b), the performance initially improves as T increases, but eventually saturates, suggesting that only a moderate number of diffusion steps is sufficient to capture the semantic transition. In contrast, Figure 2(c) and 2(d) show that λ exhibits a clear trade-off behavior, where moderate values consistently lead to better performance. This suggests that incorporating diffusion signal moderately can effectively guide negative sample selection.

5.5 Visualization

In this section, we delve into the visualization analysis to further examine the effectiveness of ADAR in capturing meaningful user preferences. Specifically, we randomly sample 2,000 users from Amazon-Beauty and project their

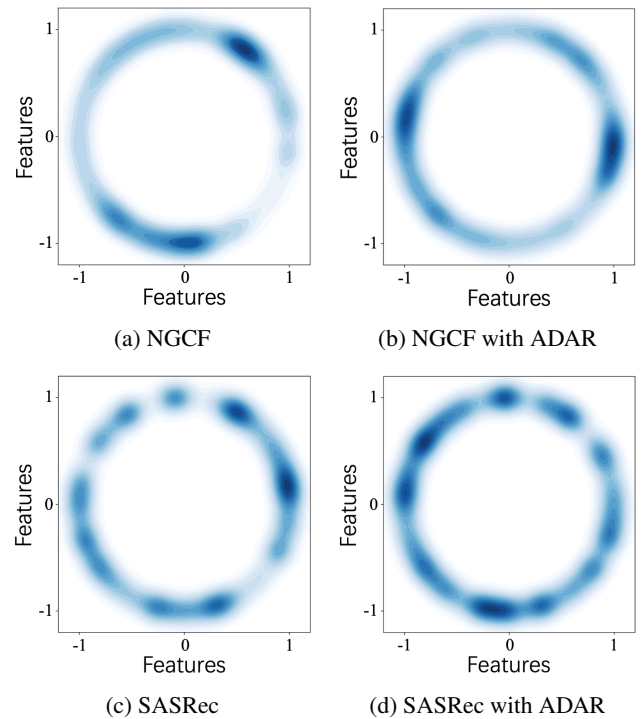


Figure 3: Distribution of user representations learned from Amazon-beauty dataset.

learned user embeddings into a two-dimensional space. This is achieved by normalizing each representation onto the unit hypersphere and applying t-SNE (Maaten and Hinton 2008) for dimensionality reduction. As shown in Figure 3, user representations learned by baseline models (NGCF and SASRec) tend to exhibit uneven and clustered patterns. In contrast, their ADAR-augmented counterparts (NGCF with ADAR and SASRec with ADAR) yield embeddings that are more uniformly distributed across the space. This suggests that incorporating diffusion-based augmentation encourages a more informative and disentangled user representation space, potentially leading to improved generalization.

6 Conclusion

To solve the challenge of unreliable negative signals and limited control over negative sampling in implicit feedback recommendation, we propose ADAR, a diffusion-based augmentation module. ADAR leverages the progressive corruption process of diffusion models to generate high-quality negative samples. By theoretically identifying the transition point from positive to negative, ADAR adaptively selects optimal sampling steps with informative and challenging negatives. Extensive experiments on various backbone models and recommendation tasks demonstrate that ADAR is both effective and broadly applicable, offering a robust solution for enhancing recommendation performance. Our code is publicly available at <https://github.com/LN-Nlaine/ADAR>.

Acknowledgments

This work was supported by Natural Science Foundation of Heilongjiang Province of China (No. LH2024F023).

References

- Arriola, M.; Gokaslan, A.; Chiu, J. T.; Yang, Z.; Qi, Z.; Han, J.; Sahoo, S. S.; and Kuleshov, V. 2025. Block Diffusion: Interpolating between Autoregressive and Diffusion Language Models. In *Proceedings of the Thirteenth International Conference on Learning Representations (ICLR)*.
- Dhariwal, P.; and Nichol, A. 2021a. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34: 8780–8794.
- Dhariwal, P.; and Nichol, A. 2021b. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34: 8780–8794.
- Ding, J.; Quan, Y.; Yao, Q.; Li, Y.; and Jin, D. 2020. Simplify and robustify negative sampling for implicit collaborative filtering. *Advances in Neural Information Processing Systems*, 33: 1094–1105.
- Fan, L.; Pu, J.; Zhang, R.; and Wu, X.-M. 2023. Neighborhood-based hard negative mining for sequential recommendation. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2042–2046.
- Glorot, X.; and Bengio, Y. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 249–256. JMLR Workshop and Conference Proceedings.
- He, X.; Deng, K.; Wang, X.; Li, Y.; Zhang, Y.; and Wang, M. 2020. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, 639–648.
- Hidasi, B.; Karatzoglou, A.; Baltrunas, L.; and Tikk, D. 2015. Session-based recommendations with recurrent neural networks. *arXiv preprint arXiv:1511.06939*.
- Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33: 6840–6851.
- Ho, J.; Saharia, C.; Chan, W.; Fleet, D. J.; Norouzi, M.; and Salimans, T. 2022. Cascaded diffusion models for high fidelity image generation. *Journal of Machine Learning Research*, 23(47): 1–33.
- Huang, T.; Dong, Y.; Ding, M.; Yang, Z.; Feng, W.; Wang, X.; and Tang, J. 2021. Mixgcf: An improved training method for graph neural network-based recommender systems. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 665–674.
- Kang, W.-C.; and McAuley, J. 2018. Self-attentive sequential recommendation. In *2018 IEEE international conference on data mining (ICDM)*, 197–206. IEEE.
- Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Lai, R.; Chen, R.; Han, Q.; Zhang, C.; and Chen, L. 2024. Adaptive hardness negative sampling for collaborative filtering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 8645–8652.
- Li, Z.; Sun, A.; and Li, C. 2023. Diffurec: A diffusion model for sequential recommendation. *ACM Transactions on Information Systems*, 42(3): 1–28.
- Ma, Y.; Chen, M.; and Li, X. 2023. DropMix: Better Graph Contrastive Learning with Harder Negative Samples. In *2023 IEEE International Conference on Data Mining Workshops (ICDMW)*, 1105–1112. IEEE.
- Maaten, L. v. d.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research*, 9(Nov): 2579–2605.
- Mao, W.; Liu, S.; Liu, H.; Liu, H.; Li, X.; and Hu, L. 2025. Distinguished quantized guidance for diffusion-based sequence recommendation. In *Proceedings of the ACM on Web Conference 2025*, 425–435.
- Nguyen, T.-K.; and Fang, Y. 2024. Diffusion-based negative sampling on graphs for link prediction. In *Proceedings of the ACM Web Conference 2024*, 948–958.
- Niu, G.; and Zhang, X. 2025. Diffusion-based hierarchical negative sampling for multimodal knowledge graph completion. *arXiv preprint arXiv:2501.15393*.
- Perez, E.; Strub, F.; De Vries, H.; Dumoulin, V.; and Courville, A. 2018. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.
- Rendle, S.; and Freudenthaler, C. 2014. Improving pairwise learning for item recommendation from implicit feedback. In *Proceedings of the 7th ACM international conference on Web search and data mining*, 273–282.
- Rendle, S.; Freudenthaler, C.; Gantner, Z.; and Schmidt-Thieme, L. 2012. BPR: Bayesian personalized ranking from implicit feedback. *arXiv preprint arXiv:1205.2618*.
- Seol, J.; and Choi, J. 2025. Context-aware Negative Sampling for Sequential Recommendation. *IEEE Access*.
- Wang, X.; He, X.; Wang, M.; Feng, F.; and Chua, T.-S. 2019. Neural graph collaborative filtering. In *Proceedings of the 42nd international ACM SIGIR conference on Research and development in Information Retrieval*, 165–174.
- Wang, Y.; Liu, X.; Huang, F.; Xiong, Z.; and Zhang, W. 2024. A multi-modal contrastive diffusion model for therapeutic peptide generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 3–11.
- Wu, X.; Yang, L.; Gong, J.; Zhou, C.; Lin, T.; Liu, X.; and Yu, P. S. 2023. Dimension independent mixup for hard negative sample in collaborative filtering. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 2785–2794.
- Xie, X.; Sun, F.; Liu, Z.; Wu, S.; Gao, J.; Zhang, J.; Ding, B.; and Cui, B. 2022. Contrastive learning for sequential recommendation. In *2022 IEEE 38th international conference on data engineering (ICDE)*, 1259–1273. IEEE.

- Yang, Y.; Ma, H.; Meng, L.; Xu, S.; Xie, R.; and Meng, X. 2025. Curriculum conditioned diffusion for multimodal recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 13035–13043.
- Yang, Z.; Shao, Z.; Dong, Y.; and Tang, J. 2024. TriSampler: a better negative sampling principle for dense retrieval. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 9269–9277.
- Yi, M.; Zhang, K.; Liu, P.; Zuo, T.; and Tian, J. 2024. DiffRAW: leveraging diffusion model to generate DSLR-comparable perceptual quality sRGB from smartphone RAW images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 6711–6719.
- Yin, M.; Wang, H.; Guo, W.; Liu, Y.; Zhang, S.; Zhao, S.; Lian, D.; and Chen, E. 2024. Dataset regeneration for sequential recommendation. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 3954–3965.
- Zhang, W.; Chen, T.; Wang, J.; and Yu, Y. 2013. Optimizing top-n collaborative filtering via dynamic negative item sampling. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, 785–788.
- Zhao, Y.; Chen, R.; Lai, R.; Han, Q.; Song, H.; and Chen, L. 2023. Augmented negative sampling for collaborative filtering. In *Proceedings of the 17th ACM Conference on Recommender Systems*, 256–266.
- Zhou, Y.; Gao, X.; Chen, Z.; and Huang, H. 2025. Attention distillation: A unified approach to visual characteristics transfer. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 18270–18280.