Diff-MSR: A Diffusion Model Enhanced Paradigm for Cold-Start Multi-Scenario Recommendation

Yuhao Wang†
City University of Hong Kong
yhwang25-c@my.cityu.edu.hk

Ziru Liu†
City University of Hong Kong
ziruliu2-c@my.cityu.edu.hk

Yichao Wang
Huawei Noah's Ark Lab
wangyichao5@huawei.com

Xiangyu Zhao ✉
City University of Hong Kong
xianzhao@cityu.edu.hk

Bo Chen
Huawei Noah's Ark Lab
chenbo116@huawei.com

Huifeng Guo ✉
Huawei Noah's Ark Lab
huifeng.guo@huawei.com

Ruiming Tang
Huawei Noah's Ark Lab
tangruiming@huawei.com

ABSTRACT

With the explosive growth of various commercial scenarios, there is an increasing number of studies on multi-scenario recommendation (MSR) which trains the recommender system with the data from multiple scenarios, aiming to improve the recommendation performance on all these scenarios synchronously. However, due to the large discrepancy in the number of interactions among domains, multi-scenario recommendation models usually suffer from insufficient learning and negative transfer especially on the cold-start scenarios, thus exacerbating the data sparsity issue. To fill this gap, in this work we propose a novel diffusion model enhanced paradigm tailored for the cold-start problem in multi-scenario recommendation in a data-driven generative manner. Specifically, based on all-domain data, we leverage the diffusion model with our newly designed variance schedule and the proposed classifier, which explicitly boosts the recommendation performance on the cold-start scenarios by exploiting the generated high-quality and informative embedding, leveraging the abundance of rich scenarios. Our experiments on Douban and Amazon datasets demonstrate two strengths of the proposed paradigm: (i) its effectiveness with a significant increase of 8.5% and 1% in accuracy on the two datasets, and (ii) its compatibility with various multi-scenario backbone models. The implementation code is available for easy reproduction[1,2].

CCS CONCEPTS

• **Information systems → Recommender systems**.

[1]https://github.com/mindspore-lab/models/tree/master/research/huawei-noah/Diff-MSR
[2]https://github.com/Applied-Machine-Learning-Lab/Diff-MSR

†Co-first author.
✉Corresponding authors.

WSDM '24, March 4–8, 2024, Merida, Mexico

ACM ISBN 979-8-4007-0371-3/24/03…$15.00
https://doi.org/10.1145/3616855.3635807

KEYWORDS

Cold-Start, Multi-Domain, Multi-Scenario Recommendation, Diffusion Model, Click-Through Rate Prediction

ACM Reference Format:
Yuhao Wang†, Ziru Liu†, Yichao Wang, Xiangyu Zhao ✉, Bo Chen, Huifeng Guo ✉, and Ruiming Tang. 2024. Diff-MSR: A Diffusion Model Enhanced Paradigm for Cold-Start Multi-Scenario Recommendation. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining (WSDM '24), March 4–8, 2024, Merida, Mexico.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3616855.3635807

1 INTRODUCTION

With the development of online service, recommender system (RS) prevails as it recommends items to interest user based on the learned preference from historical interactions. Recently, a dramatic boom in different scenarios or the so-called domains of recommendation has arisen in commercial RSs, such as the categories of product and different web pages that the users can interact with on the e-commerce platform. Therefore, multi-scenario recommendation (MSR) [13] emerges aiming at exploiting the data from multiple scenarios to improve their recommendation accuracy and tackle the data sparsity problem [48]. Generally, these scenarios have different number of historical interactions because of popularity, so it is natural to denote the one with comparatively large or low data volume as the rich or cold-start domain, respectively. Notably, the cold-start scenarios contribute to user satisfaction and business development, thus of great significance.

Nevertheless, due to the great disparity of data volume between the rich and cold-start scenarios, MSR models usually suffer from the following two limitations: (i) The domain-specific parameters tend to be insufficiently learned [44]. To be specific, it is common that the heavy-weight deep neural networks (i.e., towers) with the same architecture are built for all domains since it is exhausting and inefficient to manually search for the optimal structure of network for each domain. Consequently, the tower of cold-start domain can not be adequately trained and limited knowledge could be learned from the relatively scarce data. (ii) The domain-shared parameters tend to be dominated by the rich domains. Thus, the tower of cold-start domain is susceptible to negative transfer [34], where performance degrades when uncorrelated information is transferred across domains. Unfortunately, these limitations are still overlooked by the existing MSR methods even though they would result in the poor performance on the cold-start domains.

To address these deficiencies for cold-start MSR, we resort to the emerging diffusion model [10] which is widely known for the high-quality generation capability especially in computer vision [3]. However, there are two challenges. First, for MSR, there are complex resemblance and discrepancy among domains. By contrast, existing cold-start recommendation models only consider new items or novel users [18], which is a simpler yet different setting. Second, the existing diffusion model is only capable of capturing in-domain distribution, thus it can not establish the connection between different scenarios. To this end, we propose a novel **Diff**usion model enhanced paradigm for cold-start **M**ulti-**S**cenario **R**ecommendation (Diff-MSR), which utilizes diffusion model with the newly designed variance schedule and the proposed classifier to explicitly improve the performance of MSR backbone model on the cold-start domains. Specifically, the former focuses on capturing the characteristic of data distribution on the cold-start domain and generating high-quality embedding, while the latter aims at connecting the rich and cold-start domains through the informative outline of the noised embedding like a 'figure' in computer vision.

The contributions of this paper can be summarized as follows:

- We propose a novel paradigm Diff-MSR for the enhancement of cold-start multi-scenario recommendation, which leverages diffusion model with the designed variance schedule and classifier to explicitly improve the recommendation performance on the cold-start domains. To the best of our knowledge, this is the first effort to tackle the cold-start problem in MSR.
- The commonality and heterogeneity among scenarios is explicitly established through the noised embedding and measured by the classifier. Furthermore, high-quality and informative embeddings are generated for the cold-start domains with the help of rich domains. Consequently, the problem of insufficient learning and negative transfer is addressed in a data-driven generative manner;
- Extensive experiments show that Diff-MSR paradigm is effective and compatible with different MSR backbone models, and significantly outperforms various state-of-the-art generative methods.

2 PRELIMINARY

This section first illustrates the problem formulation of multi-scenario CTR prediction. Next, two groups of typical multi-scenario backbone model architectures are demonstrated.

2.1 Multi-Scenario CTR Prediction

Under the setting of multi-scenario recommendation, we target the Click-Trough Rate (CTR) prediction which is a binary classification task. Specifically, the recommender system takes the interaction data $(d, \boldsymbol{x})$ as input to predict y, i.e., the ground truth label indicating click ($y = 1$) or not ($y = 0$). d denotes the domain indicator $d \in \{1, 2, \ldots, D\}$ discriminating samples from D domains in total. $\boldsymbol{x}$ represents the raw features including user features and item attributes. Without loss of generality, suppose there are M features and they are all categorical, then $\boldsymbol{x} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M]$ where $\boldsymbol{x}_m$ denotes the one-hot representation of the m-th feature field. Afterward, through an embedding layer, $\boldsymbol{x}$ is mapped into

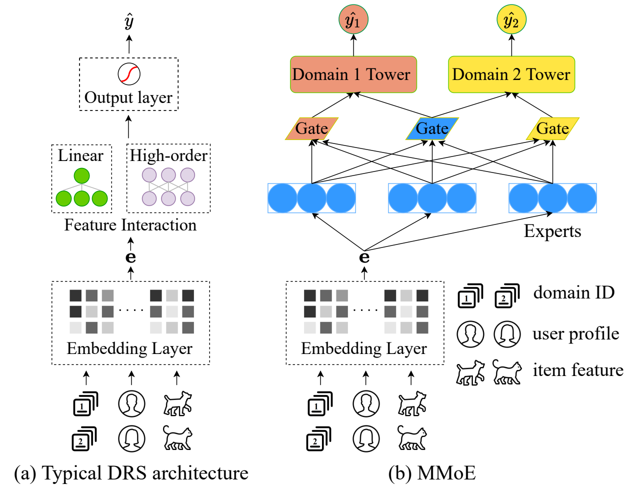

Figure 1: Illustration of multi-scenario recommendation models. Blue represents the shared parameters while red and yellow represent the scenario-specific parameters.

a low-dimensional vector $\boldsymbol{e} = [\boldsymbol{e}_1 \| \boldsymbol{e}_2 \| \ldots \| \boldsymbol{e}_M]$ where $\|$ denotes concatenation. Specifically, for the m-th feature field, $\boldsymbol{e}_m$ is obtained through a look-up operation $\boldsymbol{e}_m = \boldsymbol{E}_m \cdot \boldsymbol{x}_m$ where $\boldsymbol{E}_m \in \mathbb{R}^{u_m \times k}$ is the weight matrix, u_m is the count of feature values, and k is the embedding size. Finally, the prediction result $\hat{y}$ indicating whether a user would click on an item is calculated via $\hat{y} = f_d(\boldsymbol{e})$ where f_d denotes the recommendation model in the d-th domain. The loss function is Logloss [2] as same as the binary cross entropy loss:

$$\min_{\Theta} \mathcal{L} = -\frac{1}{B} \sum_{i=1}^{B} y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i), \tag{1}$$

where Θ is the set of learnable parameters, B is the number samples in the batch, and y_i and $\hat{y}_i$ are the true label and prediction result of the i-th sample, respectively.

2.2 Multi-Scenario Backbone Models

Most existing multi-scenario models can be categorized into two groups: ***Pre-train&Fine-tune*** and Multi-Task Learning (***MTL***) according to their training mechanism.

2.2.1 Pre-train and Fine-tune. As depicted in Figure 1(a), a typical deep recommender system (DRS) mainly adopts the feature interaction module to capture low-order and high-order interactions across feature fields followed by an output layer for the final prediction. Moreover, ***Pre-train&Fine-tune*** acts as a paradigm that it first pre-trains a DRS using all-domain data. Next it fine-tunes all parameters to adapt to each domain $d \in \{1, 2, \ldots, D\}$ using its own data and maintains an individual parameter set Θ_d.

2.2.2 Multi-Task Learning. Multi-task learning (MTL) aims at adopting a unified model to learn the related tasks to obtain their mutual improvement considering the knowledge shared [38]. Since it shares similar idea with the modeling of multi-scenario relations, a unified ***MTL*** model can be learned simply regarding different scenarios as different tasks in parallel. It models the commonality and distinction among domains through explicitly maintaining domain-shared parameters Θ_s and domain-specific parameters Θ_u. For example,

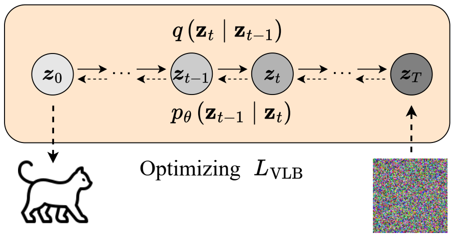

Figure 2: Illustration of diffusion model. The upper half with solid line arrow denotes forward process while the lower half with dashed arrow denotes reverse process.

as shown in Figure 1(b), MMoE [22] first leverages multiple networks at bottom named experts to extract knowledge, then feed it into the gating networks (i.e., shared and specific alike) to learn useful information for different scenarios. Finally, the information is assembled then passed into the domain-specific tower and the final prediction for all scenarios are outputted simultaneously.

2.3 Diffusion Model

Diffusion-based generative models have attracted significant attention in diverse fields including computer vision [5, 10, 25, 33] and text generation [7, 14]. They are proposed inspired by non-equilibrium thermodynamics in statistical physics [32] and they have demonstrated superior performance in high-quality data generation compared with GAN models [8], owing to their ability to achieve more stable training and diversity in generation [5]. Concretely, they operate by defining a Markov chain of diffusion steps that gradually introduce random noise to the data. Afterward, they learn the reverse of the diffusion process to generate the desired data samples from random noise. The illustration of diffusion model is depicted in Figure 2.

2.3.1 Forward Diffusion Process. In the forward diffusion process, for any data point sampled from the data distribution $\mathbf{z}_0 \sim q(\mathbf{z})$, specific Gaussian noise is added to the data in the next T steps and the distinguishable features are gradually dropped. Specifically, this process generates a noisy sequence $\{\mathbf{z}_1, \cdots, \mathbf{z}_T\}$.

$$q(\mathbf{z}_t \mid \mathbf{z}_{t-1}) = \mathcal{N}\left(\mathbf{z}_t; \sqrt{1-\beta_t}\mathbf{z}_{t-1}, \beta_t \mathbf{I}\right) \tag{2}$$

where $\{\beta_t \in (0,1)\}$ is the variance schedule that controls the forward noise injection step of the diffusion model. The key observation is that we can represent any $\mathbf{z}_t$ using the original sample $\mathbf{z}_0$ with the reparameterization trick. Define $C_t = 1 - \beta_t$ and $\bar{C}_t = \prod_{i=1}^{t} C_i$, we have:

$$\mathbf{z}_t = \sqrt{\bar{C}_t}\mathbf{z}_0 + \sqrt{1-\bar{C}_t}\epsilon \tag{3}$$

where $\epsilon \in \mathcal{N}(0, \mathbf{I})$. Then the closed form distribution is given by:

$$q(\mathbf{z}_t \mid \mathbf{z}_0) = \mathcal{N}\left(\mathbf{z}_t; \sqrt{\bar{C}_t}\mathbf{z}_0, \left(1-\bar{C}_t\right)\mathbf{I}\right) \tag{4}$$

2.3.2 Reverse Diffusion Process. The reverse process learns to recreate the data samples from Gaussian noise. Unfortunately, the estimation of the reverse process $q(\mathbf{z}_{t-1} \mid \mathbf{z}_t)$ is quite difficult since it is based on the whole dataset. Therefore, neural network is applied to estimate the corresponding conditional distribution p_θ.

$$p_\theta(\mathbf{z}_{t-1} \mid \mathbf{z}_t) = \mathcal{N}(\mathbf{z}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{z}_t, t), \Sigma_\theta(\mathbf{z}_t, t)) \tag{5}$$

where $\boldsymbol{\mu}_\theta$ and Σ_θ are learnable mean and variance.

The objective of the training is to generate a data distribution through the model that closely resembles the actual data distribution, which is equivalent to optimizing the negative log-likelihood through a variational lower bound. It can be further expressed as a combination of KL-divergence and entropy terms. Here we focus on the simplification of L_t written in the following form:

$$\begin{aligned} L_t^{\text{simple}} &= \mathbb{E}_{t\sim[1,T],\mathbf{z}_0,\epsilon_t}\left[\|\epsilon_t - \epsilon_\theta(\mathbf{z}_t, t)\|^2\right] \\ &= \mathbb{E}_{t\sim[1,T],\mathbf{z}_0,\epsilon_t}\left[\left\|\epsilon_t - \epsilon_\theta\left(\sqrt{\bar{C}_t}\mathbf{z}_0 + \sqrt{1-\bar{C}_t}\epsilon_t, t\right)\right\|^2\right] \end{aligned} \tag{6}$$

where ϵ_t is the Gaussian noise and $\epsilon_\theta(t)$ is the estimated Gaussian noise recovered from the original data z_0 at time-step t.

2.3.3 Detailed Setting. Considering the potential issues associated with approximating the diagonal variance matrix Σ_θ, such as unstable training dynamics and compromised sample quality [41], we adopt the denoising diffusion probabilistic models (DDPM) [10] where Σ_θ is held constant with $\Sigma_\theta(\mathbf{z}_t, t) = {\beta_t}^2\mathbf{I}$. Besides, we propose a novel approach that employs a piece-wise linear variance β_t to enhance the smoothness of the forward diffusion process, thereby improving the overall performance of the system. To estimate the reverse conditional distribution for the embedding under the recommendation setting, we employ the U-Net framework [28] with 1×1 convolutions, which can be regarded as a linear operation.

3 PROPOSED METHOD

In this section, the overall framework of the proposed Diff-MSR paradigm is first illustrated. Afterward, it is further detailed in four stages from Section 3.2 to 3.5.

3.1 Overall Framework

To address the limitations of existing MSR models, namely insufficient learning and negative transfer on the cold-start domains, we propose Diff-MSR as an enhanced paradigm compatible with ***Pre-train&Fine-tune*** and (***MTL***) as introduced in Section 2.2. Specifically, it is equipped with the designed piece-wise variance schedule and newly introduced classifier. Meanwhile, Diff-MSR is comprised of the following four stages: Pre-train, Diffusion, Classification, and Fine-tune. The procedure is summarized in Algorithm 1.

The illustration of Diff-MSR is depicted in Figure 3. Intuitively, from the perspective of computer vision, the hazy outline of a dog that closely resembles a cat may help the high-quality out-of-distribution (OOD) generation of figure in the cat domain. The reason is that the outline information is informative and significant in computer vision such as the typical semantic segmentation task [21, 26]. Similarly, under the recommendation setting, the embedding concatenation of user profile and item features in each instance can also be regarded as a 'figure'. Therefore, Diff-MSR is proposed based on this idea. With the help of the introduced classifier, the connection between the rich and cold-start domains is explicitly established through the noised embedding which consists of hazy while informative outline, and the cold-start domains are enhanced.

3.2 Pre-train Stage

Unlike the circumstance in computer vision that the pixel information of picture can be directly used as the common knowledge, it is challenging for the recommendation community since the raw

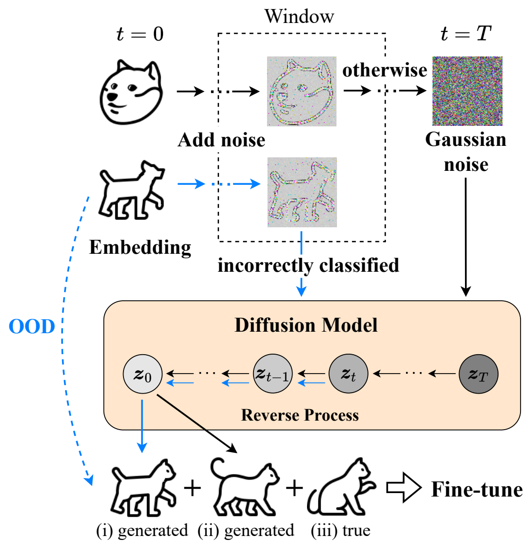

Figure 3: Illustration of Diff-MSR where the rich dog domain is used to help the OOD generation of cold-start cat domain. The user profile is omitted for simplicity. The data for fine-tuning consist of three parts as introduced in Section 3.5.

features of interaction have diverse representations in training different recommender systems. Consequently, Diff-MSR pre-trains on the data from all scenarios to obtain the shared embedding among scenarios at the first stage, then fix it as the common knowledge in the following stages. The procedure of this stage is summarized from Line 1 to Line 5 in Algorithm 1. Afterward, the pre-trained MSR backbone model is obtained, while it requires further improvement on the cold-start domains.

3.3 Diffusion Stage

It is difficult for the diffusion model to learn different interests and interaction patterns of user. Consequently, to grasp the distinctive distribution of data from the cold-start domains and conduct high-quality generation, Diff-MSR trains two diffusion models for click ($y = 1$) and unclick ($y = 0$) samples respectively on each cold-start domain $d_s \in D_s \subset \{1, 2, \ldots, D\}$ because the click and unclick signal reveals diverse user preference. Specifically, the embedding concatenation of user features and item attributes $\boldsymbol{x}$ is passed into the diffusion model as input. Particularly, the embedding of domain ID is treated as the scenario calibration and is not fed into the diffusion model for generation. Afterward, the two diffusion models are optimized via minimizing L_{VLB} in Equation (6). The procedure is listed from Line 6 to Line 13 in Algorithm 1.

Notably, we propose a new piece-wise variance schedule where $\{\beta_t\}$ remains small constant in the first few steps to gradually drop the detailed information while the high-quality outline information of embedding is maintained, then it linearly grows to ensure $\mathbf{z}_t$ is Gaussian noise at the end of forward diffusion process. It also helps to smoothen the forward diffusion process. We empirically validate its effectiveness in Section 4.4.

3.4 Classification Stage

The original diffusion model can only learn from in-domain distribution. Thus a classifier is introduced to capture different domain

Algorithm 1: Diff-MSR paradigm for multi-scenario recommendation taking CTR prediction as an example

Input: Domain ID $d \in \{1, 2, \ldots, D\}$; set of cold-start domain ID $D_s \subset \{1, 2, \ldots, D\}$; user and item features $\boldsymbol{x}$; true label of click y
Output: Well trained model for all domains
Stage 1: Pre-train
1 **while** *not converge* **do**
2   Sample a mini-batch from all-scenario training data;
3   Calculate the loss via Equation (1);
4   Take the gradient and update parameters of the multi-scenario backbone model;
5 **end**
Stage 2: Diffusion
6 **for** $d_s \in D_s$ **do**
7   **while** *not converge* **do**
8   Sample a mini-batch data samples on domain d_s;
9   Obtain the embedding by look-up operation;
10   Calculate loss via Equation (6);
11   Take the gradient and update parameters for $y = 0$ and $y = 1$, respectively;
12   **end**
13 **end**
Stage 3: Classification
14 **while** *not converge* **do**
15   Sample a mini-batch from all-scenario data;
16   Obtain the embedding by look-up operation;
17   Add Gaussian noise with random time-steps t;
18   Calculate the binary cross entropy loss;
19   Take the gradient and update parameters;
20 **end**
Stage 4: Fine-tune
21 **for** $d_s \in D_s$ **do**
22   **while** *not converge* **do**
23   Sample a mini-batch data instances on domain d_s;
24   Generate embedding from the incorrectly classified noised embedding from rich domains;
25   Generate embedding from Gaussian noise;
26   Mix the above instances;
27   Calculate loss via Equation (1);
28   Take the gradient and update specific parameters;
29   **end**
30 **end**

characteristics and serve as the link between the rich and cold-start domains. Concretely, it aims at distinguishing whether the embedding comes from the cold-start domain ($d \in D_s$) or not ($d \notin D_s$) as a binary classification task according to the user and item embedding concatenation $\boldsymbol{x}$, which contains domain-aware information like the domain-specific interest. Furthermore, it is trained with the original data and the noised embedding in different time-steps of forward process from all domains to capture the characteristic including commonality and distinction of various domains, then discriminate their domain origin. If a noised sample from the rich scenario is incorrectly judged to originates from the cold-start scenario, it would be selected and stored. Then in the next stage, it would be explicitly denoised by the reverse diffusion process of the diffusion model on the cold-start domain in a one-to-one manner because it is believed to have the outline information similar and beneficial to the cold-start domain. The procedure of this stage is illustrated from Line 14 to Line 20 in Algorithm 1.

3.5 Fine-tune Stage

Finally, to explicitly enhance the MSR backbone model, only the domain-specific parameters on top of the embedding layer are fine-tuned and enhanced on the cold-start domains without affecting the performance on the rich domains. Specifically, the instances for fine-tuning is comprised of the following three parts: **(i)** fake embedding one-to-one denoised by the diffusion models beginning with the incorrectly classified noised embedding from rich domains, **(ii)** fake embedding generated by the diffusion models starting from Gaussian noise, and **(iii)** true data from the cold-start domain as depicted at the bottom of Figure 3. For example, a noised instance $\mathbf{z}_t$ which starts from the embedding indicating a user click a dog is incorrectly judged to originate from the cold-start cat domain. It would be reversed by t steps and generate an embedding indicating a user click a cat to fine-tune the backbone model on the cat domain. By contrast, the noised embedding of clicking doge would not be selected and reversed. Overall, the procedure of this stage is demonstrated from Line 21 to Line 30 in Algorithm 1.

Notably, to ensure high-quality generation, a window of the time-step is set injecting noise to the embedding of true data from the rich domain. On the one hand, the embedding without much noise injected keeps abundant information in details from the rich domain, which has a negative effect on the generation in the reverse process of cold-start domain. On the other hand, the embedding with heavy noise added does not maintain informative outline information, which acts as a similar role as pure Gaussian noise.

4 EXPERIMENTS

We conduct extensive experiments on two public datasets to validate the effectiveness of our proposed Diff-MSR paradigm and answer the following three questions:

- **RQ1:** Is Diff-MSR effective and compatible with different multi-scenario backbone models as a paradigm?
- **RQ2:** How does Diff-MSR perform compared with the state-of-the-art generative baseline methods?
- **RQ3:** What are the effects of the introduced classifier and the proposed piece-wise variance schedule?

4.1 Experimental Settings

4.1.1 ***Datasets.*** Our experiments are conducted on two datasets: Douban [47] and Amazon 5-core [23] with three domains each. Their statistics and descriptions are summarized in **Appendix A.1**. Specifically, only users are partially overlapped in Douban while both users and items are partially overlapped in Amazon 5-core. Meanwhile, the **Music** domain in Douban and the **Beauty** domain in Amazon 5-core are regarded as the cold-start domain to be enhanced, respectively, while the other two domains are treated as the rich ones. Notably, the number of interactions of the other two rich domains is nearly 400 times and 10 times that of the cold-start domain on Douban and Amazon 5-core dataset, respectively.

4.1.2 ***Evaluation Metrics.*** We apply Area Under the ROC curve (AUC) to evaluate the performance of models on the test set, where a higher value at 0.001 level indicates significant improvement.

4.1.3 ***Backbone Models.*** We implement Diff-MSR on state-of-the-art multi-scenario backbone models as ***Pre-train&Fine-tune*** and ***MTL*** as illustrated in Section 2.2. Specifically, the first four methods belong to ***Pre-train&Fine-tune*** and the other four methods can be grouped into ***MTL***:

- **FNN** [45] introduces supervised-learning embedding layer to reduce the dimension of sparse features.
- **xDeepFM** [17] proposes Compressed Interaction Network (CIN) to explicitly learn explicit and implicit high-order interaction.
- **DCN** [36] proposes cross layers to learn cross features of bounded degree without manual feature engineering.
- **PNN** [27] designs a product layer to model inter-field interactive patterns and we choose the inner product operation.
- **MMoE** [22] leverages a group of expert networks and learns gating networks to assemble task-specific information.
- **PLE** [34] explicitly separates task-shared and task-specific experts and it employs a progressive routing mechanism in the multi-level extraction network.
- **STAR** [31] designs a star topology architecture with shared centered and domain-specific networks as a multi-scenario model.
- **AITM** [42] learns to adaptively transfer information with sequential dependence for different tasks.

4.1.4 ***Baselines.*** On one hand, intuitively, data augmentation can be leveraged to tackle the data sparsity problem of cold-start scenarios. However, existing works on data augmentation applied to recommendation community only adopt generative models under the setting of sequential recommendation [35, 43]. On the other hand, existing models of cold-start recommendation only consider new items or novel users [18], which is not applicable to multi-scenario recommendation since there are complex correlations among domains. Consequently, under the non-sequential and multi-scenario recommendation setting, we compare the proposed Diff-MSR with the following state-of-the-art generative models as baselines:

- **VAE** [12] uses a bottleneck encoder-decoder architecture and generates fake data by sampling from the latent space.
- **GAN** [8] trains a generative and discriminative model in an adversarial manner to generate plausible fake data.
- **Softmax GAN** [19] adopts a softmax cross-entropy loss rather than logistic loss for classification to stablize training for GAN.
- **WGAN** [1] adopts the Wasserstein distance between distributions and tackles the training instability of GANs.
- **WGAN-GP** [9] adds a gradient penalty and Lipschitz constraint on the aforementioned WGAN [1] to tackle the problem of non-convergence and low-quality generation.

4.1.5 ***Implementation Details.*** The implementation details of experiment are illustrated in **Appendix A.2** and the implementation code is available for easy reproduction[1, 2].

Table 1: Overall performance comparison of AUC value on the cold-start Music domain of Douban dataset. Boldface denotes the highest AUC and underline indicates the second best result. $\star$ indicates statistical significance with p-value < 0.05 in t-test.

Backbones	Mix	Fine-tune	Single	VAE	GAN	Softmax GAN	WGAN	WGAN-GP	Diff	Diff-MSR
FNN	0.5970	0.6141	0.6112	0.6220	0.6177	0.6153	0.6150	0.6064	0.6202	**0.6297***
xDeepFM	0.6029	0.6253	0.6046	0.6310	0.6255	0.6147	0.6194	0.6178	0.6366	**0.6434***
DCN	0.6130	0.6248	0.6240	0.6176	0.6212	0.6210	0.6224	0.6200	0.6298	**0.6368***
PNN	0.5998	0.6388	0.6034	0.6332	0.6395	0.6334	0.6353	0.6184	**0.6461**	0.6401*
MMoE	0.6212	/	/	0.6173	0.6194	0.6220	0.6186	0.6212	0.6260	**0.6274***
PLE	0.6236	/	/	0.6447	0.6236	0.6061	0.6590	0.6193	0.6682	**0.6769***
STAR	0.6435	/	/	0.6433	0.6450	0.6387	0.6407	0.6450	0.6453	**0.6459***
AITM	0.6294	/	/	0.6252	0.6298	0.6267	0.6265	0.6298	0.6310	**0.6577***

Table 2: Overall performance comparison of AUC value on the cold-start Beauty domain of Amazon 5-core. Boldface denotes the highest AUC and underline indicates the second best result. $\star$ indicates statistical significance with p-value < 0.05 in t-test.

Backbones	Mix	Fine-tune	Single	VAE	GAN	Softmax GAN	WGAN	WGAN-GP	Diff	Diff-MSR
FNN	0.5966	0.5975	0.5768	0.5869	0.5874	0.5895	0.5903	0.5913	0.5999	**0.6012***
xDeepFM	0.5925	0.5959	0.5796	0.5945	0.5943	0.5954	0.5958	0.5968	0.5988	**0.5994***
DCN	0.5922	0.5940	0.5751	0.5889	0.5896	0.5855	0.5874	0.5879	0.5957	**0.6000***
PNN	0.5959	0.5978	0.5760	0.5924	0.5952	0.5961	0.5961	0.5965	0.5988	**0.5993***
MMoE	0.5917	/	/	0.5871	0.5876	0.5912	0.5885	0.5905	0.5953	**0.5958***
PLE	0.6034	/	/	0.5993	0.5925	0.5974	0.5965	0.5953	0.6069	**0.6079***
STAR	0.5799	/	/	0.5811	0.5788	0.5806	0.5812	0.5802	0.5815	**0.5819***
AITM	0.5969	/	/	0.5858	0.5900	0.5928	0.5929	0.5954	0.5996	**0.6004***

4.2 Compatibility with Backbone Models (RQ1)

Particularly, Diff-MSR does not impose any restrictions on the model architecture or optimization of the multi-scenario backbone models. To answer **RQ1**, we validate Diff-MSR's compatibility with typical MSR backbone models in ***Pre-train&Fine-tune*** and ***MTL*** training paradigm. The overall performance of AUC on Douban and Amazon 5-core dataset is shown in Table 1 and 2, respectivly.

In Table 1 and 2, the first four backbones belong to ***Pre-train&Fine-tune*** while the last four belong to ***MTL***. '`Mix`' denotes training on all-domain data, '`Fine-tune`' denotes further updating all parameters of '`Mix`' on each domain, and '`Single`' represents training on the cold-start-domain data only. The last two patterns are not applicable for ***MTL*** models because they directly train a unified model. Generally, the large discrepancy in data volume among domains results in the poor performance on the cold-start scenario. Specifically, on the one hand, the phenomenon of negative transfer is common on Douban dataset as '`Mix`' generally performs worse than '`Single`'. This is because only users are partially overlapped among domains. However, it is not seen on Amazon 5-core dataset because both users and items are partially overlapped thus more useful information is transferred and improves the recommendation accuracy on the cold-start Beauty domain. On the other hand, insufficient learning occurs on both datasets since '`Fine-tune`' always achieves better performance than '`Mix`'. This is because the model is adequately trained and further adapted to the in-domain distinction of the cold-start domain.

We also have the following observations. First, Diff-MSR significantly outperforms '`Mix`' and '`Fine-tune`' because the information of commonality and heterogeneity across domains is captured by the classifier owing to the capability to discriminate domain origin, then it is explicitly transferred from the rich domains through the reverse process of diffusion model to enhance the cold-start domain. Second, the improvement on Douban dataset is much larger than that on Amazon 5-core dataset, because the data volume of the Music domain is relatively much lower than the Beauty domain among all scenarios. Meanwhile, less information is shared on Douban dataset because only the knowledge of overlapped users can be transferred. Thus Diff-MSR can relieve the problem of insufficient learning and negative transfer more remarkably on Douban dataset.

To summarize, we can safely draw the conclusion that Diff-MSR is able to improve the poor recommendation performance on the cold-start domains without affecting the accuracy of recommendation on other rich domains since only the domain-specific parameters of the cold-start domains are updated. Moreover, Diff-MSR is compatible with most existing multi-scenario backbone models and acts as an effective and powerful enhanced paradigm.

4.3 Comparison with Baseline Methods (RQ2)

To answer **RQ2**, we compare Diff-MSR with five generative models as baseline methods detailed in Section 4.1.4, which has achieved excellent performance especially on computer vision tasks [41]. However, these state-of-the-art generative methods including the original diffusion model suffer from the severe limitation that they are only capable of generating data from the learned in-domain distribution, while they are unable to leverage the data from rich domains to help the high-quality out-of-distribution generation of the cold-start domain. Their comparison results on Douban and Amazon 5-core dataset are also shown in Table 1 and 2.

First, on Douban dataset, the baseline methods successfully improve the recommendation performance on the cold-start domain in 35% (14/40) of all the cases, indicating their unstable performance to generate in-domain embedding for the cold-start domain. Furthermore, on Amazon 5-core dataset, all these baseline methods fail to achieve a comparable performance of 'Mix' on most backbone models except STAR, let alone improving 'Fine-tune'. This means that they are unable to comprehend the distribution of embedding in a more complex situation. By contrast, the proposed Diff-MSR significantly enhances the recommendation accuracy based on all these backbone models where the maximum improvement is up to 8.5% and 1.0% on the two datasets. The reason is that Diff-MSR employs the high-quality generation capability of diffusion model and explicitly leverages the informative outline information of the embedding to capture the commonality and difference between the rich and cold-start domains. Meanwhile, the generation and training process of diffusion model is more stable than VAE and GAN, though it requires more time for training and sampling.

4.4 Ablation Study (RQ3)

To answer **RQ3**, we conduct the ablation study from two aspects. On the one hand, to investigate the effect of the classifier, we compare Diff-MSR with 'Diff', which is not equipped with the classifier and only adopts the diffusion model to generate high-quality embedding for the cold-start domain. As shown in Table 1 and 2, 'Diff' achieves better results than 'Mix' and 'Fine-tune' based on all backbone models on both datasets while other generative models can not. Therefore it demonstrates the high-quality generation capability of the diffusion model. In addition, Diff-MSR significantly outperforms 'Diff' except on the PNN model on Douban dataset. The reason is that the classifier acts as a bridge connecting the rich and cold-start domains, and it explicitly selects the incorrectly classified embedding to be denoised through the reverse diffusion process of the cold-start domain. By contrast, even though the diffusion model possesses the capability of high-quality generation, limited knowledge of data distribution could be learned from the original cold-start domains while the informative outline instances from rich domains deserve exploiting and mining in a data-driven manner.

On the other hand, to verify the effectiveness of the designed piece-wise variance schedule in the diffusion model, we compare it with the commonly used linear and cosine variance schedule. The result in Figure 4(a) shows that the piece-wise schedule significantly outperforms the other two methods on FNN and MMoE as two cases of the backbone model, indicating that the designed piece-wise variance schedule is more suitable for high-quality embedding generation in recommendation. This is because it properly maintains the informative embedding outline information in the first few steps of the forward diffusion process as shown in Figure 5(a) and (b). Specifically, we can see that as more noise is injected to the original embedding of the backbone model, it becomes more difficult for the classifier to discriminate the domain origin as the embedding turns out to be pure Gaussian noise in around 200 steps with an AUC of 0.5. The slope of the piece-wise schedule is less steeper than the other two schedules in the first 100 steps, meaning that it keeps the discriminable outline while gradually drops the details in a better way of learning the generation of embedding in the recommendation community.

4.5 Hyper-parameter Analysis

We also take FNN and MMoE as the representatives of the MSR backbone models in the two groups to analyze the effect of an important hyper-parameter, i.e., the choice of the training objective

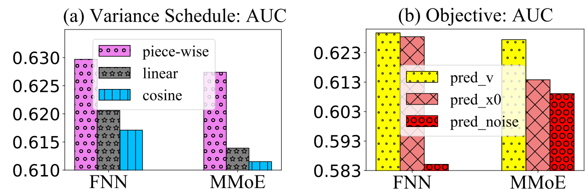

Figure 4: Ablation study and hyper-parameter analysis of Diff-MSR on Douban dataset. The result is the AUC value on the cold-start Music domain.

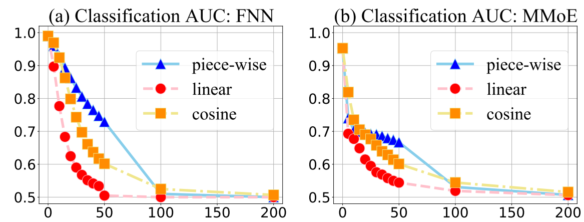

Figure 5: Performance comparison of the classifier on the noised embedding in different forward time-steps, which is evaluated by the AUC metric on all-domain data of Douban.

of the diffusion model on Douban dataset. Specifically, 'pred_v' denotes the v-parameterization proposed in [29], 'pred_x0' denotes predicting the start of embedding, and 'pred_noise' denotes predicting the noise [10]. We can see from Figure 4(b) that 'pred_v' outperforms the other two learning objectives on both backbone models while 'pred_noise' achieves the worst recommendation accuracy. Hence 'pred_v' is suggested for embedding generation.

5 RELATED WORK

In this section, we provide a summary of the recent work on multi-scenario recommendation and the diffusion model applied to the recommendation community.

5.1 Multi-Scenario Recommendation

Multi-scenario recommendation (MSR) [48] aims at using the historical interactions from multiple domains to simultaneously improve their recommendation performance. It is also known as multi-domain and multi-target cross-domain recommendation in industry and research community. For example, SAR-Net [30] adopts attention layers to capture the cross-scenario interest. HMoE [13] uses a multi-gate mixture-of-experts to implicitly learn the connection between domains and explicitly learn task relations with a stacked model. In addition, it is common to regard different scenarios as different tasks and implement a unified multi-task framework [46].

However, the existing MSR methods mainly focus on learning the commonality and heterogeneity across domains by designing different parameter-sharing patterns (e.g., shared and domain-specific parameters) in a model-driven manner. Nevertheless, they overlook the imbalanced data across domains and the significance of cold-start scenarios, leading to insufficient learning and negative transfer revealed by the poor performance on these cold-start scenarios [15, 40]. By contrast, our proposed Diff-MSR paradigm is the first to tackle these problems and achieve enhancement on the cold-start domains for different MSR backbone models. Not only can it grasp the characteristic of in-domain data, but it also establishes the connection between the rich and cold-start domains through the noised embedding of samples and the designed classifier.

5.2 Diffusion Model with Recommendation

In recent years, diffusion model has shown its remarkable generation capabilities in computer vision [3] and natural language processing [14]. In the recommendation community, several works focus on incorporating diffusion models to enhance the recommendation quality and improve user experience [6, 16, 20, 37, 39]. For example, DiffuRec [16] employs the diffusion model with an approximator to model the latent representation of items and the multi-interest of users. DiffRec [37] is proposed which reduces the cost for prediction and enables temporal modeling of the interaction sequence. Notably, the personalized information is retained at end of the forward diffusion process $\mathbf{z}_T$ allowing for more accurate and personalized recommendations.

However, all these methods only apply the diffusion model to the sequential recommendation, which relies on the specific sequence dependency as in natural language processing. Meanwhile, they only focus on dealing with either user embedding or item representation only and simply leverage the diffusion model to introduce randomness. By contrast, in this paper, we are the first to investigate a non-sequential and multi-scenario recommendation setting regarding the embedding concatenation of both user profiles and item attributes as a 'figure' in computer vision, which retains the label information of domain ID and the click signal.

6 CONCLUSION

This paper is the first to address the cold-start problem in multi-scenario recommendations. Specifically, we propose an enhanced paradigm Diff-MSR which leverages the diffusion model with the newly designed piece-wise variance schedule and the introduced classifier to explicitly improve the performance on the cold-start scenarios. Extensive experiments on Douban and Amazon 5-core datasets show that our proposed method is significantly effective and compatible with different multi-scenario backbone models. It also surpasses various state-of-the-art generative baseline methods and achieves more stable performance. In future work, to produce high-quality embedding for recommender systems more efficiently, DDIM [33] and guided diffusion model [24] can be considered.

A EXPERIMENTAL SETTINGS

A.1 Datasets

The statistics of Douban and Amazon 5-core datasets are summarized in Table 3. Specifically, we extract the most recent three interacted items for each user in the cold-start domain to simulate the great difference in the number of historical interactions between domains in industrial multi-scenario recommendation. Sparsity denotes the proportion of negative samples with $y = 0$, which is uncorrelated with the data volume.

- **Douban**[3] dataset is crawled from Douban in which only users are partially overlapped. For all three domains, we randomly divide the data into training, validation, and test set with the ratio of 8:1:1. User ID, item ID, and domain ID are used as feature fields to predict whether a user gives a rating (ranged from 1 to 5) higher than 3 to an item as click.

Table 3: The statistics of Douban and Amazon 5-core. † represents the cold-start domain.

Dataset	Douban			Amazon 5-core		
Domains	Music†	Book	Movie	Clothing	Beauty†	Health
Users	1672	2110	2718	39387	22363	38609
Items	5567	6777	9565	23033	12101	18534
Interactions	3123	96041	1133420	278677	62455	346355
Sparsity	21.71%	27.94%	39.09%	20.48%	23.20%	19.22%

Table 4: The hyper-parameter settings of experiments

Dataset	Douban	Amazon 5-core
Embedding dimension (k)	16	16
Batch size	512	2048
Hidden layers of backbone	[256,256,256]	[512,512,512]
Hidden layers of U-net	[16,32,64,128]	[16,32,64,128]
Hidden layers of classifier	[64,64]	[128,128]
Diffusion step	500	500
Learning rate of train	1e-3	5e-4
Learning rate of test	5e-4	1e-4
Learning rate of diffusion	2e-3	2e-3
Learning rate of classifier	2e-3	1e-3

- **Amazon 5-core**[4] is a dense subset from Amazon in which all users and items have at least 5 interactions. We follow the experiment setting of HeroGRAPH [4] and select three related domains with both overlapped users and items. Besides, the data is split into training, validation, and test set by the time of 1st March 2014 and 30th April 2014.

A.2 Implementation Details

The hyper-parameters are listed in Table 4. Moreover, Adam [11] optimizer is used to minimize the binary cross entropy loss. For diffusion, the U-net with Gaussian diffusion is implemented with two channels for user and item embedding respectively. For piece-wise variance schedule, it remains 2e-4 in the first 50 steps and linearly increases to 0.05. For the window of time-step to add noise, it is empirically set to be $(30, 50)$ for ***Pre-train&Fine-tune*** and $(20, 40)$ for ***MTL***. Moreover, the classifier is simply a fully connected neural network and the training data from rich and cold-start domains need to be balanced. To enhance the multi-scenario backbone model, the ratio of the true data to the generated embedding is 3:1. All experimental results are averaged over 3 runs on Tesla V100 GPU.

ACKNOWLEDGEMENT

This research was partially supported by Huawei Innovation Research Program, APRC - CityU New Research Initiatives (No.9610565, Start-up Grant for New Faculty of City University of Hong Kong), CityU - HKIDS Early Career Research Grant (No.9360163), Hong Kong ITC Innovation and Technology Fund Midstream Research Programme for Universities Project (No.ITS/034/22MS), Hong Kong Environmental and Conservation Fund (No. 88/2022), and SIRG - CityU Strategic Interdisciplinary Research Grant (No.7020046, No.7020074).

[3] https://github.com/FengZhu-Joey/GA-DTCDR/tree/main/Data

[4] https://cseweb.ucsd.edu/~jmcauley/datasets.html#amazon_reviews

REFERENCES

[1] Martin Arjovsky, Soumith Chintala, and Léon Bottou. 2017. Wasserstein generative adversarial networks. In *International conference on machine learning*. PMLR, 214–223.

[2] Joseph Berkson. 1944. Application of the logistic function to bio-assay. *Journal of the American statistical association* 39, 227 (1944), 357–365.

[3] Florinel-Alin Croitoru, Vlad Hondru, Radu Tudor Ionescu, and Mubarak Shah. 2023. Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023).

[4] Qiang Cui, Tao Wei, Yafeng Zhang, and Qing Zhang. 2020. HeroGRAPH: A Heterogeneous Graph Framework for Multi-Target Cross-Domain Recommendation.. In *ORSUM@ RecSys*.

[5] Prafulla Dhariwal and Alexander Nichol. 2021. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems* 34 (2021), 8780–8794.

[6] Hanwen Du, Huanhuan Yuan, Zhen Huang, Pengpeng Zhao, and Xiaofang Zhou. 2023. Sequential Recommendation with Diffusion Models. *arXiv preprint arXiv:2304.04541* (2023).

[7] Shansan Gong, Mukai Li, Jiangtao Feng, Zhiyong Wu, and LingPeng Kong. 2022. Diffuseq: Sequence to sequence text generation with diffusion models. *arXiv preprint arXiv:2210.08933* (2022).

[8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2020. Generative adversarial networks. *Commun. ACM* 63, 11 (2020), 139–144.

[9] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. 2017. Improved training of wasserstein gans. *Advances in neural information processing systems* 30 (2017).

[10] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems* 33 (2020), 6840–6851.

[11] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv:1412.6980* (2014).

[12] Diederik P Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).

[13] Pengcheng Li, Runze Li, Qing Da, An-Xiang Zeng, and Lijun Zhang. 2020. Improving multi-scenario learning to rank in e-commerce by exploiting task relationships in the label space. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 2605–2612.

[14] Xiang Li, John Thickstun, Ishaan Gulrajani, Percy S Liang, and Tatsunori B Hashimoto. 2022. Diffusion-lm improves controllable text generation. *Advances in Neural Information Processing Systems* 35 (2022), 4328–4343.

[15] Xiaopeng Li, Fan Yan, Xiangyu Zhao, Yichao Wang, Bo Chen, Huifeng Guo, and Ruiming Tang. 2023. HAMUR: Hyper Adapter for Multi-Domain Recommendation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. 1268–1277.

[16] Zihao Li, Aixin Sun, and Chenliang Li. 2023. DiffuRec: A Diffusion Model for Sequential Recommendation. *arXiv preprint arXiv:2304.00686* (2023).

[17] Jianxun Lian, Xiaohuan Zhou, Fuzheng Zhang, Zhongxia Chen, Xing Xie, and Guangzhong Sun. 2018. xdeepfm: Combining explicit and implicit feature interactions for recommender systems. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 1754–1763.

[18] Blerina Lika, Kostas Kolomvatsos, and Stathes Hadjiefthymiades. 2014. Facing the cold start problem in recommender systems. *Expert systems with applications* 41, 4 (2014), 2065–2073.

[19] Min Lin. 2017. Softmax gan. *arXiv preprint arXiv:1704.06191* (2017).

[20] Qidong Liu, Fan Yan, Xiangyu Zhao, Zhaocheng Du, Huifeng Guo, Ruiming Tang, and Feng Tian. 2023. Diffusion Augmentation for Sequential Recommendation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. 1576–1586.

[21] Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3431–3440.

[22] Jiaqi Ma, Zhe Zhao, Xinyang Yi, Jilin Chen, Lichan Hong, and Ed H Chi. 2018. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 1930–1939.

[23] Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton Van Den Hengel. 2015. Image-based recommendations on styles and substitutes. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*. 43–52.

[24] Alex Nichol, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob McGrew, Ilya Sutskever, and Mark Chen. 2021. Glide: Towards photorealistic image generation and editing with text-guided diffusion models. *arXiv preprint arXiv:2112.10741* (2021).

[25] Alexander Quinn Nichol and Prafulla Dhariwal. 2021. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*. PMLR, 8162–8171.

[26] Hyeonwoo Noh, Seunghoon Hong, and Bohyung Han. 2015. Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision*. 1520–1528.

[27] Yanru Qu, Han Cai, Kan Ren, Weinan Zhang, Yong Yu, Ying Wen, and Jun Wang. 2016. Product-based neural networks for user response prediction. In *2016 IEEE 16th international conference on data mining (ICDM)*. IEEE, 1149–1154.

[28] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 234–241.

[29] Tim Salimans and Jonathan Ho. 2022. Progressive distillation for fast sampling of diffusion models. *ICLR* (2022).

[30] Qijie Shen, Wanjie Tao, Jing Zhang, Hong Wen, Zulong Chen, and Quan Lu. 2021. SAR-Net: A Scenario-Aware Ranking Network for Personalized Fair Recommendation in Hundreds of Travel Scenarios. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 4094–4103.

[31] Xiangrong Sheng, Liqin Zhao, Guorui Zhou, Xinyao Ding, Binding Dai, Qiang Luo, Siran Yang, Jingshan Lv, Chi Zhang, Hongbo Deng, et al. 2021. One model to serve all: Star topology adaptive recommender for multi-domain ctr prediction. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 4104–4113.

[32] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*. PMLR, 2256–2265.

[33] Jiaming Song, Chenlin Meng, and Stefano Ermon. 2020. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502* (2020).

[34] Hongyan Tang, Junning Liu, Ming Zhao, and Xudong Gong. 2020. Progressive layered extraction (ple): A novel multi-task learning (mtl) model for personalized recommendations. In *Fourteenth ACM Conference on Recommender Systems*. 269–278.

[35] Qinyong Wang, Hongzhi Yin, Hao Wang, Quoc Viet Hung Nguyen, Zi Huang, and Lizhen Cui. 2019. Enhancing collaborative filtering with generative augmentation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 548–556.

[36] Ruoxi Wang, Bin Fu, Gang Fu, and Mingliang Wang. 2017. Deep & cross network for ad click predictions. In *Proceedings of the ADKDD'17*. 1–7.

[37] Wenjie Wang, Yiyan Xu, Fuli Feng, Xinyu Lin, Xiangnan He, and Tat-Seng Chua. 2023. Diffusion Recommender Model. *arXiv preprint arXiv:2304.04971* (2023).

[38] Yuhao Wang, Ha Tsz Lam, Yi Wong, Ziru Liu, Xiangyu Zhao, Yichao Wang, Bo Chen, Huifeng Guo, and Ruiming Tang. 2023. Multi-Task Deep Recommender Systems: A Survey. *arXiv preprint arXiv:2302.03525* (2023).

[39] Yu Wang, Zhiwei Liu, Liangwei Yang, and Philip S Yu. 2023. Conditional Denoising Diffusion for Sequential Recommendation. *arXiv preprint arXiv:2304.11433* (2023).

[40] Yuhao Wang, Xiangyu Zhao, Bo Chen, Qidong Liu, Huifeng Guo, Huanshuo Liu, Yichao Wang, Rui Zhang, and Ruiming Tang. 2023. PLATE: A Prompt-Enhanced Paradigm for Multi-Scenario Recommendations. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1498–1507.

[41] Zhengwei Wang, Qi She, and Tomas E Ward. 2021. Generative adversarial networks in computer vision: A survey and taxonomy. *ACM Computing Surveys (CSUR)* 54, 2 (2021), 1–38.

[42] Dongbo Xi, Zhen Chen, Peng Yan, Yinger Zhang, Yongchun Zhu, Fuzhen Zhuang, and Yu Chen. 2021. Modeling the sequential dependence among audience multi-step conversions with multi-task learning in targeted display advertising. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 3745–3755.

[43] Zhe Xie, Chengxuan Liu, Yichi Zhang, Hongtao Lu, Dong Wang, and Yue Ding. 2021. Adversarial and contrastive variational autoencoder for sequential recommendation. In *Proceedings of the Web Conference 2021*. 449–459.

[44] Senrong Xu, Liangyue Li, Yuan Yao, Zulong Chen, Han Wu, Quan Lu, and Hanghang Tong. 2023. MUSENET: Multi-Scenario Learning for Repeat-Aware Personalized Recommendation. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*. 517–525.

[45] Weinan Zhang, Tianming Du, and Jun Wang. 2016. Deep learning over multi-field categorical data. In *European conference on information retrieval*. Springer, 45–57.

[46] Yuanliang Zhang, Xiaofeng Wang, Jinxin Hu, Ke Gao, Chenyi Lei, and Fei Fang. 2022. Scenario-Adaptive and Self-Supervised Model for Multi-Scenario Personalized Recommendation. In *Proceedings of CIKM 2022*.

[47] Feng Zhu, Chaochao Chen, Yan Wang, Guanfeng Liu, and Xiaolin Zheng. 2019. Dtcdr: A framework for dual-target cross-domain recommendation. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 1533–1542.

[48] Feng Zhu, Yan Wang, Chaochao Chen, Jun Zhou, Longfei Li, and Guanfeng Liu. 2021. Cross-domain recommendation: challenges, progress, and prospects. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence*.