SeeDRec: Sememe-based Diffusion for Sequential Recommendation

Haokai Ma 1, Ruobing Xie 3*, Lei Meng 1,2*, Yimeng Yang 1, Xingwu Sun 3, Zhanhui Kang 3
1 School of Software, Shandong University, China
2 Shandong Research Institute of Industrial Technology, China
3 Tencent, China

mahaokai@mail.sdu.edu.cn, ruobingxie@tencent.com, lmeng@sdu.edu.cn, y_yimeng@mail.sdu.edu.cn, sunxingwu01@gmail.com, kegokang@tencent.com

Abstract

Inspired by the power of Diffusion Models (DM) verified in various fields, some pioneering works have started to explore DM in recommendation. However, these prevailing endeavors commonly implement diffusion on item indices, leading to the increasing time complexity, the lack of transferability, and the inability to fully harness item semantic information. To tackle these challenges, we propose SeeDRec, a sememe-based diffusion framework for sequential recommendation (SR). Specifically, inspired by the notion of sememe in NLP, SeeDRec first defines a similar concept of recommendation sememe to represent the minimal interest unit and upgrades the specific diffusion objective from the item level to the sememe level. With the Sememe-to-Interest Diffusion Model (S2IDM), SeeDRec can accurately capture the user’s diffused interest distribution learned from both local interest evolution and global interest generalization while maintaining low computational costs. Subsequently, an Interest-aware Prompt-enhanced (IPE) strategy is proposed to better guide each user’s sequential behavior modeling via the learned user interest distribution. Extensive experiments on nine SR datasets and four cross-domain SR datasets verify its effectiveness and universality. The code is available in https://github.com/hulkima/SeeDRec.

1 Introduction

Recommender system (RS) has become essential for many real-world applications owing to its ability to accurately mine users’ personalized interests [Wang et al., 2015; Fan et al., 2023; Meng et al., 2020; Ma et al., 2021; Ma et al., 2023a]. Recently, numerous advanced techniques in NLP have been assimilated into RS, continuing to demonstrate their impact [Kang and McAuley, 2018; Sun et al., 2019; Ma et al., 2023b]. The sequential modeling process of user behaviors in sequential recommendation (SR) bears semblance to the language modeling task prevalent in NLP, that is, SR seeks to recommend the next-item that user may be interested in by modeling the sequential dependencies of user’s temporal behaviors [de Souza Pereira Moreira et al., 2021].

As motivated by the outstanding distribution generative performance of the Diffusion Model (DM) in various domains (image synthesis [Ho et al., 2020], audio processing [Kong et al., 2021] and semantic segmentation [Brempong et al., 2022]), some pioneering studies attempt to explore the effectiveness of DM in recommendation. DiffRec [Wang et al., 2023e] conducts DM on item indices to infer users’ interaction probabilities in a denoising manner, CDDRec [Wang et al., 2023f], DiffuRec [Li et al., 2023], DreamRec [Yang et al., 2023] and DDRM [Zhao et al., 2024] consider injecting uncertainty into item embeddings via DM under the generative paradigm. PDRRec [Ma et al., 2024a] enhances DiffRec by leveraging the pre-trained DM on item indices as plugins to improve SR, which exhibits universality and analogous inference costs to the sequence encoders. However, the ID-based methods typically exhibit: (1) a linear increased complexity with the growing number of corpus, leading to the worse scalability in large-scale RS, (2) an excessive dependency on item indices, lacking the cross-domain transferability, (3) a limitation in fully utilizing the semantic correlations within items, thereby resulting in sub-optimal performance.

To tackle these challenges, an intuitive idea is to conduct DM on other objectives that are atomic, multi-domain transferable, and encapsulate the semantic correlations requisite for SR, rather than item indices. We notice the definition of sememe, which is regarded as the minimum semantic unit in NLP [Niu et al., 2017]. NLP researchers believe that each word can be decomposed into a limited set of manually-defined and language-independent sememes [Qi et al., 2022]. Inspired by this, we creatively propose the concept of recommendation sememe (abbreviated as sememe) to represent users’ minimal interest unit in recommendation. As shown in Fig. 1, all items can be represented as the combination of...

Figure 1: Illustration of the proposed SeeDRec. Our sememe-based diffusion enhances the scalability, transferability, and usage of item semantics, while keeping good performance and universality.
several sememes, thereby reducing the quantity of DM objectives, accomplishing the transferability across diverse domains and incorporating semantic correlations among items in SR. To do this, we present a model-agnostic Sememe-based Diffusion framework for Sequential Recommendation (SeeDRec). It regards each user’s original interest on sememes observed by the system as a “Seed”, nourishes it with both local and global user preferences, and cultivates the “Blooming flower” of generalized interest distribution via our sememe-based diffusion. Specifically, we first propose the Sememe-to-Interest Diffusion Model (S2IDM) to upgrade the existing DM-based recommenders from item index to sememe, mining each user’s generalized interest distribution by synergistically considering temporal, frequency, and co-occurrence information interlinking sememes. The user generalized interest distribution learned by S2IDM contains both the user’s local personalized behavioral preference and the global sememe correlations implied by all user behaviors. Furthermore, to incorporate the generalized interest distribution into the sequential modeling, we design an Interest-aware Prompt-enhanced (IPE) strategy to guide the sequential modeling towards the direction for better personalized behavior understanding. With the sememes powered by S2IDM and IPE, SeeDRec could achieve improvement on various base models as a plugin with better scalability and interest generalization. Moreover, the knowledge of user interest diffusion encoded in S2IDM can be transferred to other domains via our proposed recommendation sememe anchors.

We conduct extensive experiments on nine SR datasets and four cross-domain SR (CDSR) datasets to demonstrate the superiority of SeeDRec. For simplification and universality, we also use existing taxonomies and words as sememes to simulate practical scenarios. We further conduct various ablation studies, universality analyses, few-shot analyses, and model analyses to validate the effectiveness of the proposed S2IDM and IPE. The contributions of this work are as follows:

- We have verified the feasibility of conducting diffusion on the recommendation sememe and incorporating it in SR. To the best of our knowledge, we are the first to conduct DM on sememes to enhance SR and CDSR tasks.
- We propose S2IDM to mine diffused user generalized interest distribution by simultaneously considering global sememe correlations and users’ local sequential behaviors.
- We creatively present the IPE strategy to enable the precise interest transfer from the discrete sememe interest distribution to the continuous representation to improve the personalized sequential modeling via a prompt learning paradigm.
- We conduct extensive evaluations on 13 real-world datasets to verify that the proposed SeeDRec is effective, universal, and easy-to-deploy. We also design comprehensive analyses to demonstrate the effective mechanism of SeeDRec and validate it in the more challenging few-shot scenarios.

2 Related Work

Sequential Recommendation Benefiting from the advancement of sequential modeling techniques, SR is proposed to format and encode user temporal behaviors to infer their dynamic interests [Sun et al., 2022; Wang et al., 2023g; Wang et al., 2023d; Sun et al., 2024]. Within its evolutionary trajectory, early SR methods reason users’ short-term preference through the Markov Chains [Rendle et al., 2010]. GRU4Rec [Hidasi et al., 2016] leverages the Gate Recurrent Unit as the sequential encoder to capture users’ long-term dependencies. Caser [Tang and Wang, 2018] imports the Convolutional Neural Network (CNN) to extract sequential patterns from user behaviors at different time intervals. Inspired by the success of self-attention mechanism, SASRec [Kang and McAuley, 2018] incorporates it to jointly model users’ short- and long-term preferences. CLASRec [Xie et al., 2022] and MStein [Fan et al., 2023] additionally utilize the mutual information on self-supervised signals to improve the sequential representations. Unlike the complex network structures and stochastic augmentations in previous SR works, SeeDRec smartly leverages the prevailing diffusion models to incorporate the generalized user interests in sequential pattern modeling. Furthermore, SeeDRec can be effortlessly deployed on diverse SR models with advanced techniques to bring significant and consistent performance improvements (see Sec. 4.5).

Diffusion Model for Recommendation Inspired by the remarkable performance in image generation [Ho et al., 2020; Nichol and Dhariwal, 2021; Wang et al., 2023b], scene text editing [Wang et al., 2023a], and machine translation [Chen et al., 2023], some researchers attempt to incorporate DM in recommendation. CODIGEM [Walker et al., 2022] is the pioneer DM-based recommender which generates improved collaborative signals via the Denoising Diffusion Probabilistic Model (DDPM). DiffRec [Wang et al., 2023e] reduces the scheduled noises into the reverse process to infer user interaction probabilities in a denoising manner. PDRRec [Ma et al., 2024a] is the state-of-the-art DM-based recommender, which takes the DM trained on all corpus as the plugin in SR to fully leverage the diffusion-based user preferences. Moreover, a series of related DM-based works (e.g., DiffuRec [Li et al., 2023], DiffRec* [Du et al., 2023] and DreamRec [Yang et al., 2023]) employ DM on the continuous item embedding space with additional transitions and noising strategies, which shares the same modeling pipeline with classical SR methods. Disparate objectives make them inherently non-comparable with this work, and they can serve as the base SR model within our SeeDRec. In particular, we have employed PDRRec as the backbone to substantiate the efficacy of SeeDRec over DM-based recommenders.

In conclusion, existing DM-based methods typically conduct DM on the discrete item indices or the continuous embedding of each item to achieve the uncertainty injection in recommendation. Therefore, the augmentation of the number of items exacerbates the multiplication of their time and space complexities, rendering them arduous to deploy in real-world million-level recommendation systems. Moreover, these DM-based recommenders solely rely on item indices, demonstrating the lack of scalability and the inability to fully leverage the semantic correlations within items. Every instance of encountering a new dataset compels them to be trained from scratch without leveraging the transferability across multiple domains, leading to the recurrent process that demands substantial compility and time consumption.
3 Methods

3.1 Problem Statement

In this article, we focus on exploring the DM-enhanced recommendation task in multiple recommendation scenarios, with an illustration of the example of SR. We define the behavior sequence $S_p^u = \{i_1, i_2, \ldots, i_p\}$ of user $u \in U$, where $i_j$ is the $j$-th interacted item of user $u$ and $p$ denote the historical behavior length. Given $S_p^u$, we adopt SASRec [Kang and McAuley, 2018] as the sequential encoder to predict the target item $i_{p+1}$ that will be preferred by user $u$.

3.2 Overall Structure

In this section, we elaborate on the proposed model-agnostic Sememe-based Diffusion for sequential recommendation, which utilizes the DM on sememes to explore each user’s atomized multi-interest units for the multiple recommendation scenarios. As illustrated in Fig. 2, the proposed SeeDRec consists of two main components, including the Sememe-to-Interest Diffusion Model (S2IDM) and the Interest-aware Prompt-enhanced (IPE) strategy. Specifically, SeeDRec first proposes S2IDM, which creatively conducts diffusion on the sememe level rather than the item level used in conventional works, with the aim to model users’ local minimal interest units from the global interest diffusion aspect. To adeptly guide the attention of sequential modeling towards the user’s personalized interests, SeeDRec designs an IPE strategy to convert the diffused generalized interest distribution into informative prompts via multiple personalized prompt generators. It enables the accurate extraction of the interest-centred information from each user’s historical behavior sequences, thereby achieving precise user dynamic interest modeling. Furthermore, SeeDRec continues to manifest its superiority in SR and CDSR tasks where the explicit item-sememe hierarchical taxonomy information is even absent (as a more challenging setting with increased noise). Detailed experimental results and analyses can be found in Sec. 4.2 and Sec. 4.4.

3.3 Sememe-to-Interest Diffusion Model

In this section, we describe our Sememe-to-Interest Diffusion Model based on DDPM [Ho et al., 2020], named S2IDM. The overall structure of S2IDM is illustrated in the left part of Fig. 2. In contrast to the direct diffusion employed by DiffRec [Wang et al., 2023] and PDRRec [Ma et al., 2024] at the ID indices, S2IDM creatively deploys DM at the sememe level, concomitantly refining it in accordance with the inter-correlation of sememes, the sememe overlap of item and the multiple interest drift of behavioral sequences.

Forward Process

In the forward process, S2IDM corrupts the original sememe distribution of each user by injecting the Gaussian noises step by step in a discrete manner. To be specific, given the behavioral sequence $S_p^u$ in the index level, S2IDM first extracts the prior sememe behaviors, re-weights it via the frequency and time interval of each sememe to obtain the original interest distribution $s_0^u$ and finally corrupts it in a discrete manner.

Diffusion on sememe With the behavioral sequence $S_p^u = \{i_1, i_2, \ldots, i_p\}$ of user $u$, we first LOOKUP the prior sememe behaviors $s_u = \{s_1^u, s_2^u, \ldots, s_p^u\}$ of user $u$ as the initial input $s_0$ from the item-sememe dependency matrix $T^u \in \mathbb{R}^{S \times q}$, where $s_k^u \in S$ denotes the $k$-th sememe of item $i_p$ and $q = |S|$ denotes the total length of the sememe set $S$. This enables the transfer of DM from the item domain to that of sememe, yielding a threefold advantage: (1) It entails a reduction in the space and time complexity of DM as stipulated by $|S| \ll |I|$, enabling the practical scalability of our sememe-based diffusion model. (2) By virtue of its foundational definition as the minimal unit of interest, the inherent multi-domain sharing characteristic of sememe endows it with transferability in CDSR tasks. (3) It brings in the semantic interrelationship inherent in sememes to furnish comprehensive support to DM.

Frequency-time aware reweighting In light of the pervasive one-to-many association between items and sememes,
Discrete noising Discrete noising is one of the core phases in DM that injects uncertainty into the original interest distribution. The unadorned utilization of the reweighting strategies in TI-DiffRec is untenable, obscuring users’ authentic interests. To this end, we first count the frequency of each sememe in user behaviors to obtain the $s_{0}$, then we conduct the time-interval reweighting strategy in TI-DiffRec to generate the time-interval weights $w^{i}_j = w^{i}_j = \cdots = w^{i}_k = w^{i}_j = w_{\min} + \frac{t - t_{j}}{t_{p} - t_{j}} (w_{\max} - w_{\min})$ for all sememes $s^{i}_1, s^{i}_2, \cdots, s^{i}_k$ of item $i_k$ with the item-level timestamp sequence $T_i = \{t_1, t_2, \cdots, t_p\}$. Here $s^{i}_j$ is the $j$-th sememe of item $i_k$ in $s_{0}$. Ultimately, we assign these weights to each sememe, amalgamating the weight of identical sememes to derive the original interest distribution $s_{0}$.

Reverse Process
The task of the reverse process in DM is to remove the added Gaussian noise step by step to recover the interest distribution $s_{0}$ from the perturbed sememe distribution $s'_{l}$. The related reverse transition $p_{0}(s'_{l-1} | s'_{l})$ is defined as: $p_{0}(s'_{l-1} | s'_{l}) = \mathcal{N}(s'_{l-1}; \mu_{0}(s'_{l}, t); \Sigma_{0}(s'_{l}, t))$, where $\mu_{0}(s'_{l}, t)$ and variance $\Sigma_{0}(s'_{l}, t)$ can be modeled by a deep neural network due to the fact that the relatively small Gaussian noise ensures the transition kernel $p_{0}(s'_{l-1} | s'_{l})$ follows a Gaussian distribution. The iterative process can be formulated as follows: 

$$s'_{l} \xrightarrow{p_{0}(s'_{l-1} | s'_{l})} s'_{l-1} \xrightarrow{p_{0}(s'_{l-2} | s'_{l-1})} \cdots \xrightarrow{p_{0}(s'_{1} | s'_{2})} s'_{0}. \quad (1)$$

Optimization Strategy
The training objective of S2IDM is to force the reverse transition $p_{0}(s'_{l-1} | s'_{l})$ to closely approximate the posterior distribution $q(s'_{l-1} | s'_{l}, s_{0})$, which is achieved by minimizing the Kullback-Leibler (KL) divergence. With the simplification of the generated interest distribution $s'_{0}$ as the inherently noise-enriched $\tilde{s}_{0}$, proceeding directly with the reverse process on it as $s'_{l} \rightarrow s'_{l-1} \rightarrow \cdots \rightarrow s'_{0}$. This not only retains more personalized information to improve the precision of DM but also ensures that the inference process is not initiated from a fully disordered state, leading to a more robust inference process that precisely aligns with the intrinsic nature of recommendation tasks.

3.4 Interest-aware Prompt-enhanced Strategy
It is intuitive that there exists an insurmountable variation between the diffusion-based user interest and the temporal preference, which motivates us to explore how to incorporate the generalized interest distribution into the sequential modeling. In parallel, the interest distribution obtained from S2IDM (a) exhibits correlation with the generalized user interests, so as to provide some attention clues that are somewhat relevant to SR tasks, and (b) enables the multi-domain inter-connectivity at the sememe level, thereby leveraging its semantic relationships to uncover nuanced user preferences and yield enhanced information gain. To this end, we follow the prompt learning paradigm [Lester et al., 2021; Wu et al., 2024] to propose the Interest-aware Prompt-enhanced (IPE) strategy to convert the generalized interest distribution on discrete sememes into multiple continuous prompts. This can smooth the bias between the semantic and behavioral intentions and guide the sequential modeling towards understanding personalized behavioral patterns to enrich the user representation.

Specifically, given the generalized interest distribution $\tilde{s}_{0}$, we first project it into the same feature space via multiple prompt generators $f_{0}(\cdot)$, which is built with a two-layer fully-connected network with the Tanh activation. After obtaining the multi-interest prompts $M = \{m_{1}, m_{2}, \cdots, m_{k}\}$, we place these prompt-enhanced knowledge before the original input matrix $D^{\mathbb{Z}}$ to inject the diffused user generalized interests into the self-attention functions in pre-trained sequential models. Here we designate the length $k$ of the multi-interest prompts as 3 for streamlining. Beyond the rectification of diffusion objective within S2IDM, the lower computational complexity of SeeDRec is also evident in various facets. First is that the generalized interest distribution $\tilde{s}_{0}$ used in IPE only needs to be generated once via the inference process of S2IDM before model training, yielding the asymptotic similar time complexity to its sequence encoder in online serving. Secondly, the implementation of IPE only requires a limited number of additional prompt generators to be trained, which also mitigates the space complexity of the existing DM-based recommenders. Note that the proposed IPE is a universal strategy, devoid of the tailored design in sequence modeling in SR. Thus, IPE can harness the possible
advancement in future sequence modeling, thereby extending the lifecycle of SeeDRec (see Sec. 4.2 and Sec. 4.5).

**Optimization Objective** Following [Kang and McAuley, 2018], we opt the Binary Cross-Entropy Loss $L$ of each user-item pair $(u, i)$ in training set $R$ as the optimization objective:

$$L = - \sum_{(u,i) \in R} [y_{u,i} \log \hat{y}_{u,i} + (1 - y_{u,i}) \log (1 - \hat{y}_{u,i})],$$  \hspace{1cm} (3)

where $\hat{y}_{u,i} = u^T e_i$ is the predicted probability between user representation $u$ and item embedding $e_i$, $y_{u,i} = 1$ and $y_{u,i} = 0$ denote the positive and negative samples respectively.

### 3.5 In-depth Model Discussion

The most related work of SeeDRec in existing DM-based recommenders is PDRec [Ma et al., 2024a], which fully leverages the diffused preference on item indices to improve SR. Since the DM within PDRec is trained on item indices, its time complexity will be limited by the number of item corpus, thereby encountering the scalability challenge. Secondly, due to the ID-based diffusion, PDRec relies on overlapped users to re-train the DM from scratch in CDSR tasks. Finally, the ID-based DM proves challenging in fully harnessing the associative information within the item semantic hierarchy, potentially resulting in sub-optimal performance.

In contrast, SeeDRec proposes $S2IDM$ on the discrete sememes to mine each user’s generalized interest distribution through minimal interest units modeling. Hence, the time complexity on the DM side can be simplified from $|Z| \cdot \mathcal{O}(D)$ to $|S| \cdot \mathcal{O}(D)$, where $\mathcal{O}(D)$ denotes the time complexity of all other facets within DM. According to the consensus of $|S| \ll |Z|$, $S2IDM$ successfully addresses the long-standing scalability challenge that has persistently plagued PDRec. Moreover, SeeDRec exhibits the capability to comprehensively leverage the semantic associative relationships and cross-domain transferability under the sememe hierarchy (see the complexity comparison in Sec. 4.3). This effectively addresses the cross-domain generalization and semantic correlation challenges inherent in PDRec. The experimental results in Sec. 4.2 concurrently affirm that SeeDRec can yield significant and consistent performance improvements upon PDRec. Besides, we also employ the masked discrete DM [Austin et al., 2021] in S2IDM, which yields comparable SR performance. So we just implement S2IDM in the regular manner to simplify the computational complexity.

### 4 Experiments

We conduct extensive experiments on nine SR datasets and four CDSR datasets to answer the following questions:

- **RQ1**: Does the proposed SeeDRec outperform the base SR models and the SOTA DM-based SR methods?
- **RQ2**: How does SeeDRec perform in datasets where explicit item taxonomies (e.g., categories) are absent?
- **RQ3**: How does each component proposed in SeeDRec impact the recommendation performance?
- **RQ4**: Is our SeeDRec effective and universal enough with different base SR models and cross-domain SR tasks?
- **RQ5**: How does SeeDRec function in the interest distribution transfer and the few-shot scenarios?

### 4.1 Experimental Settings

**Datasets** As shown in Table 1, we construct five SR datasets from two platforms (i.e., Amazon [Lin et al., 2022] and PixelRec [Cheng et al., 2023]) with existing categories viewed as sememes. To verify that SeeDRec could work well without existing taxonomies, we also build four SR datasets and four CDSR datasets (with similar sizes) respectively, intentionally overlooking the inherent categories and using words as natural sememes via certain tokenization, lemmatization, and filtering on item titles with the NLTK library.

**Baselines** We implement SeeDRec on two SR models (SASRec [Kang and McAuley, 2018] and CL4SR [Xie et al., 2022]) and a SOTA DM-based model (PDRec [Ma et al., 2024a]) to verify its effectiveness and universality. Besides, we also compare with two DM-based CF models (T-DiffRec [Wang et al., 2023c] and TI-DiffRec [Ma et al., 2024a]).

**Parameter settings** We conduct a comprehensive grid search to select the optimal hyper-parameters. That is, the learning rate is tuned from 0.001 to 0.05. The batch size and the maximum sequence length are defined as 512 and 200 for fair comparisons. It is imperative to underscore that SeeDRec essentially possesses very few parameters (e.g., $k$ in IPE). We just assign $k = 3$ via our empirical knowledge. For $S2IDM$, we define $\omega_{\text{min}} = 0.1$, $\omega_{\text{max}} = 0.5$ and the step $T$ as 10 for all datasets. We use the early-stop strategy to avoid overfitting. Following [Wang et al., 2023c], we randomly sample 999 negative items for each positive instance to speed up the evaluation. All reported results are the average values of five runs with different seeds on the same NVIDIA Tesla V100.

### 4.2 Performance on SR (RQ1 & RQ2)

In this section, we leverage NDCG@10 (N@10), Hit rate@10 (H@10) and AUC as the evaluation metrics. From Table 2, we can observe that: (1) SeeDRec achieves significant improvements on all metrics and datasets compared to its base sequential models, with the significance level $p < 0.05$. It indicates that SeeDRec can precisely mine the user’s interest distribution and successfully incorporate it into SR via the tailored multi-intention prompts on different types of base SR models. (2) Upon a lateral examination across various datasets, SeeDRec is more beneficial on relatively denser datasets. This aligns with its foundational assumption, that is, $S2IDM$ can intrinsically better capture users’ generalized interest distributions from the diffusion on minimal interest units on denser datasets where user behaviors are abundant. (3) SeeDRec attains the peak results across most of the datasets on the basic of PDRec. It confirms that SeeDRec exhibits the different underlying mechanisms of DM utilization with PDRec, positioning it as a stalwart support for the evolution of more advanced DM-based recommenders in the future.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>PixelRec</th>
<th>Home</th>
<th>Electronic</th>
<th>CD</th>
<th>Toy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Users</td>
<td>24,972</td>
<td>22,788</td>
<td>56,727</td>
<td>110,805</td>
<td>116,677</td>
</tr>
<tr>
<td>Items</td>
<td>44,643</td>
<td>23,603</td>
<td>45,279</td>
<td>105,841</td>
<td>77,760</td>
</tr>
<tr>
<td>Sememe</td>
<td>108</td>
<td>1,265</td>
<td>781</td>
<td>476</td>
<td>525</td>
</tr>
<tr>
<td>#Inter.</td>
<td>45,6813</td>
<td>187,778</td>
<td>507,373</td>
<td>1,342,060</td>
<td>1,018,540</td>
</tr>
</tbody>
</table>

Table 1: Statistics of five real-world SR datasets.
Table 3: Computational complexity comparison between TI-DiffRec and S2IDM on CD and Pixel. “Par.,” “Tra.” and “Eval.” denote parameters, training time and evaluation time for one epoch.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Algorithm</th>
<th>#Par. (M)</th>
<th>GPU (MB)</th>
<th>#Tra. (s)</th>
<th>#Eval. (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD</td>
<td>TI-DiffRec</td>
<td>201.99</td>
<td>6052</td>
<td>409.11</td>
<td>2193.92</td>
</tr>
<tr>
<td></td>
<td>S2IDM</td>
<td>0.92</td>
<td>1408</td>
<td>364.9</td>
<td>827.8</td>
</tr>
<tr>
<td>Pixel</td>
<td>TI-DiffRec</td>
<td>85.2</td>
<td>3252</td>
<td>38.79</td>
<td>291.79</td>
</tr>
<tr>
<td></td>
<td>S2IDM</td>
<td>0.21</td>
<td>1270</td>
<td>35.94</td>
<td>178.25</td>
</tr>
</tbody>
</table>

Figure 4: Ablation results of SeeDRec on SASRec (solid) and CLASRec (slash) on five SR datasets (using word as sememe).
Table 4: Results of SeeDRec and its ablation versions on four CDSR datasets. All improvements over the backbone are significant.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Pixel</th>
<th>Home</th>
<th>Electronic</th>
<th>Metric</th>
<th>N@10</th>
<th>H@10</th>
<th>N@10</th>
<th>H@10</th>
<th>N@10</th>
<th>H@10</th>
<th>N@10</th>
<th>H@10</th>
<th>N@10</th>
<th>H@10</th>
<th>N@10</th>
<th>H@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>SASRec</td>
<td>0.0857 0.1583</td>
<td>0.0962 0.1267</td>
<td>0.1172 0.1865</td>
<td>21.0%</td>
<td>19.0%</td>
<td>7.7%</td>
<td>9.5%</td>
<td>14.5%</td>
<td>14.2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SeeDRec</td>
<td>0.1037 0.1883</td>
<td>0.1036 0.1387</td>
<td>0.1342 0.2130</td>
<td>0.1600 0.2109</td>
<td>0.1606 0.2109</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Few Shot</th>
<th>Metric</th>
<th>N@10</th>
<th>H@10</th>
<th>N@10</th>
<th>H@10</th>
<th>N@10</th>
<th>H@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>SASRec</td>
<td>0.0894 0.1618</td>
<td>0.1714 0.2072</td>
<td>0.1579 0.2269</td>
<td>41.0%</td>
<td>30.8%</td>
<td>10.9%</td>
<td>11.1%</td>
</tr>
<tr>
<td>SeeDRec</td>
<td>0.1261 0.2115</td>
<td>0.1900 0.2301</td>
<td>0.1835 0.2624</td>
<td>0.2727 0.3431</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Results of SeeDRec on full and few-shot settings.

4.5 Universality Analyses (RQ4)

We conduct SeeDRec on different classical SR models in five SR datasets to verify its universality. From Table 2 and Fig. 4, we have: (1) SeeDRec is a model-agnostic method that simultaneously exhibits its superiority on SASRec, CL4SRec, and even diffusion-based PDRRec. Meanwhile, each component in SeeDRec achieves its incremental improvement on most datasets with different base models. (2) CL4SRec and PDRRec outperform SASRec by advanced techniques such as contrastive learning and diffusion model, which is also reflected when armed with SeeDRec. It implies the compatibility between the sememe-based diffusion of SeeDRec and advanced sequential modeling techniques. Consequently, SeeDRec is likely to retain its universality and effectiveness in cooperating with future sophisticated sequential models.

4.6 In-depth Model Analyses (RQ5)

SeeDRec in Few-shot Setting  The diffusion-based generalization exhibited by S2IDM serves to facilitate the exploration of associated interests for users with fewer behaviors or discovered interests. It motivates us to investigate SeeDRec’s potential in few-shot scenarios. Hence, we set a sememe threshold for each dataset to simulate users that only have very few interests discovered by the real RS application and regard the few-shot performance in Table 5.

We observe that: (1) Comparing with the full evaluation, SeeDRec achieves more significant improvements over SASRec on the few-shot setting. It verifies the feasibility of SeeDRec as a few-shot recommender thanks to the diffused interests. (2) Comparing the relative improvements across diverse datasets, it becomes evident that SeeDRec exhibits superior few-shot performance on Pixel, where the average number of sememes per item is comparatively lower. This may stem from the fact that SeeDRec elegantly addresses the twofold challenges of behavioral sparsity and preference unitary in this particular setting, thereby implying some ingenious usages of SeeDRec for new users (e.g., asking new users to pick several “seed” interests and generalize them via SeeDRec).

Table 5: Results of SeeDRec on full and few-shot settings.

5 Conclusion and Future Work

This paper proposes a Sememe-based Diffusion framework (SeeDRec), which captures each user’s sememe-based diffusion-generalized interest distribution to enhance SR. With the proposed recommendation sememe powered by S2IDM and IPE, SeeDRec is verified to be effective, transferable, and scalable on thirteen SR and CDSR datasets. The proposed SeeDRec could also be easily adopted with different types of sequential models without much additional inference computation costs, which will be welcomed by the industry.

Future investigations should encompass the design of more cogent minimal interest units and the exploration of the semantic correlations among sememes during the diffusion process. Furthermore, we believe that SeeDRec indicates the future direction for the subsequent DM-based recommenders beyond the diffusion on indices. The proposed recommendation sememe and S2IDM can facilitate seamless integration into diverse scenarios to bring consistent improvements.
Acknowledgments
This work is supported in part by the National Key R&D Program of China (Grant no. 2021YFC3300203), the TaiShan Scholars Program (Grant no. tsqn202211289), the Shandong Province Excellent Young Scientists Fund Program (Overseas) (Grant no. 2022HWYQ-048), the Oversea Innovation Team Project of the "20 Regulations for New Universities" funding program of Jinan (Grant no. 2021GXRC073) and the Young Elite Scientists Sponsorship Program by CAST (2023QNRCC001). ChatGPT and Grammarly are utilized to improve grammar and correct spelling. Corresponding Authors: Lei Meng and Ruobing Xie.

References


[Sun et al., 2024] Weilin Sun, Manyi Li, Peng Li, Xiao Cao, Xiangxu Meng, and Lei Meng. Sequential selection and calibration of video frames for 3d outdoor scene reconstruction. CAAI Transactions on Intelligence Technology, pages 1–15, 2024.


[Wu et al., 2024] Yiqing Wu, Ruobing Xie, Yongchun Zhu, Fuzhen Zhuang, Xu Zhang, Leyu Lin, and Qing He. Personalized prompt for sequential recommendation. IEEE Transactions on Knowledge and Data Engineering, 2024.

