Unleashing the Potential of Diffusion Models Towards Diversified Sequential Recommendations

Zhuo Cai
zhuo.cai@student.uts.edu.au
University of Technology Sydney
Sydney, Australia

Shoujin Wang*
shoujin.wang@uts.edu.au
University of Technology Sydney
Sydney, Australia

Victor W. Chu
wingyan.chu@uts.edu.au
University of Technology Sydney
Sydney, Australia

Usman Naseem
usman.naseem@mq.edu.au
Macquarie University
Sydney, Australia

Yang Wang
yang.wang@uts.edu.au
University of Technology Sydney
Sydney, Australia

Fang Chen
fang.chen@uts.edu.au
University of Technology Sydney
Sydney, Australia

Abstract

Sequential recommender systems (SRSs) aim to recommend the next items to well match users' preferences. In addition to recommendation accuracy, diversity is another critical aspect in evaluating SRSs. Recently, the emerging diffusion models (DMs) have been widely adopted in SRSs. Their employed learning-to-generate paradigm allows them to cover a much broader range of users' preferences and thus generate more diversified items. However, existing DM-based SRSs still face two significant gaps that prevent them from further improving the recommendation diversity: (1) they often rely on **non-diversified users' preferences as guidance** to direct the training of diffusion networks, restricting networks' ability to generate diverse items; and (2) they are based on **a homogeneous diffusion inference mechanism** to generate the next items and thus can only accommodate users' major preferences. Such a practice neglects users' heterogeneous preferences towards various types of items, further limiting recommendation diversity. To bridge these two critical gaps and to further unleash the potential of DMs in enhancing the recommendation diversity of SRSs, we propose a novel diversity-guided diffusion model for sequential recommendations, called *DiffDiv* for short. To be specific, first, a new diversity-aware guidance learning mechanism is devised to direct the training of DMs to effectively capture users' diversified preferences from their historical interactions. Then, a novel heterogeneous diffusion inference mechanism is designed to generate diversified items to accommodate users' heterogeneous preferences, further boosting the recommendation diversity. Extensive experiments on real-world datasets validate the effectiveness of DiffDiv in terms of both recommendation accuracy and diversity.

*Corresponding author

Conference'17, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

CCS Concepts

• **Information systems** → **Retrieval tasks and goals**.

Keywords

Sequential Recommendations, Diversity, Diffusion Models

ACM Reference Format:
Zhuo Cai, Shoujin Wang, Victor W. Chu, Usman Naseem, Yang Wang, and Fang Chen. 2025. Unleashing the Potential of Diffusion Models Towards Diversified Sequential Recommendations. In . ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Sequential recommender systems (SRSs) aim to predict the next item a user may want to interact with according to their historical interactions with items [18, 36, 53]. Despite remarkable progress, most of the existing SRSs primarily focus on enhancing recommendation accuracy, neglecting recommendation diversity [49, 61]. This will lead to significant issues, such as filter bubbles [32], where users are confined to a narrow range of items they are familiar with. This significantly reduces the utility of recommendations as well as user experience. Therefore, it is of great significance to develop novel SRSs to effectively generate diversified recommendations [61].

Although some approaches have been proposed to enhance the recommendation diversity of SRSs [6, 25, 47], they generally follow the learning-to-classify paradigm to match candidate items with users' recently interacted items. Such practice inevitably push them to prioritize recommending items which are similar to those ones which have been already interacted with by users [60], limiting their ability to effectively capture users' diverse preferences.

As an emerging type of generative models, diffusion models (DMs) have been increasingly adopted in SRSs in recent years [24, 60]. Specifically, DM-based SRSs generate the next items for users from scratch (e.g., from Gaussian noise) during a progressive process [24, 60]. This particular process enables them to discover a broader range of users' preferences, thereby providing more diversified recommendation results compared to traditional diversified SRSs [24]. Generally, existing DM-based SRSs can be divided into two classes: (1) approaches for target item generation [24, 33, 60], and (2) approaches for augmented data generation [26, 29, 56]. The

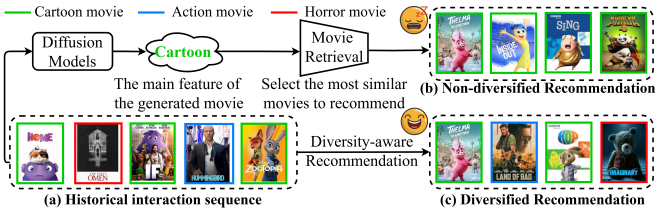

Figure 1: A toy example in movie recommendation scenario. From (a) to (b) demonstrates how existing DM-based SRSs generate recommendations. The process from (a) to (c) illustrates what a diversified SRS is supposed to achieve.

former directly generates the next items that match users' preferences while the latter generate additional data to enhance the learning of users' preferences. This paper focuses on the former.

Typically, a DM-based SRS involves two main stages: (1) Stage 1: diffusion model training, where users' preferences learned by a sequence encoder from their historical interactions are utilized to guide the model training process; (2) Stage 2: next-item inference, where the diffusion inference is employed for generating the next item with the trained diffusion model [60]. However, existing DM-based SRSs have significant gaps in both stages, preventing them from delivering more diversified recommendations.

Gap 1 (In Stage 1): Non-diversified users' preferences-guided diffusion model training. Existing DM-based SRSs typically utilize users' preferences learned by sequence encoders as a guidance to guide the training of DMs [24, 33, 60]. However, the learned users' preferences are encoded in the form of deterministic embeddings (i.e., fix-point vectors). Deterministic embeddings often express limited scope of information embedded in users' interaction sequences, making them insufficient for capturing users' diverse preferences [24]. Therefore, relying on such non-diversified users' preferences as the guidance to train DMs hinders the ability of existing DM-based SRSs to generate diversified recommendations. The restricted ability of deterministic embeddings to learn data diversity has also been demonstrated in other domains, such as image recognition [34] and cross-modal retrieval [9].

Gap 2 (In Stage 2): Homogeneous diffusion inference mechanism for generating the next item. Existing DM-based SRSs employ a homogeneous diffusion inference mechanism, i.e., generating a single item embedding to best match a user's major preference towards a certain type of items only, while overlooking their preferences towards other types of items. Subsequently, the top several candidate items that are similar to the generated item are recommended to the user [60]. As a result, such an inference mechanism causes existing DM-based SRSs to recommend non-diversified items which are all highly similar to the single generated item. For instance, the process from (a) to (b) in Figure 1 illustrates the core idea of current DM-based SRSs. Suppose a user has recently watched three cartoon movies, one action movie, and one horror movie. In this case, cartoon elements (i.e., major preference) dominate the watching sequence, guiding DMs to generate a cartoon movie. As a result, a movie list dominantly comprised of cartoon movies will be selected as the recommendation result according to this generated one. Clearly, such recommendation neglects user's heterogeneous and diversified preferences for action and horror movies.

To bridge these two significant gaps to further unleash the capability of DMs in enhancing the diversity of SRSs, in this work, we propose a novel diversity-guided diffusion model for SRSs, termed *DiffDiv*. To be specific, to tackle ***Gap 1***, we design a new diversity-aware guidance learning module (DAGL). DAGL first employs a sequence encoder to extract users' preferences from their historical interaction sequences. Afterwards, a new diversity-aware guidance is constructed by capturing uncertainty and variations in users' preferences from a probabilistic perspective. Benefiting from such diversity-aware guidance, DiffDiv can be trained to comprehensively learn users' more diversified preferences. In addition, to avoid improving diversity at an unacceptable cost to accuracy in this process, we further design a new accuracy-diversity balanced optimization strategy (ADBO). ADBO is built on controllable robust divergences to strike a balance between recommendation accuracy and diversity. To overcome ***Gap 2***, we devise a novel heterogeneous diffusion inference mechanism (HDI) to capture heterogeneous user preferences. Specifically, HDI generates various types of items during the diffusion inference process guided by diversified preferences learned from users' historical interactions. Each generated item embedding will result in a particular type of items to be selected to accommodate users' certain preference. Finally, the various generated item embeddings lead to the recommendation of multi-type items to facilitate users' diversified preferences. Such a novel design further boosts the capability of DiffDiv in providing more diversified recommendation results.

The main contributions of this paper are summarized as follows:

- To further unleash the potential of DMs in enhancing the recommendation diversity of SRSs, we propose a novel diversity-guided diffusion model, termed DiffDiv.
- A new diversity-aware guidance learning module (DAGL) is designed to construct a diversity-aware guidance by capturing the uncertainty and diversity of users' preferences from a probabilistic perspective. Such guidance directs the training of DiffDiv towards more diversified generation.
- A new accuracy-diversity balanced optimization strategy (ADBO) is devised to strike a balance between recommendation accuracy and diversity through controllable robust divergences.
- A new heterogeneous diffusion inference (HDI) is devised to generate multi-type items to accommodate users' diversified preferences.

Extensive experiments on two real-world datasets validate the superiority of DiffDiv over representative and/or state-of-the-art methods in terms of both recommendation accuracy and diversity.

2 Related Work

2.1 Sequential Recommendations

Sequential recommender systems (SRSs) have been widely explored due to its significant real-world application values [41, 42, 68]. Sequential recommendation methods can be technically categorized into three categories: traditional sequential models, latent representation models, and neural network-based models [44, 46]. Traditional sequential models utilize sequential pattern mining or Markov chain models to model item dependencies in a sequence [13]. While simple, they cannot model long- and short-term dependencies. Latent representation models learn latent factors for

users and items, typically through techniques like matrix or tensor factorization [15]. However, they can only capture simple relationships. In recent years, many kinds of neural networks have been applied to SRSs. Within these, recurrent neural networks (RNNs) [53], convolutional neural networks (CNNs) [62], and Transformers [18, 36] are the most commonly used ones that capture complex interaction relationships. Additionally, some works also explore graph neural networks (GNNs) for SRSs, representing items in sequences as nodes in a graph and learn user-item features through graph-based aggregation [5, 65]. Recently, researchers have begun extending DMs to SRSs, leveraging DMs to generate the target item or augment data [24, 56, 60].

2.2 Diversified Recommendations

Diversity in recommender systems can be categorized into individual diversity and aggregate diversity, focusing on recommendation list for each individual user or all users as a whole, respectively [61, 67]. In this paper, we focus on the former. The diversified recommendation methods can be classified into multi-stage and one-stage approaches based on whether the recommendation lists are generated iteratively or all at once. Many diversified methods employ a re-ranking strategy, such as a greedy algorithm (e.g., maximal marginal relevance (MMR) [3]). MMR recommends items one-by-one and regulates that current items are not too similar with previously recommended ones [69]. These methods belong to the multi-stage category, as they iteratively adjust the recommendation lists rather than generating them at once. Some learning-to-rank methods incorporate diversity-aware objectives into the training process to guide models towards high diversity during training [21]. Clustering-based methods recommend items from various clusters, assuming that similar items will be grouped into the same cluster [2]. These methods belong to the one-stage category, producing the recommendation results in a single step.

In SRSs, diversification methods can also be divided into multi-stage paradigm [4, 6, 7, 27] and one-stage paradigm [25, 39, 47]. Since the multi-stage paradigm requires iterative adjustments to the recommendation lists and thus will significantly increase time consumption, this paper focuses on the one-stage paradigm.

Existing diversified SRSs generally follow a learning-to-classify paradigm and thus recommend items highly similar to those items which have already been interacted with by users. This significantly limits their ability to provide diversified recommendations.

2.3 Diffusion Models for SRSs

With the remarkable success of DMs in image synthesis [10, 16], recent research has explored extending DMs to various recommendation tasks, including Top-K recommendations [51, 66] and multimodal recommendations [22, 31]. In SRSs, DM-based methods can be categorized into two main types: models for target item generation [24, 33, 60] and models for sequence data augmentation [26, 29, 56]. The former one relies on users' historical sequential interactions as guidance to direct the generation of the target item [17, 20, 23, 30, 50, 52, 57]. For example, some methods employ Transformers to capture users' overall preferences and use the output of Transformers to guide the reverse phrase of DMs [24, 60]. Models for sequence data augmentation leverage DMs to generate additional items and augment the original sequences, alleviating the sequence sparsity problem [26, 29, 56].

Although DMs have significant potential to enhance recommendation diversity, existing DM-based SRSs tend to prioritize improving the recommendation accuracy, frequently neglecting the crucial aspects of modeling recommendation diversity. This gap motivates us to develop DM-based methods that aim to improve both recommendation accuracy and diversity in SRSs.

3 Problem Formulation

The sequential recommendation task aims to leverage users' historical interaction sequences to predict what they are likely to engage with next [40, 43, 45]. The whole sequence set is denoted as $\mathcal{S} = \{s_1, ..., s_{|S|}\}$, where each sequence $s = \{i_1, ..., i_n\}$ ($s \in \mathcal{S}$) contains n items ordered by the timestamps at which they were interacted with by a user. The number of items is M. The sequence longer than n will be truncated while shorter ones will be padded. The task involves utilizing previous $n-1$ items ($\{i_1, \cdots, i_{n-1}\}$, also called context items) and their embeddings $\{\mathbf{e}_1, \cdots, \mathbf{e}_{n-1}\}$ to train a model to predict the target item i_n (i.e., the last item of the sequence). i_n is available during the training stage for model optimization and is unavailable during the test stage, where it needs to be predicted. For clarity and brevity, we use a single sequence to explain our model throughout the paper.

4 The DiffDiv Model

As shown in Figure 2, DiffDiv consists of four main components: (1) sequence data processing, (2) diversity-guided diffusion training, (3) heterogeneous diffusion inference, and (4) item recommendations. First, given a user's sequentially interacted items, which are divided into context items and a target item, we map them into embedding vectors. Then, the diversity-aware guidance learning module extracts diversity-aware guidance from the user's context items to guide the training of diffusion network towards more diversified generation. During this process, the accuracy-diversity balanced optimization is applied to balance recommendation accuracy and diversity. Subsequently, the heterogeneous diffusion inference generates various types of items based on the trained diffusion network. Finally, item recommendations that align with the user's diversified preferences are produced based on the generated items.

4.1 Diversity-guided Diffusion Training

4.1.1 Diversity-aware guidance learning. At the beginning of diversity-guidance learning, a sequence encoder is applied to learn user's overall preferences from his/her sequentially interacted items:

$$\mathbf{c} = \textbf{SeqEncoder}(\{\mathbf{e}_1, \cdots, \mathbf{e}_{n-1}\}), \tag{1}$$

where $\mathbf{e} \in \mathbb{R}^d$ refers to the embedding of an item in a sequence. $\mathbf{c} \in \mathbb{R}^d$ denotes contextual embedding that represents the user's overall preferences implied in a sequence. d is the embedding dimension. We select the Transformer as the **SeqEncoder** in this paper, which has shown promising performance in recent research [18, 36]. The Transformer architecture applied in this paper follows the standard configuration [1], including multi-head attention, position-wise feed-forward network, layer normalization and dropout. Since the design of the Transformer network is not the focus of this paper, we omit the details of Transformer network for simplicity.

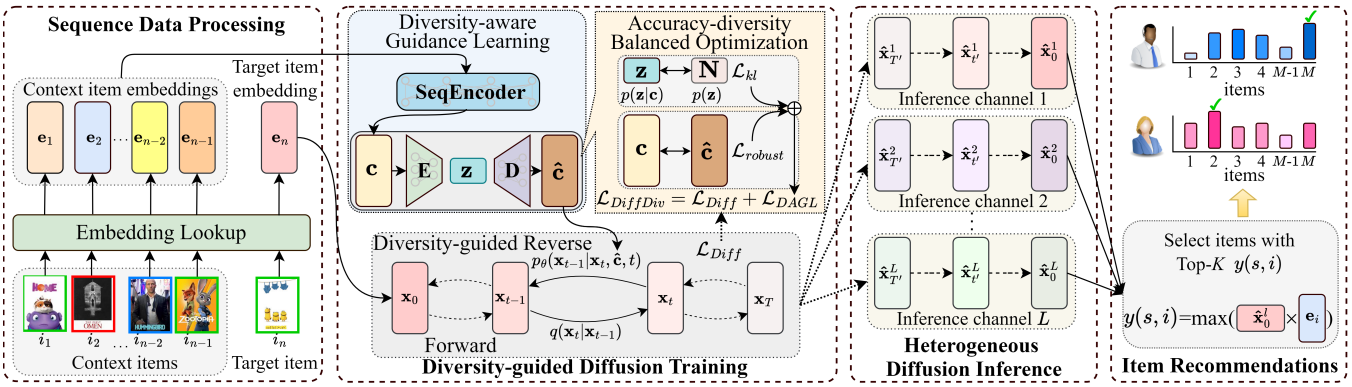

Figure 2: Framework of DiffDiv. DiffDiv begins by utilizing context items to construct a diversity-aware guidance. This guidance then is utilized to guide the direction of diffusion training. To balance accuracy and diversity, an accuracy-diversity balanced optimization strategy is employed. After training, a heterogeneous diffusion inference mechanism generates items that align with users' diversified preferences. Finally, these generated items are utilized to make recommendations.

Existing DM-based SRSs generally utilize $\mathbf{c}$ as the guidance to direct the generation process. However, $\mathbf{c}$ is encoded in the form of deterministic embedding. Such deterministic embedding often expresses limited semantic information in user behaviors, making it insufficient to capture the uncertainty and diversity of users' preferences. As a result, using such non-diversified guidance to direct the item generation fails to produce diversified recommendations.

To tackle this problem, we devise a diversity-aware guidance learning module (DAGL). Previous studies have shown that modeling data in a distribution space can enhance the ability to learn data diversity [9, 34, 55]. Motivated by these works, we explore to construct a diversity-aware guidance from a probabilistic perspective. Formally, we first learn a latent variable $\mathbf{z}$ from $\mathbf{c}$:

$$q(\mathbf{z}|\mathbf{c}) = \mathcal{N}(\mathbf{z}; \boldsymbol{\mu}_\phi(\mathbf{c}), \boldsymbol{\sigma}^2_\phi(\mathbf{c})), \tag{2}$$

where $\boldsymbol{\mu}_\phi(\mathbf{c})$ and $\boldsymbol{\sigma}^2_\phi(\mathbf{c})$ are mean and variance estimated by a encoder network ($\mathbb{R}^d \rightarrow \mathbb{R}^{d'}$, where d' is the dimension of latent layer) with parameters ϕ. Then, the latent variable can be derived through a reparameterization trick: $\mathbf{z} = \boldsymbol{\mu}_\phi(\mathbf{c}) + \boldsymbol{\sigma}_\phi(\mathbf{c})\boldsymbol{\epsilon}$ for $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Subsequently, a decoder network ($\mathbb{R}^{d'} \rightarrow \mathbb{R}^d$) with parameter Ψ is applied to get the diversity-aware guidance: $\hat{\mathbf{c}} = f_\Psi(\mathbf{z}) \in \mathbb{R}^d$.

Our constructed probabilistic guidance $\hat{\mathbf{c}}$ offers higher diversity compared to $\mathbf{c}$, since it represents each sequence information as a probability distribution rather than a deterministic embedding. This probabilistic nature allows for multiple possible embeddings to be sampled from the distribution, capturing different variations and uncertainties in the sequence.

To optimize DAGL's parameters, the following loss is applied:

$$\mathcal{L}_{DAGL} = \underbrace{-\mathbb{E}_{\mathbf{z}}[logp(\hat{\mathbf{c}}|\mathbf{z})]}_{\text{reconstruction term}} + \underbrace{D_{KL}[q(\mathbf{z}|\mathbf{c})||p(\mathbf{z})]}_{\text{prior matching term}}, \tag{3}$$

where D_{KL} is the KL-divergence between two distributions.

Discussion: The proposed DAGL module is straightforward yet effective. More complex and sophisticated network architectures could also be employed for this purpose. We leave it to our future work. To note that although DMs can also represent embeddings in the probabilistic form, we do not use DMs to construct $\hat{\mathbf{c}}$. This is because the inference of DMs requires multiple steps and would be applied at each reverse step of DiffDiv. Consequently, this nesting structure would lead to exponential increase in time consumption.

4.1.2 Diversity-guided diffusion. After constructing the diversity-aware guidance, we next illustrate how to utilize it to guide the training direction of DiffDiv to learn more diversified users' preferences. To improve training efficiency, DiffDiv implements the diffusion process on the embedding of target item i_n. Specifically, let $\mathbf{x}_0 \leftarrow \mathbf{e}_n \in \mathbb{R}^d$ denotes the initial variable of the forward stage.

Forward Stage. Given the initial variable $\mathbf{x}_0 \sim q(\mathbf{x}_0)$, DiffDiv progressively adds Gaussian noise into $\mathbf{x}_0$ at each step until reaching the maximum diffusion step T. Formally, the transition is as follows:

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}), \tag{4}$$

where $\mathbf{x}_t$ and $\mathbf{x}_{t-1}$ are variables at step t and $t-1$. $\mathcal{N}(\mathbf{x}_t; \boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ indicates that $\mathbf{x}_t$ follows a Gaussian distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma}^2$. β_t controls the degree of noise added at the t-th step, and $\mathbf{I}$ refers to an identity matrix. β_t is pre-defined using a linear schedule, with values ranging between 0.0001 and 0.02 [16].

Following Markov chain principle, $\mathbf{x}_t$ can be derived from $\mathbf{x}_0$:

$$q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\overline{\alpha}_t}\mathbf{x}_0, (1-\overline{\alpha}_t)\mathbf{I}), \tag{5}$$

where $\overline{\alpha}_t = \prod_{t'=1}^{t} \alpha_{t'}$ and $\alpha_{t'} = 1 - \beta_{t'}$. Then, a reparameterization trick is applied: $\mathbf{x}_t = \sqrt{\overline{\alpha}_t}\mathbf{x}_0 + \sqrt{(1-\overline{\alpha}_t)}\boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

Diversity-guided Reverse Stage. In reverse stage, the goal of DiffDiv is to generate more diversified items that reflect various aspects of users' preferences. To this end, the diversity-aware guidance $\hat{\mathbf{c}}$ is utilized to direct the reverse process of DiffDiv:

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, \hat{\mathbf{c}}) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, \hat{\mathbf{c}}, t), \Sigma_\theta(\mathbf{x}_t, \hat{\mathbf{c}}, t)), \tag{6}$$

where $\Sigma_\theta(\mathbf{x}_t, \hat{\mathbf{c}}, t)$ is fixed to $\sigma^2(t) = \frac{1-\overline{\alpha}_{t-1}}{1-\overline{\alpha}_t}\beta_t$ as in previous work [51, 60]. $\boldsymbol{\mu}_\theta(\mathbf{x}_t, \hat{\mathbf{c}}, t)$ is the predicted mean from a network $\boldsymbol{f}_\theta(\cdot)$:

$$\boldsymbol{\mu}_\theta(\mathbf{x}_t, \hat{\mathbf{c}}, t) = \frac{\sqrt{\alpha_t}(1-\overline{\alpha}_{t-1})}{\sqrt{1-\overline{\alpha}_t}}\mathbf{x}_t + \sqrt{\overline{\alpha}_{t-1}}\boldsymbol{f}_\theta(\mathbf{x}_t, \hat{\mathbf{c}}, t), \tag{7}$$

where $\boldsymbol{f}_\theta(\mathbf{x}_t, \hat{\mathbf{c}}, t)$ utilizes a network with parameters θ to predict $\mathbf{x}_0$, based on $\mathbf{x}_t$, the diversity-aware guidance $\hat{\mathbf{c}}$ and step t. For simplification, we use the MLP instead of more complex networks such as U-net [35]. Since user interaction data is much simpler than

image data, using overly complex network architectures can lead to overfitting and unnecessarily increase training time [60].

Diffusion Optimization. To optimize the parameters of the diffusion network, the following learning object is applied:

$$\begin{aligned}\mathcal{L}_{Diff} &= \underbrace{\mathbb{E}_q[-log p_\theta(\mathbf{x}_0|\mathbf{x}_1)]}_{\mathcal{L}_0\text{: reconstruction term}} \\ &+ \sum_{t=2}^{T} \underbrace{\mathbb{E}_q[D_{KL}(q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t,\hat{\mathbf{c}}))]}_{\mathcal{L}_{t-1}\text{: denoising matching term}}.\end{aligned} \tag{8}$$

Note that the prior matching term in the complete loss is omitted because it is a constant and does not influence the optimization process [16]. A step t is sampled for each data point $\mathbf{x}_0$ to minimize the denoising matching term, which can be simplified as the mean square error between $\mathbf{x}_0$ and its prediction $f_\theta(\mathbf{x}_t, \hat{\mathbf{c}}, t)$ at step t:

$$\mathcal{L}_{Diff} = \mathbb{E}_{t\sim U(1,T)}[||\mathbf{x}_0 - f_\theta(\mathbf{x}_t, \hat{\mathbf{c}}, t)||_2^2], \tag{9}$$

where t is sampled from a uniform distribution $\mathrm{U}(1, T)$.

4.1.3 Accuracy-diversity balanced optimization. Although the diversity-aware guidance can significantly improve the model's capacity in capturing data diversity, it may also inevitably introduce some noisy information due to its probabilistic nature. To be specific, as $\hat{\mathbf{c}}$ becomes more diverse, the distance between $\hat{\mathbf{c}}$ and a user's overall preferences $\mathbf{c}$ increases. The guidance that diverges too far from a user's overall preferences can impair recommendation accuracy. Hence, **to improve recommendation diversity while ensuring that the accuracy does not decrease**, we apply robust divergences [11, 19] to replace the reconstruction term in Eq.3:

$$\mathcal{L}_{DAGL} = \underbrace{\mathbb{E}_{\mathbf{z}}[D_{\alpha,\beta,\gamma}(\hat{p}(\mathbf{c})||p(\hat{\mathbf{c}}|\mathbf{z}))]}_{\text{robust divergence term}} + \underbrace{D_{KL}[q(\mathbf{z}|\mathbf{c})||p(\mathbf{z})]}_{\text{prior matching term}}, \tag{10}$$

where $\hat{p}(\mathbf{c})$ refers to the empirical distribution of $\mathbf{c}$. The robust divergence term can be calculated using the following formulas ($\hat{p}(\mathbf{c})||p(\hat{\mathbf{c}}|\mathbf{z})$) is omitted due to space limitation):

$$D_\alpha = \frac{1}{\alpha-1}\log\left(\frac{1}{(2\pi\sigma_\phi^2(\mathbf{c}))^{\alpha/2}}\exp\left(\frac{-1}{2\sigma_\phi^2(\mathbf{c})}||\mathbf{c}-\hat{\mathbf{c}}||_2^2\right)^{1-\alpha}\right), \tag{11}$$

$$D_\beta = -\frac{\beta+1}{\beta}\left(\frac{1}{(2\pi\sigma_\phi^2(\mathbf{c}))^{\beta/2}}\exp\left(\frac{-1}{2\sigma_\phi^2(\mathbf{c})}||\mathbf{c}-\hat{\mathbf{c}}||_2^2\right)^{\beta} - 1\right), \tag{12}$$

$$D_\gamma = \frac{\gamma+1}{\gamma}\left(\frac{1}{(2\pi\sigma_\phi^2(\mathbf{c}))^{\gamma/2}}\exp\left(\frac{-1}{2\sigma_\phi^2(\mathbf{c})}||\mathbf{c}-\hat{\mathbf{c}}||_2^2\right)^{\gamma}\right), \tag{13}$$

where α, β and γ are coefficients for α-, β- and γ-divergence, respectively. The working principle behind robust divergence involves down-weighting the contributions of samples with smaller densities, as the probability densities of outliers (i.e., noisy information introduced by probabilistic guidance) are generally much smaller than those of inliers. By controlling these values, the model can adjust the balance between recommendation accuracy and diversity.

Finally, by combining $\mathcal{L}_{Diff}$ and $\mathcal{L}_{DAGL}$, the accuracy-diversity balanced optimization (ADBO) loss for DiffDiv is:

$$\mathcal{L}_{DiffDiv} = \mathcal{L}_{Diff} + \mathcal{L}_{DAGL}. \tag{14}$$

4.2 Heterogeneous Diffusion Inference Mechanism for Generating Diversified Items

Current DM-based SRSs typically employ a homogeneous inference mechanism to generate a single item that reflects the user's main preference. Then, the top several similar items to this generated item are selected for recommendation. Such a mechanism often results in a highly homogeneous set of selected items, as the generated item reflects only the user's primary preference to a certain type of items, overlooking other preferences to other types.

To address this issue, we design a heterogeneous diffusion inference mechanism (HDI) that contains multiple inference channels, to capture users' multifaceted preferences. Each inference channel is directed by a specific guidance signal learned by the DAGL module (e.g., $\hat{\mathbf{c}}^l$ to direct channel l). In this way, we reinforce each channel to generate items to reflect the users' preference towards one type of items. $\hat{\mathbf{c}}^l$ is constructed by sampling different Gaussian noise to create the latent variable in Eq.2. Due to the nature of probabilistic guidance, the guidance in each channel can express relatively distinct semantic information, thereby reflecting various aspects of users' preferences. In each inference channel, we initialize it with $\hat{\mathbf{x}}_{T'}^l = \boldsymbol{\epsilon}^l \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and follow the reverse chain $\hat{\mathbf{x}}_{T'}^l \rightarrow \hat{\mathbf{x}}_{T'-1}^l \rightarrow \cdots \rightarrow \hat{\mathbf{x}}_0^l$ to obtain the output of the l-th channel:

$$\begin{aligned}\hat{\mathbf{x}}_{t'-1}^l &= \boldsymbol{\mu}_\theta(\hat{\mathbf{x}}_{t'}^l, \hat{\mathbf{c}}^l, t') + \sqrt{\Sigma_\theta(\hat{\mathbf{x}}_{t'}^l, \hat{\mathbf{c}}^l, t')}\boldsymbol{\epsilon}^l \\ &= \sqrt{\overline{\alpha}_{t-1}} f_\theta(\hat{\mathbf{x}}_{t'}^l, \hat{\mathbf{c}}^l, t') + \frac{\sqrt{\alpha_t'}(1-\overline{\alpha}_{t'-1})}{\sqrt{1-\overline{\alpha}_t'}}\hat{\mathbf{x}}_{t'}^l + \sqrt{\frac{1-\overline{\alpha}_{t'-1}}{1-\overline{\alpha}_t'}}\beta_t'\boldsymbol{\epsilon}^l.\end{aligned} \tag{15}$$

Afterwards, we can get a set of item embedding $\{\hat{\mathbf{x}}_0^1, \cdots, \hat{\mathbf{x}}_0^l, \cdots \hat{\mathbf{x}}_0^L\}$, where each represents one certain aspect of the user's preferences. L is the number of inference channels.

In practice, a user may interact with items that align well with one of his/her preferences, rather than all of his/her preferences. Therefore, we apply a max operation to predict the likelihood that a user will interacted with a item next:

$$y(s, i) = \max_{1\le l\le L}(\hat{\mathbf{x}}_0^{l\top} \cdot \mathbf{e}_i), \tag{16}$$

where $y(s, i)$ is a score that quantifies the relevance between item i and the given user whose interaction sequence is s. $\mathbf{e}_i$ refers to the embedding of item i. Finally, the top-K items with the highest scores are recommended to the user.

Discussion: During the experiments, we found that setting the maximum reverse step T' to a value between 3 and 10 for DiffDiv can produce comparable results compared to setting $T' = T$, where T generally ranges from 500 to 2000 in [33, 60]. It may be explained that DiffDiv utilizes the diversity-aware guidance to direct the training, making quicker capture of users' wide range of preferences and thus accelerating the generation process. Therefore, in DiffDiv, we set $T' \ll T$, which makes the inference time of DiffDiv to remain comparable to existing SRSs [33, 60].

5 Experiments

In this section, we conduct extensive experiments to answer the following four research questions:

RQ1: How does DiffDiv perform compared with other competitive methods in terms of accuracy and diversity?

RQ2: How does each component affect the performance of DiffDiv?

Table 1: Statistics of Datasets.

Datasets	#ori_seq	#items	#training	#validation	#test
Zhihu	11,714	4,838	61,641	8,160	7,911
YooChoose	92,090	7,261	427,128	57,061	55,244

RQ3: How does different hyper-parameters affect DiffDiv?
RQ4: How efficient is DiffDiv compared with other methods?

5.1 Experimental Setup

5.1.1 Datasets. We select two commonly used publicly available datasets: Zhihu and YooChoose to verify the effectiveness of our proposed method. **Zhihu** [12] dataset was collected from a knowledge-sharing platform. Users are recommended with a list of Q&A, and they may read what they are interested in. **YooChoose** dataset comes from ACM RecSys Challenge 2015 [1]. It consists of sequences or sessions of click or purchase behaviors.

We follow [59, 60] to preprocess the datasets: first sort all sequences in chronological order and then split them into training, validation and test sets at the ratio of 8:1:1. The statistics of original datasets and the number of sequences in training, validation and test sets after processing are reported in Table 1.

5.1.2 Baselines. We choose competitive methods from four classes:
General SRS methods:

- **GRU4Rec (ICLR'16)** [14] utilizes Gated Recurrent Units (GRUs) to model the sequential relationships in user behaviors.
- **LRURec (WSDM'24)** [63] designs linear and lightweight recurrent units for sequential recommendations.
- **SASRec (ICDM'18)** [18] employs the Transformer to model the dependencies between items in a sequence.
- **BERT4Rec (CIKM'19)** [36] utilizes a bidirectional Transformer architecture to model complex item dependencies.

Contrastive learning-based SRS methods:

- **CL4SRec (ICDE'22)** [58] employs contrastive learning to address the data sparsity problem in SRSs. Transformer is applied as the sequence encoder of CL4SRec in our implement.
- **ContraRec (TOIS'23)** [38] proposes context-target contrast and context-context contrast to enhance the performance of contrastive learning-based SRSs. Transformer is used as the sequence encoder of ContraRec in our implement.

Diversified SRS methods:

- **MCPRN (IJCAI'19)** [47] proposes a mixture-channel model to learn users' multiple purposes for SRSs.
- **ComiRec (KDD'20)** [4] employs an attention mechanism and dynamic routing technique to model users' multiple interests.
- **TEDDY (IPM'24)** [25] proposes an adaptive masking module to disentangle diversity from users' main interest.

Diffusion model-based SRS methods:

- **DiffuRec (TOIS'23)** [24] employs a diffusion model to represent item embeddings in a distribution space and then utilizes a approximator to generate target item representations.
- **DreamRec (NeurIPS'23)** [60] utilizes the Transformer to learn users' preferences, which then serve as the diffusion guidance for generating the oracle item for each user.
- **DiffRIS (WWW'24)** [33] combines CNN and LSTM to learn a diffusion guidance to direct the item generation process.

5.1.3 Parameter Settings. For fair comparisons, we carefully tune all models, including both competitive methods and our proposed DiffDiv. Specifically, for each competitive method, we first initialize hyper-parameter values according to the instructions provided in the original papers and then fine-tune them on our used datasets to ensure their best performances. To be fair, the dimension of item embeddings is fixed to 64 across all models. The dimensions of MLP utilized to predict $\mathbf{x}_0$ is: 192→64. We conduct grid searching strategy to select the optimal model hyperparameters. Specifically, the dimension of hidden layer in DAGL is tuned within {8, 16, 32, 48}. The learning rate is tuned within {0.0001, 0.001, 0.005, 0.01, 0.02, 0.05}. The maximum diffusion step T is searched within {500, 1000, 1500, 2000}, and the maximum reverse step T' is selected within $\{3, 4, \cdots, 10\}$. We tune L from {2, 4, 8, 16, 32}. The values of α, β, and γ are selected from {0.01, 0.05, 0.1, 0.5}. The batch size is tuned in {128, 256, 512, 1024}. To ensure fairness, we use the same item embedding initialization method, $\mathcal{N}(\mathbf{0}, \mathbf{I})$, across all models, as the embedding initialization can significantly affect the diversity metrics. The maximum number of epochs is set to 1000. Adamw [28] is used to optimize the model parameters. Each method was run five times, and the average results are reported.
Experimental Environment: All methods are implemented using an NVIDIA L4 GPU with 24GB memory.

5.1.4 Evaluation Metrics. We select metrics from two perspectives: (1) **Accuracy:** Normalized Discount Cumulative Gain (NDCG) and Mean Reciprocal Rank (MRR); (2) **Diversity:** Since the datasets lack categorical information, we select Intra-List Distance (ILD) and History-Recommendation Distance (HRD) to evaluate the recommendation diversity. ILD measures the average distance between each two items in the recommendation lists and has been commonly used to measure recommendation diversity [61, 64]. HRD is a new metric proposed in this paper to measure the average distance between recommended items and users' historically interacted items. HRD can assess the recommendation diversity and the ability to mitigate filter bubbles. The higher the ILD and HRD, the greater the diversity. Their calculation formulas are as follows:

$$\text{ILD@}K = \frac{1}{|\mathcal{R}_K|(|\mathcal{R}_K|-1)} \sum_{i \in \mathcal{R}_K} \sum_{j \in \mathcal{R}_K \setminus i} d_{ij}, \tag{17}$$

$$\text{HRD@}K = \frac{1}{|\mathcal{R}_K|} \sum_{i \in \mathcal{R}_K} d_{im}, \tag{18}$$

where $\mathcal{R}_K$ and $\mathcal{R}_K \setminus i$ refer to the whole recommendation list and the list without item i. d_{ij} is Euclidean distance between embeddings of item i and j. d_{im} is Euclidean distance between embedding of item i and the mean embedding of a user's historical items.

Inspired from F1-score [8, 48], we aggregate accuracy and diversity metrics into two **unified metrics**: NI and MH. These metrics assess a method's ability to balance accuracy and diversity. Their calculation formulas are as follows:

$$\text{NI@}K = \frac{2 \times \text{NDCG@}K \times (\text{ILD@}K)/\tau}{\text{NDCG@}K + (\text{ILD@}K)/\tau}, \tag{19}$$

$$\text{MH@}K = \frac{2 \times \text{MRR@}K \times (\text{HRD@}K)/\tau}{\text{MRR@}K + (\text{HRD@}K)/\tau}, \tag{20}$$

[1] https://recsys.acm.org/recsys15/challenge/

Table 2: Performance comparison of DiffDiv and competitive methods on Zhihu and YooChoose datasets. The best performances are in bold, and the second-best performances are underlined. *the improvement is significant at p<0.05.

Datasets	Models	NDCG↑		MRR↑		ILD↑		HRD↑		NI↑		MH↑	
		10	20	10	20	10	20	10	20	10	20	10	20
Zhihu	GRU4Rec	0.00384	0.00544	0.00266	0.00309	5.58115	5.61662	4.28040	4.28076	0.00323	0.00370	0.00237	0.00253
	LRURec	0.00333	0.00543	0.00213	0.00268	4.10125	4.08863	5.52386	5.59038	0.00254	0.00297	0.00240	0.00274
	SASRec	0.00369	0.00620	0.00235	0.00303	5.81507	5.89957	4.43189	4.45563	0.00325	0.00400	0.00228	0.00257
	BERT4Rec	0.00448	0.00588	0.00293	0.00331	4.33884	4.36479	3.35486	3.34901	0.00292	0.00318	0.00213	0.00222
	CL4SRec	0.00462	0.00709	0.00278	0.00345	5.37627	5.43871	4.16747	4.17547	0.00340	0.00393	0.00238	0.00260
	ContraRec	0.00495	0.00732	0.00317	0.00380	5.67480	5.64416	4.29346	4.23113	0.00361	0.00407	0.00256	0.00272
	MCPRN	0.00443	0.00647	0.00280	0.00335	5.66089	5.70987	7.54057	7.49499	0.00345	0.00396	0.00321	0.00354
	ComiRec	0.00416	0.00716	0.00236	0.00316	7.33751	7.17321	5.65078	5.56705	0.00390	0.00478	0.00257	0.00296
	TEDDY	0.00479	0.00689	0.00310	0.00365	6.26600	6.48836	5.13651	5.20878	0.00379	0.00441	0.00281	0.00304
	DiffuRec	0.00475	0.00718	0.00280	0.00345	4.86578	4.83749	4.76312	4.73844	0.00322	0.00362	0.00258	0.00281
	DreamRec	0.00570	0.00742	0.00336	0.00381	10.85885	11.01705	7.81129	7.88157	0.00556	0.00632	0.00361	0.00388
	DiffRIS	0.00391	0.00675	0.00235	0.00312	10.73667	10.82377	7.99747	7.82952	0.00452	0.00601	0.00296	0.00347
	DiffDiv	**0.00588***	**0.00858***	**0.00403***	**0.00476***	**11.94270***	**11.77684***	**8.52111***	**8.37391***	**0.00592***	**0.00698***	**0.00414***	**0.00445***
YooChoose	GRU4Rec	0.01247	0.01537	0.00896	0.00975	3.23436	3.24180	2.55922	2.53235	0.00852	0.00912	0.00652	0.00667
	LRURec	0.01281	0.01604	0.00774	0.00863	2.84228	2.74068	5.63873	5.71688	0.00788	0.00817	0.00918	0.00983
	SASRec	0.01234	0.01546	0.00753	0.00838	2.59397	2.59537	2.10353	2.07503	0.00730	0.00777	0.00540	0.00555
	BERT4Rec	0.01155	0.01417	0.00783	0.00854	2.43285	2.46163	2.03550	2.02278	0.00685	0.00731	0.00536	0.00549
	CL4SRec	0.01200	0.01501	0.00743	0.00826	3.70284	3.71801	3.20326	3.13796	0.00916	0.00994	0.00688	0.00713
	ContraRec	0.01373	0.01684	0.00868	0.00953	4.02736	3.89121	3.24740	3.07876	0.01015	0.01064	0.00743	0.00748
	MCPRN	0.01421	0.01720	0.00998	0.01080	2.64548	2.66976	4.75372	4.72608	0.00771	0.00815	0.00974	0.01008
	ComiRec	0.01507	0.01764	0.01047	0.01118	4.22483	4.14645	3.81138	3.62569	0.01083	0.01128	0.00882	0.00880
	TEDDY	0.01550	0.01794	0.01099	0.01166	5.18791	5.49729	5.25560	5.23578	0.01243	0.01363	0.01075	0.01103
	DiffuRec	0.01477	0.01669	0.01042	0.01095	3.03897	3.09295	2.72284	2.70464	0.00861	0.00903	0.00715	0.00724
	DreamRec	0.01955	0.02221	0.01260	0.01333	11.24598	11.10570	8.38581	8.21427	0.02092	0.02221	0.01439	0.01472
	DiffRIS	0.01734	0.02009	0.01056	0.01130	11.05444	11.04874	8.16722	8.14155	0.01943	0.02105	0.01283	0.01334
	DiffDiv	**0.02205***	**0.02697***	**0.01588***	**0.01722***	**11.88060***	**11.76550***	**8.69080***	**8.57745***	**0.02288***	**0.02513***	**0.01660***	**0.01719***

Table 3: Performance comparison of DiffDiv and its variants without one of the key components.

Datasets	Models	NDCG		MRR		ILD		HRD		NI		MH	
		10	20	10	20	10	20	10	20	10	20	10	20
Zhihu	w/o DAGL	0.00531	0.00734	0.00357	0.00413	11.62495	11.54263	8.34158	8.24507	0.00555	0.00646	0.00385	0.00413
	w/o ADBO	0.00493	0.00738	0.00318	0.00384	**12.06756**	**11.87712**	**8.67164**	**8.48404**	0.00543	0.00658	0.00367	0.00403
	w/o HDI	0.00519	0.00725	0.00353	0.00408	11.33497	11.17212	8.22859	7.99547	0.00542	0.00631	0.00380	0.00404
	DiffDiv	**0.00588**	**0.00858**	**0.00403**	**0.00476**	11.94270	11.77684	8.52111	8.37391	**0.00592**	**0.00698**	**0.00414**	**0.00445**
YooChoose	w/o DAGL	0.02182	0.02472	0.01542	0.01619	11.32020	11.29961	8.43169	8.35417	0.02222	0.02361	0.01611	0.01645
	w/o ADBO	0.01541	0.02057	0.01090	0.01229	**11.93148**	**11.82510**	8.65367	8.54469	0.01873	0.02200	0.01338	0.01430
	w/o HDI	0.02079	0.02514	0.01571	0.01687	11.26834	11.28909	8.41865	8.35876	0.02163	0.02379	0.01625	0.01679
	DiffDiv	**0.02205**	**0.02697**	**0.01588**	**0.01722**	11.88060	11.76550	**8.69080**	**8.57745**	**0.02288**	**0.02513**	**0.01660**	**0.01719**

where τ rescales diversity metrics to align with accuracy metrics. Without this rescaling, the unified metrics would be disproportionately influenced by the significantly smaller values of accuracy metrics. Specifically, τ is set to 2000 and 500 for Zhihu and YooChoose datasets, respectively.

5.2 Overall Performance (RQ1)

From the results in Table 2, we have the following observations:
(1) DiffDiv significantly outperforms competitive methods in terms of both accuracy (NDCG, MRR), diversity (ILD, HRD), and unified metrics (NI, MH). Specifically, DiffDiv demonstrates average improvements of 18.6%, 6.4%, and 15.9% over the two datasets in terms of NDCG@20, ILD@20, and MH@20, respectively, compared to the second best method. This can be attributed to three key components of DiffDiv: DAGL, ADBO and HDI. It should be noted that the improvement in recommendation diversity achieved by DiffDiv does not lead to the decline of recommendation accuracy. This observation aligns with the findings in [54, 61], which suggest that the relationship between accuracy and diversity can be a "win-win". The simultaneous improvement in accuracy and diversity suggests that users, in practice, seek to interact with a diverse range of items rather than homogeneous ones. This also validates the significance of our research motivation to enhance the diversity of SRSs.
(2) Contrastive learning-based methods (CL4SRec and ContraRec) outperform general methods in both accuracy and diversity. This is because constructing multiple sequence views can alleviate the data sparsity problem and capture more of users' potential intents.
(3) Diversified methods (MCPRN, ComiRec, TEDDY) generally perform better in recommendation diversity metrics than non-diversified SRS methods. These methods typically apply multi-interest modeling modules or disentangle diversity factors from the main interest factors to capture users' diverse preferences.
(4) DM-based methods (DreamRec, DiffRIS, DiffDiv) consistently outperform general methods in recommendation accuracy and diversity. DMs generate the next item from scratch, leading to more diverse outcomes. An exception is DiffuRec, which does not exhibit better diversity modeling capacity than general methods, primarily

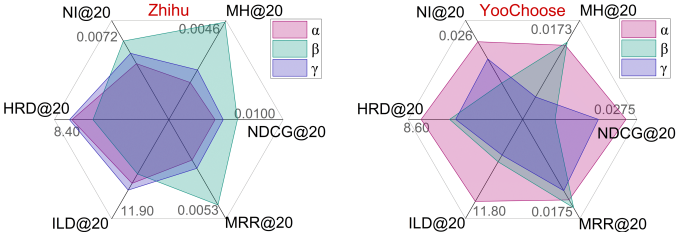

Figure 3: Recommendation performances under different type of robust divergences (α–, β– and γ–divergence).

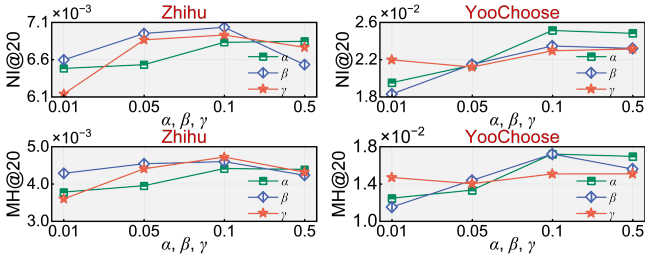

Figure 4: Recommendation performances of DiffDiv under different values of α, β and γ.

because it is based on a classification loss function that pushes generated item embeddings to resemble users' historical interactions.

5.3 Ablation Study (RQ2)

5.3.1 Evaluation of different model components. To validate the effectiveness of the key components of DiffDiv, we compare DiffDiv with three variants: w/o DAGL (replacing diversity-aware guidance $\hat{\mathbf{c}}$ with $\mathbf{c}$), w/o ADBO (replacing accuracy-diversity balanced optimization with commonly used diffusion optimization), and w/o HDI (replacing heterogeneous diffusion inference with single-channel inference). From Table 3, we have the following observations:
(1) w/o DAGL and w/o HDI deteriorate both accuracy and diversity of DiffDiv. This demonstrates that our designed diversity-aware guidance and heterogeneous diffusion inference modules actually benefit the model in terms of both accuracy and diversity.
(2) w/o ADBO raises model diversity but significantly degrades model accuracy. Without ADBO, the noise introduced by DAGL cannot be suppressed, leading to increased diversity at the expense of a huge decline in accuracy, which is unacceptable in practice.

5.3.2 Evaluation of different types of robust-divergence. To evaluate DiffDiv's performance with different types of robust divergence (α-, β-, and γ-divergence) employed in the accuracy-diversity balanced optimization, we conduct experiments using each divergence type individually. Figure 3 shows that β-divergence performs best on Zhihu dataset, while α-divergence excels on YooChoose dataset. This suggests that different robust divergence methods are suited to different scenarios and should be selected according to the specific characteristics of the datasets.

5.4 Analysis of Parameter Sensitivity (RQ3)

5.4.1 Performance w.r.t. values of α, β and γ. The values of α, β, and γ can be utilized to balance the recommendation accuracy and diversity. Smaller values tend to bias the model towards diversity, while larger values favor accuracy. Figure 4 shows that the general trend for NI@20 and MH@20 initially increases and then declines, indicating that by appropriately tuning α, β, and γ, DiffDiv can effectively balance the recommendation accuracy and diversity.

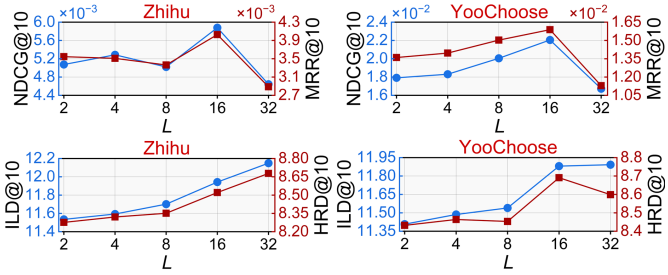

Figure 5: Recommendation performances of DiffDiv under different numbers of inference channel L.

5.4.2 Performance w.r.t. number of inference channel L. The number of inference channels, L, is a crucial hyperparameter that significantly impacts model performances. Figure 5 shows that increasing L consistently improves recommendation diversity, although there is some fluctuation. This suggests that our proposed heterogeneous diffusion inference does actually improve the model diversity. The recommendation accuracy initially improve with an increase in L, reaching its peak at $L = 16$, before declining. This may be explained that an excessively large L can blur user intents, potentially diminishing recommendation performance to some extent.

5.4.3 Performance and time consumption w.r.t. number of reverse steps T'. We select the reverse step T' from 3 to 10 for efficiency consideration. DiffDiv leverages a diversity-aware guidance to direct the training process, enabling the model to quickly capture a wide range of users' preferences. As a result, fewer reverse steps are required compared to forward steps. To validate this, we compare recommendation performances and the inference time of DIffDiv under different T' values from range {5, 10, 50, 100, 200, 500}.

As shown in Figure 6, the recommendation performance with $T' = 5$ is comparable to that with higher T' values. However, as T' increases, the inference time grows significantly. To be specific, while setting $T' = 50$ yields slightly better recommendation performance than $T' = 5$, the corresponding increase in inference time is impractical. Therefore, DiffDiv generally selects a smaller T' value, such as 3 or 5, to enhance its generalizability and scalability without incurring an unacceptable performance drop.

5.5 Efficiency Analysis (RQ4)

We measured the average training and inference time per epoch for DiffDiv and competitive methods. Since the sequence encoder influences time efficiency significantly, we compare DiffDiv with Transformer-based methods and design a GRU version of DiffDiv for comparison with RNN-based methods. To ensure fairness, we set $L = 16$ which leads to the best performances and fix the training batch size at 256 and the test batch size at 100 across all methods. Notably, the inference stage, which includes the calculation of all evaluation metrics, is generally longer than the training stage.

As shown in Table 4, the training time of DiffDiv-G and DiffDiv is comparable to competitive methods. Additionally, DiffDiv's inference time is shorter than other DM-based models, as the maximum

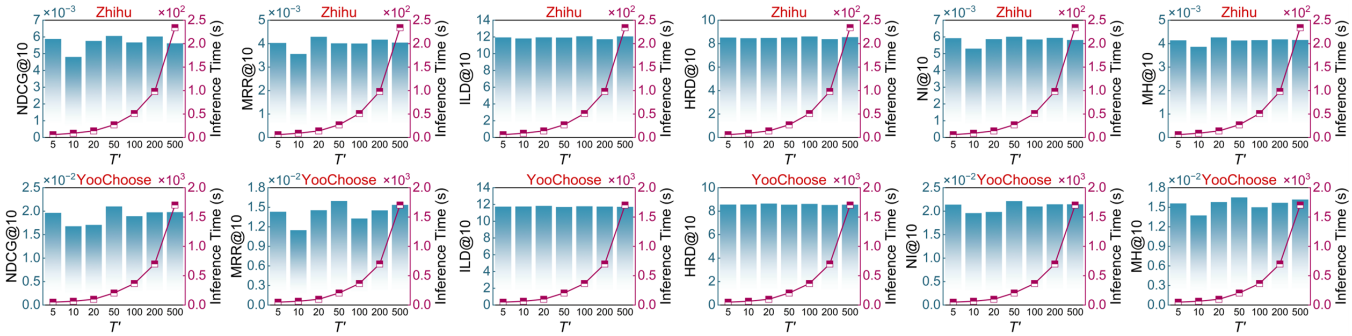

Figure 6: Recommendation performances of DiffDiv under different number of reverse step T' from {5, 10, 20, 50, 100, 200, 500}.

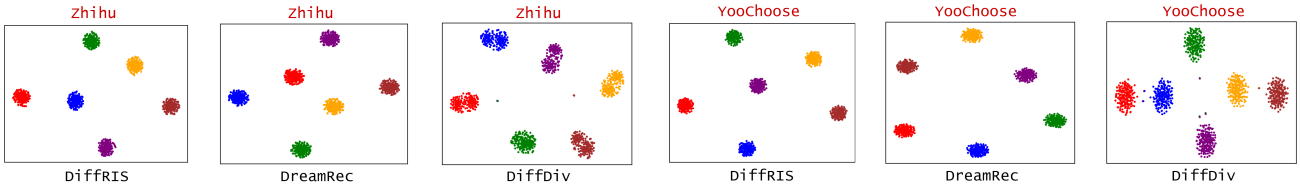

Figure 7: Visualization of the embeddings of the top 200 items recommended for six randomly selected users (sequences) using T-SNE. Each color represents a different user (sequence). The more dispersed the points, the greater the distance between items in the recommendation lists, indicating higher diversity in the recommendation results.

Table 4: Efficiency comparison of DiffDiv and competitive methods. DiffDiv-G is a GRU version of DiffDiv.

Zhihu				
Models (RNNs)	GRU4Rec	MCPRN	DiffRIS	DiffDiv-G
Training time (s)	0.834	13.772	1.229	1.046
Inference time (s)	4.346	5.118	15.324	6.712
Models (Transformer)	SASRec	DiffuRec	DreamRec	DiffDiv
Training time (s)	4.038	3.398	4.014	4.264
Inference time (s)	4.454	27.936	24.167	6.795
YooChoose				
Models (RNNs)	GRU4Rec	MCPRN	DiffRIS	DiffDiv-G
Training time (s)	12.428	176.001	15.242	13.956
Inference time (s)	31.069	38.724	108.465	41.483
Models (Transformer)	SASRec	DiffuRec	DreamRec	DiffDiv
Training time (s)	34.681	35.044	35.509	36.024
Inference time (s)	32.060	133.956	168.394	42.004

reverse step of DiffDiv is much smaller. While DiffDiv-G and DiffDiv have slightly longer inference time than GRU4Rec and SASRec, this is totally acceptable given the substantial improvements in both recommendation accuracy and diversity.

Furthermore, as shown in Table 1, the YooChoose dataset contains approximately 9.4 times more interactions than Zhihu (737,163 vs 77,712). From Table 4, we observe that the training time of DiffDiv on YooChoose is approximately 8.4 times that on Zhihu (36.024s vs 4.264s). This linear relationship between time consumption and dataset size highlights the scalability of DiffDiv for larger datasets.

5.6 Case Study

Since both two datasets used do not contain item information such as categories, text, or images, we were unable to conduct a case study showcasing the textual or visual features of the recommended items. Instead, to validate DiffDiv's effectiveness in enhancing diversity, we conduct a case study by randomly selecting six users and visualizing the embeddings of the top 200 items recommended to each user using the T-SNE technique [37]. We present the results of three methods: DiffDiv, and two competitive methods: DiffRIS [33] and DreamRec [60]. These two methods have demonstrated superior performances in diversity metrics among all baselines. The greater the dispersion of points within each group, the higher the diversity in recommendations.

As shown in Figure 7, the embeddings of the recommended items provided by DiffRIS are slightly more crowded than those of DreamRec. This observation aligns with the results in Table 2, where DreamRec outperforms DiffRIS in diversity metrics. Compared with both DiffRIS and DreamRec, the embeddings of the recommended items provided by DiffDiv are more decentralized, confirming that DiffDiv can provide more diversified recommendation results.

6 Conclusion and Future Work

In this paper, we proposed a diversity-guided diffusion model, DiffDiv, to unleash the potential of DMs towards diversified sequential recommendations. First, we designed a new diversity-aware guidance learning module (DAGL) to construct a diversity-aware guidance to guide the training of DiffDiv towards capturing more diversified preferences of users. Additionally, we devised a new accuracy-diversity balanced optimization strategy (ADBO) to trade off between recommendation accuracy and diversity. Afterwards, we designed a novel heterogeneous diffusion inference mechanism (HDI) to generate multiple types of items to accommodate users' heterogeneous preferences. Extensive experiments conducted on two real-world datasets demonstrate the effectiveness of DiffDiv in terms of both recommendation accuracy and diversity. In future work, our proposed diversity-guided diffusion model can be extended to other recommendation scenarios (e.g., multimodal recommendations) to enhance their recommendation diversity.

References

[1] Vaswani Ashish, Shazeer Noam, Parmar Niki, Uszkoreit Jakob, Jones Llion, N. Gomez Aidan, Kaiser Lukasz, and Polosukhin Illia. 2017. Attention is all you need. In *Conference on Neural Information Processing Systems*. 5998–6008.

[2] Chems Eddine Berbague, Nour El-islem Karabadji, Hassina Seridi, Panagiotis Symeonidis, Yannis Manolopoulos, and Wajdi Dhifli. 2021. An overlapping clustering approach for precision, diversity and novelty-aware recommendations. *Expert Systems with Applications* 177 (2021), 114917.

[3] Jaime Carbonell and Jade Goldstein. 1998. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *ACM SIGIR Conference on Research and Development in Information Retrieval*. 335–336.

[4] Yukuo Cen, Jianwei Zhang, Xu Zou, Chang Zhou, Hongxia Yang, and Jie Tang. 2020. Controllable multi-interest framework for recommendation. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2942–2951.

[5] Jianxin Chang, Chen Gao, Yu Zheng, Yiqun Hui, Yanan Niu, Yang Song, Depeng Jin, and Yong Li. 2021. Sequential recommendation with graph neural networks. In *ACM SIGIR Conference on Research and Development in Information Retrieval*. 378–387.

[6] Wanyu Chen, Pengjie Ren, Fei Cai, Fei Sun, and Maarten de Rijke. 2020. Improving end-to-end sequential recommendations with intent-aware diversification. In *ACM International Conference on Information and Knowledge Management*. 175–184.

[7] Wanyu Chen, Pengjie Ren, Fei Cai, Fei Sun, and Maarten De Rijke. 2021. Multi-interest diversification for end-to-end sequential recommendation. *ACM Transactions on Information Systems* 40, 1 (2021), 1–30.

[8] Davide Chicco and Giuseppe Jurman. 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics* 21 (2020), 1–13.

[9] Sanghyuk Chun, Seong Joon Oh, Rafael Sampaio De Rezende, Yannis Kalantidis, and Diane Larlus. 2021. Probabilistic embeddings for cross-modal retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition*. 8415–8424.

[10] Prafulla Dhariwal and Alexander Nichol. 2021. Diffusion models beat gans on image synthesis. In *Conference on Neural Information Processing Systems*, Vol. 34. 8780–8794.

[11] Futoshi Futami, Issei Sato, and Masashi Sugiyama. 2018. Variational inference based on robust divergences. In *International Conference on Artificial Intelligence and Statistics*. 813–822.

[12] Bin Hao, Min Zhang, Weizhi Ma, Shaoyun Shi, Xinxing Yu, Houzhi Shan, Yiqun Liu, and Shaoping Ma. 2021. A large-scale rich context query and recommendation dataset in online knowledge-sharing. *arXiv preprint arXiv:2106.06467* (2021).

[13] Ruining He and Julian McAuley. 2016. Fusing similarity models with markov chains for sparse sequential recommendation. In *IEEE International Conference on Data Mining*. 191–200.

[14] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2015. Session-based recommendations with recurrent neural networks. *arXiv preprint arXiv:1511.06939* (2015).

[15] Balázs Hidasi and Domonkos Tikk. 2016. General factorization framework for context-aware recommendations. *Data Mining and Knowledge Discovery* 30, 2 (2016), 342–371.

[16] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. In *Conference on Neural Information Processing Systems*, Vol. 33. 6840–6851.

[17] Guoqing Hu, Zhengyi Yang, Zhibo Cai, An Zhang, and Xiang Wang. 2024. Generate and instantiate what you prefer: Text-guided diffusion for sequential recommendation. *arXiv preprint arXiv:2410.13428* (2024).

[18] Wang-Cheng Kang and Julian McAuley. 2018. Self-attentive sequential recommendation. In *IEEE International Conference on Data Mining*. 197–206.

[19] Tung Kieu, Bin Yang, Chenjuan Guo, Razvan-Gabriel Cirstea, Yan Zhao, Yale Song, and Christian S Jensen. 2022. Anomaly detection in time series with robust variational quasi-recurrent autoencoders. In *IEEE International Conference on Data Engineering*. 1342–1354.

[20] Hanbyul Lee and Junghyun Kim. 2024. EDiffuRec: An enhanced diffusion model for sequential recommendation. *Mathematics* 12, 12 (2024), 1795.

[21] Chang Li, Haoyun Feng, and Maarten de Rijke. 2020. Cascading hybrid bandits: Online learning to rank for relevance and diversity. In *ACM Conference on Recommender Systems*. 33–42.

[22] Jin Li, Shoujin Wang, Qi Zhang, Shui Yu, and Fang Chen. 2025. Generating with Fairness: A Modality-Diffused Counterfactual Framework for Incomplete Multimodal Recommendations. *arXiv preprint arXiv:2501.11916* (2025).

[23] Wuchao Li, Rui Huang, Haijun Zhao, Chi Liu, Kai Zheng, Qi Liu, Na Mou, Guorui Zhou, Defu Lian, Yang Song, et al. 2024. DimeRec: a unified framework for enhanced sequential recommendation via generative diffusion models. *arXiv preprint arXiv:2408.12153* (2024).

[24] Zihao Li, Aixin Sun, and Chenliang Li. 2023. Diffurec: A diffusion model for sequential recommendation. *ACM Transactions on Information Systems* 42, 3 (2023), 1–28.

[25] Zihao Li, Yunfan Xie, Wei Emma Zhang, Pengfei Wang, Lixin Zou, Fei Li, Xiangyang Luo, and Chenliang Li. 2024. Disentangle interest trend and diversity for sequential recommendation. *Information Processing & Management* 61, 3 (2024), 103619.

[26] Qidong Liu, Fan Yan, Xiangyu Zhao, Zhaocheng Du, Huifeng Guo, Ruiming Tang, and Feng Tian. 2023. Diffusion augmentation for sequential recommendation. In *ACM International Conference on Information and Knowledge Management*. 1576–1586.

[27] Shihai Liu and Yonghui Yang. 2024. Multi-interest distribution based diversified recommendation. In *Asia Conference on Algorithms, Computing and Machine Learning*. 176–183.

[28] Ilya Loshchilov and Frank Hutter. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101* (2017).

[29] Haokai Ma, Ruobing Xie, Lei Meng, Xin Chen, Xu Zhang, Leyu Lin, and Zhanhui Kang. 2024. Plug-in diffusion model for sequential recommendation. In *AAAI Conference on Artificial Intelligence*, Vol. 38. 8886–8894.

[30] Haokai Ma, Ruobing Xie, Lei Meng, Yimeng Yang, Xingwu Sun, and Zhanhui Kang. 2024. Seedrec: sememe-based diffusion for sequential recommendation. In *International Joint Conference on Artificial Intelligence*. 1–9.

[31] Haokai Ma, Yimeng Yang, Lei Meng, Ruobing Xie, and Xiangxu Meng. 2024. Multimodal conditioned diffusion model for recommendation. In *Companion Proceedings of the ACM on Web Conference*. 1733–1740.

[32] Tien T Nguyen, Pik-Mai Hui, F Maxwell Harper, Loren Terveen, and Joseph A Konstan. 2014. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *International World Wide Web Conference*. 677–686.

[33] Yong Niu, Xing Xing, Zhichun Jia, Ruidi Liu, Mindong Xin, and Jianfu Cui. 2024. Diffusion recommendation with implicit sequence influence. In *Companion Proceedings of the ACM on Web Conference*. 1719–1725.

[34] Seong Joon Oh, Kevin P Murphy, Jiyan Pan, Joseph Roth, Florian Schroff, and Andrew C Gallagher. 2019. Modeling uncertainty with hedged instance embeddings. In *International Conference on Learning Representations*.

[35] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer Assisted Intervention*. 234–241.

[36] Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. 2019. BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer. In *ACM International Conference on Information and Knowledge Management*. 1441–1450.

[37] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9 (2008), 2579–2605.

[38] Chenyang Wang, Weizhi Ma, Chong Chen, Min Zhang, Yiqun Liu, and Shaoping Ma. 2023. Sequential recommendation with multiple contrast signals. *ACM Transactions on Information Systems* 41, 1 (2023), 1–27.

[39] Chenyang Wang, Zhefan Wang, Yankai Liu, Yang Ge, Weizhi Ma, Min Zhang, Yiqun Liu, Junlan Feng, Chao Deng, and Shaoping Ma. 2022. Target interest distillation for multi-interest recommendation. In *ACM International Conference on Information and Knowledge Management*. 2007–2016.

[40] Nan Wang, Shoujin Wang, Yan Wang, Quan Z Sheng, and Mehmet Orgun. 2020. Modelling local and global dependencies for next-item recommendations. In *Web Information Systems Engineering–WISE 2020: 21st International Conference, Amsterdam, The Netherlands, October 20–24, 2020, Proceedings, Part II 21*. 285–300.

[41] Nan Wang, Shoujin Wang, Yan Wang, Quan Z Sheng, and Mehmet A Orgun. 2022. Exploiting intra-and inter-session dependencies for session-based recommendations. *World Wide Web* 25, 1 (2022), 425–443.

[42] Rongyao Wang, Shoujin Wang, Wenpeng Lu, and Xueping Peng. 2022. News recommendation via multi-interest news sequence modelling. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 7942–7946.

[43] Shoujin Wang, Longbing Cao, Liang Hu, Shlomo Berkovsky, Xiaoshui Huang, Lin Xiao, and Wenpeng Lu. 2020. Hierarchical attentive transaction embedding with intra-and inter-transaction dependencies for next-item recommendation. *IEEE Intelligent Systems* 36, 4 (2020), 56–64.

[44] Shoujin Wang, Longbing Cao, Yan Wang, Quan Z Sheng, Mehmet A Orgun, and Defu Lian. 2021. A survey on session-based recommender systems. *Comput. Surveys* 54, 7 (2021), 1–38.

[45] Shoujin Wang, Liang Hu, and Longbing Cao. 2017. Perceiving the next choice with comprehensive transaction embeddings for online recommendation. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2017, Skopje, Macedonia, September 18–22, 2017, Proceedings, Part II 17*. 285–302.

[46] Shoujin Wang, Liang Hu, Yan Wang, Longbing Cao, Quan Z Sheng, and Mehmet Orgun. 2019. Sequential recommender systems: challenges, progress and prospects. In *International Joint Conference on Artificial Intelligence*. 6332–6338.

[47] Shoujin Wang, Liang Hu, Yan Wang, Quan Z Sheng, Mehmet Orgun, and Longbing Cao. 2019. Modeling multi-purpose sessions for next-item recommendations via mixture-channel purpose routing networks. In *International Joint Conference on Artificial Intelligence*.

[48] Shoujin Wang, Wentao Wang, Xiuzhen Zhang, Yan Wang, Huan Liu, and Fang Chen. 2024. A hierarchical and disentangling interest learning framework for unbiased and true news recommendation. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 3200–3211.

[49] Shoujin Wang, Xiuzhen Zhang, Yan Wang, and Francesco Ricci. 2024. Trustworthy recommender systems. *ACM Transactions on Intelligent Systems and Technology* 15, 4 (2024), 1–20.

[50] Weidong Wang, Yan Tang, and Kun Tian. 2024. LeadRec: Towards personalized sequential recommendation via guided diffusion. In *International Conference on Intelligent Computing*. 3–15.

[51] Wenjie Wang, Yiyan Xu, Fuli Feng, Xinyu Lin, Xiangnan He, and Tat-Seng Chua. 2023. Diffusion recommender model. In *ACM SIGIR Conference on Research and Development in Information Retrieval*. 832–841.

[52] Yu Wang, Zhiwei Liu, Liangwei Yang, and Philip S Yu. 2024. Conditional denoising diffusion for sequential recommendation. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. 156–169.

[53] Chao-Yuan Wu, Amr Ahmed, Alex Beutel, Alexander J Smola, and How Jing. 2017. Recurrent recommender networks. In *ACM International Conference on Web Search and Data Mining*. 495–503.

[54] Qiong Wu, Yong Liu, Chunyan Miao, Binqiang Zhao, Yin Zhao, and Lu Guan. 2019. PD-GAN: Adversarial learning for personalized diversity-promoting recommendation. In *International Joint Conference on Artificial Intelligence*. 3870–3876.

[55] Zhangkai Wu, Longbing Cao, and Lei Qi. 2024. evae: Evolutionary variational autoencoder. *IEEE Transactions on Neural Networks and Learning Systems* (2024).

[56] Zihao Wu, Xin Wang, Hong Chen, Kaidong Li, Yi Han, Lifeng Sun, and Wenwu Zhu. 2023. Diff4rec: Sequential recommendation with curriculum-scheduled diffusion augmentation. In *ACM International Conference on Multimedia*. 9329–9335.

[57] Wenjia Xie, Rui Zhou, Hao Wang, Tingjia Shen, and Enhong Chen. 2024. Bridging User Dynamics: Transforming Sequential Recommendations with Schr\" odinger Bridge and Diffusion Models. *arXiv preprint arXiv:2409.10522* (2024).

[58] Xu Xie, Fei Sun, Zhaoyang Liu, Shiwen Wu, Jinyang Gao, Jiandong Zhang, Bolin Ding, and Bin Cui. 2022. Contrastive learning for sequential recommendation. In *IEEE International Conference on Data Mining*. 1259–1273.

[59] Zhengyi Yang, Xiangnan He, Jizhi Zhang, Jiancan Wu, Xin Xin, Jiawei Chen, and Xiang Wang. 2023. A generic learning framework for sequential recommendation with distribution shifts. In *ACM SIGIR Conference on Research and Development in Information Retrieval*. 331–340.

[60] Zhengyi Yang, Jiancan Wu, Zhicai Wang, Xiang Wang, Yancheng Yuan, and Xiangnan He. 2023. Generate what you prefer: Reshaping sequential recommendation via guided diffusion. In *Conference on Neural Information Processing Systems*, Vol. 36.

[61] Qing Yin, Hui Fang, Zhu Sun, and Yew-Soon Ong. 2023. Understanding diversity in session-based recommendation. *ACM Transactions on Information Systems* 42, 1 (2023), 1–34.

[62] Fajie Yuan, Alexandros Karatzoglou, Ioannis Arapakis, Joemon M Jose, and Xiangnan He. 2019. A simple convolutional generative network for next item recommendation. In *ACM International Conference on Web Search and Data Mining*. 582–590.

[63] Zhenrui Yue, Yueqi Wang, Zhankui He, Huimin Zeng, Julian McAuley, and Dong Wang. 2024. Linear recurrent units for sequential recommendation. In *ACM International Conference on Web Search and Data Mining*. 930–938.

[64] Mi Zhang and Neil Hurley. 2008. Avoiding monotony: improving the diversity of recommendation lists. In *ACM Conference on Recommender Systems*. 123–130.

[65] Mengqi Zhang, Shu Wu, Xueli Yu, Qiang Liu, and Liang Wang. 2022. Dynamic graph neural networks for sequential recommendation. *IEEE Transactions on Knowledge and Data Engineering* 35, 5 (2022), 4741–4753.

[66] Jujia Zhao, Wang Wenjie, Yiyan Xu, Teng Sun, Fuli Feng, and Tat-Seng Chua. 2024. Denoising diffusion recommender model. In *ACM SIGIR Conference on Research and Development in Information Retrieval*. 1370–1379.

[67] Yuying Zhao, Yu Wang, Yunchao Liu, Xueqi Cheng, Charu C Aggarwal, and Tyler Derr. 2023. Fairness and diversity in recommender systems: a survey. *ACM Transactions on Intelligent Systems and Technology* (2023).

[68] Peilin Zhou, You-Liang Huang, Yueqi Xie, Jingqi Gao, Shoujin Wang, Jae Boum Kim, and Sunghun Kim. 2024. Is contrastive learning necessary? a study of data augmentation vs contrastive learning in sequential recommendation. In *Proceedings of the ACM Web Conference 2024*. 3854–3863.

[69] Cai-Nicolas Ziegler, Sean M McNee, Joseph A Konstan, and Georg Lausen. 2005. Improving recommendation lists through topic diversification. In *International World Wide Web Conference*. 22–32.