

基于引导扩散的序列推荐方法

李博 莫先

宁夏大学信息工程学院 银川 750021

(wjswlb21@163.com)

摘要 随着用户行为偏好的动态变化,传统序列推荐方法面临着难以捕捉用户意图转变的挑战。为了解决这一问题,提出了一种基于引导扩散的序列推荐方法(GDRec),旨在通过将目标项目表示嵌入到扩散模型中,实现对用户当前意图的精准捕捉。具体地,GDRec模型包括以下关键组件:序列编码器、交叉注意力条件去噪解码器和交叉散度目标。序列编码器逐步生成用户偏好表示,捕捉历史序列与当前目标的动态关系;交叉注意力条件去噪解码器去除嵌入表示中的噪声,提高对下一目标项目的预测精度;交叉散度目标则赋予模型排序能力,确保表示的高质量,并在扩散过程中嵌入目标项目表示进行引导。最后,在Amazon的Office和Tools数据集上进行的大量实验证明了GDRec在多个评价指标上均优于现有的先进方法,显示出其在序列推荐任务中的优越性能。此外,消融实验和超参数分析进一步验证了模型的有效性和稳定性。

关键词: 序列推荐;引导扩散;信息嵌入;用户意图

中图分类号 TP393.1

Guided Diffusion Sequence Recommendation Methods

LI Bo and MO Xian

College of Information Engineering, Ningxia University, Yinchuan 750021, China

Abstract With the dynamic change of user behavior preference, traditional sequential recommendation methods face the challenge of difficult to capture the change of user intention. In order to solve this problem, this study proposes guided diffusion sequence recommendation method(GDRec), which aims to achieve accurate capture of the user's current intention by embedding the target item representation into the diffusion model. Specifically, the GDRec model includes the following key components: a sequence encoder, a cross-attention conditional denoising decoder, and a cross-divergence objective. The sequence encoder gradually generates the user preference representation to capture the dynamic relationship between the historical sequence and the current target. The cross-attention conditional denoising decoder removes the noise in the embedded representation and improves the prediction accuracy of the next target item. The cross-divergence objective, on the other hand, empowers the model with ranking capabilities, ensuring high quality representations, and embedding target item representations to guide the diffusion process. Finally, a large number of experiments on Amazon Office and Tools datasets prove that GDRec is superior to the existing advanced methods in multiple evaluation indicators, showing its superior performance in sequence recommendation tasks. In addition, ablation experiments and hyperparameter analysis further verify the effectiveness and stability of the model.

Keywords Sequential recommendation, Guided diffusion, Information embedding, User intent

1 引言

序列推荐(Sequential Recommendation, SR)因特别关注用户与项目之间交互的时间性,能够依据用户的历史交互记录有效捕捉用户潜在意图和兴趣变化,而成为广泛研究的热点^[1-4],并在新闻、短视频、购物和电商等领域展现出有效性。

随着深度学习的快速发展,基于深度神经网络的序列推荐研究^[5-6]因其强大的扩展性和出色的性能提升效果而备受关注。Bi-LSTM^[7]是一种双向长短期记忆网络,能够同时捕捉序列数据的前后上下文信息,提升文本分类和序列标注等任务的性能,具有较强的序列特征提取能力。PL-GPR^[8]旨

在为用户提供优化后的学习内容和路径推荐。该模型利用长短期记忆网络的结构,有效捕捉用户学习行为中的时间依赖性和上下文信息。处理过程中,模型通过分析用户的历史学习交互数据,动态生成个性化的推荐,以适应用户的学习需求和偏好。

此外,注意力机制因其有效性及与多种模型的融合能力,被广泛应用于序列推荐任务中。CSAN^[9]引入了自注意力机制,以便动态关注用户行为中的重要特征,并结合用户历史数据生成上下文嵌入。处理过程中,模型通过对序列中的各个位置应用自注意力,增强了信息的相关性与个性化,从而有效提升了用户体验和推荐效果。PACA^[10]引入了位置感知

基金项目:国家自然科学基金(62306157);宁夏自然科学基金(2024AAC05011)

This work was supported by the National Natural Science Foundation of China (62306157) and Natural Science Foundation of Ningxia (2024AAC05011).

通信作者:莫先(mxian168@nxu.edu.cn)

机制,以便在推荐过程中考虑用户行为的时间顺序信息,从而增强上下文理解。处理时,模型为每个位置生成一个嵌入,并将位置嵌入与会话嵌入相乘,构建出会话的总特征。

近期,许多研究将对比学习引入序列推荐任务,旨在通过分析不同数据之间的相似性和差异性来提升学习效果。DuoRec^[11]引入了对比学习机制,以便在训练过程中识别和强化有效的表示,同时抑制无关特征的干扰。处理方法上,模型通过构建正负样本对,利用对比损失函数来优化用户行为的表示,确保学习到的特征能够更好地反映用户的真实偏好。CLASrec^[12]利用对比学习机制,通过分析用户行为序列中的相似性和差异性来增强特征表示。该方法在结构上结合了对比损失函数,使得正样本和负样本之间的距离得以优化。处理过程中,模型通过生成不同的用户行为视图,强化有效信息的提取,从而学习到更具辨别性的表示。

尽管上述研究在序列推荐任务中取得了不错效果,但现有方法仍存在许多问题。首先,研究表明^[13],随着网络层数的增加,模型性能反而下降。这通常是由于特征表示在嵌入空间中趋于一致,导致维度塌陷,项目表示退化^[14],降低了推荐效果。其次,文献^[15]发现了一种称为“排名高原”的问题,即尽管模型生成的项目表示质量很高,但顶端项目的排名分数差异很小,难以区分。例如,DouRec模型的最佳项目表示得分仅比第二名高出1%。这种现象不仅可能归因于项目表示的退化,还涉及用户与项目交互之外的其他因素。此外,从用户的历史交互数据可以看出,部分用户的意图会发生显著变化,而序列推荐通常依据过去的交互行为加权处理,不注重候选项目表示的质量变化,致使模型无法及时捕捉用户动态意图变化,降低了推荐效果。因此,如何设计出一个项目表示质量高、目标项目区分能力强、动态意图捕捉准的序列推荐方法,是当前研究的一个重要挑战。

为应对这一挑战,本文提出了一种基于引导扩散的序列推荐方法(Guided Diffusion Sequence Recommendation Methods, GDRec)。该模型包含多个关键模块,包括序列编码器、交叉注意力条件去噪解码器以及交叉散度目标。在扩散过程,模型会对目标项目的表示进行建模,捕捉项目序列的潜在表征分布,并将目标商品的嵌入表示与高斯噪声融合,生成被污染的目标表示,引导扩散过程。在反向过程,交叉注意力条件去噪解码器对目标项目进行重构。在历史交互记录和目标项目的引导下,序列解码器可以逐步除去历史交互序列表示的噪声,并生成基于用户当前意图引导的参与表示。模型中添加用于引导扩散的目标项目可以促进高质量的目标表示生成,而模型扩散和反向过程中的逐步点积运算可以及时且正确地捕捉用户动态意图变化,从而提高模型推荐效果。本文的主要贡献总结如下:

- 1) 创新性地提出基于引导扩散的序列推荐方法 GDRec,并通过引入去噪扩散概率模型到序列推荐任务,在扩散过程中添加项目目标引导。与现有模型相比,GDRec 能够更加精准地捕捉用户兴趣的变化,并提供更强的可解释性。
- 2) 通过在两个真实世界的数据集上进行广泛的实验,验证了 GDRec 在序列推荐任务上与当前先进的模型竞争时的优越性。各项指标上的显著提升充分证明了其有效性。

2 相关工作

2.1 序列推荐

将用户的历史交互记录建模为马尔可夫决策过程(Markov Decision Process, MDP)使得马尔可夫链在序列推荐中发挥了重要作用^[16-18]。然而,由于该方法假设下一个项目的表示仅依赖于前一个项目的表示,导致其表达能力受限。随着深度神经网络及其多种变体的出现,这一局限性逐渐被突破。例如,Caser^[19]是一种基于卷积序列嵌入的个性化前N项序列推荐模型,模型通过卷积神经网络提取用户行为序列中的特征,将用户历史行为转化为序列嵌入,并通过卷积操作生成有效的特征表示,生成下一个项目的推荐。在此基础上,BERT4Rec是一种基于Transformer的序列推荐模型,旨在通过双向编码器表示提升推荐的准确性和个性化。该模型提出了基于自注意力机制的序列推荐方法,认为未来信息对于当前项目的预测同样重要,并通过双向自注意力机制将未来信息引入模型,使得模型能够同时捕捉用户行为序列中的前后上下文信息,取得了优异的性能。此外,图神经网络(Graph Neural Networks, GNN)能够有效地将高阶关系融入序列建模中,因此在序列推荐领域也得到了广泛应用^[20]。

2.2 去噪扩散概率模型

文献^[21]提出了去噪扩散概率模型(Denoising Diffusion Probabilistic Model, DDPM),并在连续空间任务中取得了显著成功,例如图像生成等应用。最近,一些研究尝试将DDPM扩展到离散任务领域,例如文本生成和序列推荐。在文本生成方面,DiffSeq^[22]是一种基于扩散模型的序列到序列文本生成模型,旨在提高文本生成的质量和多样性。在前向过程中部分污染答案,将用户的问题视为未被噪声污染的部分,而答案则作为被污染的部分,随后通过非自回归的反向传递重构答案,实现了预期的文本生成。在序列推荐任务中,文献^[23]通过构建扩散过程来表示用户行为序列,从而生成更加丰富的特征表示。处理过程中,模型将用户的历史行为视为图结构,利用图卷积网络进行特征学习,增强了对用户偏好的理解,从而实现推荐任务。

与上述方法不同,GDRec在目标项目表示上添加高斯噪声,通过增加损坏的目标项目表示和历史交互记录,来引导扩散过程。此外,GDRec采用自回归策略实现条件生成,能够在较少的步骤中生成高保真度的项目表示,从而提升模型在序列推荐任务中的表现。

3 问题定义

序列推荐是一种智能化推荐方法,旨在通过分析用户的历史行为模式来预测并推荐用户可能感兴趣的下一个项目。在这个过程中,系统会考虑用户与项目之间的历史交互序列,以此来捕捉用户的兴趣演变和偏好。

在序列推荐系统中,用户的交互序列被定义为 $S_x = \{x^1, x^2, \dots, x^n\}$,在训练过程中, x^n 代表目标项目。 e^n 是项目 x^n 的嵌入表示, e_t^n 是 x^n 及其之前项目的历史交互序列表示。扩散过程中的噪声序列被定义为 $S_T = \{x_0^n, x_1^n, \dots, x_T^n\}$, x_t^n 是目标项目 x^n 添加了 t 步噪声后的项目表示, T 是最大扩散步骤。反向过程中的还原序列被定义为 $\hat{S}_T = \{\hat{x}_T^n, \hat{x}_{T-1}^n, \dots, \hat{x}_0^n\}$, \hat{x}_t^n 是消除第 t 步之后噪声的项目表示, \hat{x}_0^n 是消除所有

噪声的项目表示,即预测的目标项目。

4 本文方法

本章将详细描述 GDRec 模型,如图 1 所示。GDRec 模型由 4 个核心组成。1)带引导的扩散过程:在扩散过程中,嵌入被高斯噪声干扰的目标项目表示以引导扩散。2)序列

编码器:学习历史序列和当前目标的嵌入表示,并逐步生成用户偏好表示,引导模型捕捉用户兴趣的动态变化。3)交叉注意力条件去噪解码器:逐步去除条件序列中的高斯噪声,生成过渡性的目标偏好表示,从而提高模型对下一目标项目的预测精度。4)交叉散度目标:赋予模型排序能力,保持高质量的表示,避免目标项目表示退化。

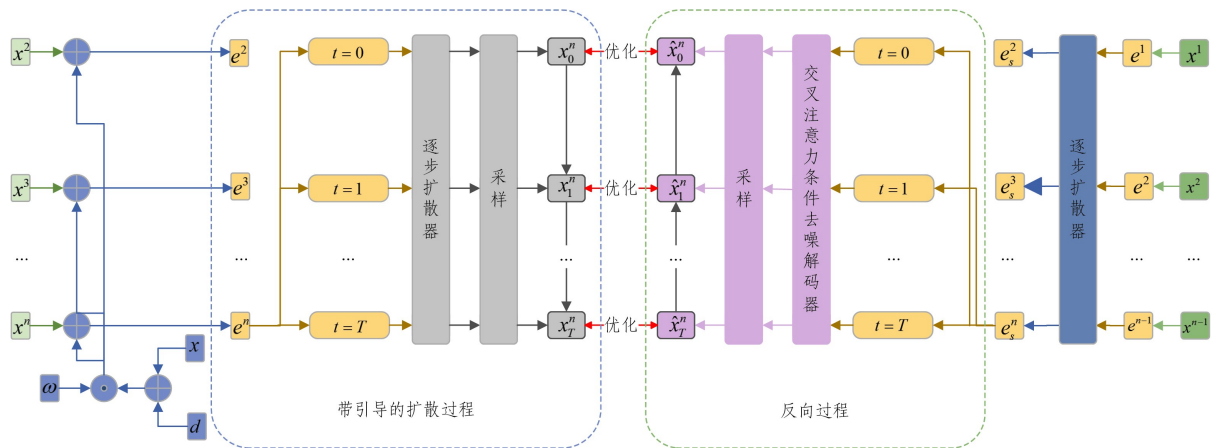


图 1 GDRec 模型框架

Fig. 1 Framework of GDRec model

4.1 带引导的扩散过程

扩散过程中向目标项目引入高斯噪声,并在每一步中创建被损坏的目标表示,由此促进去噪解码器的学习。在给定的扩散步骤 t 中添加预定义的噪声尺度 β_t 和相应的扩散转移分布 $q(x_t^n | x_{t-1}^n) \sim N(x_t^n; \sqrt{1-b_t}x_{t-1}^n, b_t\mathbf{I})$, 当第一个扩散步长被用作条件时,当前扩散步长的分布可以表示为:

$$q(x_t^n | x_0^n) = N(x_t^n; \sqrt{\bar{\alpha}_t}x_0^n, (1-\bar{\alpha}_t)\mathbf{I}) \quad (1)$$

$$\alpha_t = 1 - \beta_t, \bar{\alpha}_t = \prod_{i=1}^t \alpha_i$$

其中 $x_0^n = e^n$ 是扩散过程初始化的项目表示。因此通过 x_0^n 可以对任意扩散步骤 t 的损坏目标 x_t^n 进行采样:

$$x_t^n = \sqrt{\bar{\alpha}_t}x_0^n + \sqrt{1-\bar{\alpha}_t}\epsilon, \epsilon \sim N(0, \mathbf{I}) \quad (2)$$

为了在扩散过程有着更好的项目表示,本文将嵌入噪声目标和目标项目表示为以下形式:

$$e^i = x^i + \omega \otimes (x + d) \quad (3)$$

其中, x^i 是第 i 个历史项目表示; ω 是可学习的超参数,在扩散过程, ω 控制着噪声注入的引导程度; 符号 \otimes 是逐元素乘积; d 是依照文献[24]创建的相应扩散步骤的嵌入; x 表示噪声项目嵌入。

4.2 序列编码器

在序列推荐任务中,相比于以往采用非自回归生成的方法从随机初始化的高斯噪声中去噪, GDREC 采用条件自回归的方式进行下一项目表示的生成,并使用 SASRec 作为序列编码器,以捕获历史交互序列的隐藏表示。

4.3 交叉注意力条件编码器

根据 4.1 节描述的扩散步骤分布 $q(x_t^n | x_0^n)$, $q(x_{t-1}^n | x_0^n)$ 和 $q(x_t^n | x_{t-1}^n)$, 反向去噪分布可以使用贝叶斯法则计算后验概率分布的解析形式, 即 $q(x_{t-1}^n | x_t^n, x_0^n)$ 。可以将反向去噪步骤近似为分布 $p_\theta(\hat{x}_{t-1}^n | \hat{x}_t^n) \sim N(\mu_\theta(\hat{x}_t^n, t), \beta_t \mathbf{I})$, 其中 $\hat{\beta}_t = [(1-\hat{\alpha}_{t-1})/(1-\hat{\alpha}_t)]\beta_t$ 是方差的封闭形式。在步骤 $t-1$ 以上一步去噪表示 \hat{x}_t^n 为条件, 学习去噪表示 \hat{x}_{t-1}^n 。为了能基于

历史交互记录自回归地预测后续项目, 本文将去噪器集成到条件生成框架中, 形成条件去噪解码器, 同时将反向去噪步骤与先前序列进行条件化表示, 表达式为: $p_\theta(\hat{x}_t^n | e_s, t) \sim N(\mu_\theta(e_s, t), \beta_t \mathbf{I})$, 其中 e_s 表示通过序列编码器生成的历史交互, $\mu_\theta(e_s, t)$ 表示去噪后的均值。考虑到序列推荐任务的稀疏性, 模型的去噪解码器采用交叉注意力架构, 在尽可能保留更多信息的同时, 还可以将整个序列表示和对应的步长作为输入。具体来说, 对于给定的序列嵌入 $e_s^{1:n}$ 和对应的去噪步骤 t , 条件去噪解码器可以预测扩散过程中被损坏的目标项目的去噪均值。依据步骤 t 从嵌入记录中查询可学习嵌入 e_t , 并扩展到 $R^{(n-1) \times d}$ 维度, 保证每个隐藏的嵌入都可以感知到一致的去噪步骤。交叉注意力 (Cross-Attention, CA) 的定义如下:

$$m_q^{1:n}(e_s^{1:n}, t) = CA(e_s^{1:n}, e_t) = \text{Softmax}\left(\frac{(e_t W^Q)(e_s^{1:n} W^K)^T}{\sqrt{d}}\right)(e_s^{1:n} W^V) \quad (4)$$

给定预计算的后延方差和预测的去噪均值, 可以对步骤 t 的用户偏好进行采样:

$$\hat{x}_t^n = \mu_\theta^{\hat{x}} + \hat{\beta}_{t+1} \epsilon, \epsilon \sim N(0, \mathbf{I}) \quad (5)$$

4.4 优化

传统扩散概率模型目标是学习对应步骤上的去噪函数 $p_\theta(x_{t-1} | x_t)$, 并最小化 KL 散度 $D_{\text{KL}}[q(x_{t-1} | x_t, x_0) \| p_\theta(x_{t-1} | x_t)]$, 以最大化预测的项目嵌入与被损坏的输入数据之间的相似性。然而, 序列推荐任务不仅要求对项目进行有效排序, 还要求目标项目得分高于非兴趣项目。因此, 仅重建输入序列无法满足这些要求。此外, 因为序列推荐任务中的项目嵌入都是随机初始化并在执行过程中动态优化的, 所以模型可能会学习到琐碎的项目表示, 导致项目嵌入高度相似且得分差距微小。为了解决这些问题, 本文要求预测项目嵌入与目标项目嵌入之间的 KL 散度小于预测项目嵌入和随机负采样项目嵌入之间的 KL 散度, 并在每个去噪步骤 t 上引

入使用 KL 散度衡量非相似性的交叉散度损失函数：

$$L_{\text{cd}}^t = \frac{1}{N} \sum_n \log(\sigma(-D_{\text{KL}}[q(x_i^n | x_{i+1}^n, x_0^n) \| p_\theta(\hat{x}_i^n | e_s, t)])) + \log(1 - \sigma(-D_{\text{KL}}[q(x_i^n | x_{i+1}^n, x_0^n) \| p_\theta(\hat{x}_i^n | e_s, t)])) \quad (6)$$

其中, x_0^n 是用户交互中从未出现的随机负采样项目的嵌入, 根据式(2)和式(4), 本文对被高斯噪声损坏的目标项目嵌入 x_i^n 和生成的用户偏好目标 \hat{x}_i^n 进行了抽样。

为了防止生成的项目表示崩塌, 本文依据 CDDRec 采用了一种基于 InfoNCE 损失^[25] 的内外视角对比损失函数。内视角的 InfoNCE 在最小化用户偏好和目标项目嵌入之间距离的同时, 还扩大了不同用户或项目之间的距离：

$$L_{\text{in}}^t = \frac{1}{N} \sum_{i=1}^N \log \frac{\exp(\hat{x}_i^{tT} x_i^t / \tau)}{\sum_j \exp(\hat{x}_i^{tT} x_j^t / \tau) + \sum_j \exp(\hat{x}_i^{tT} \hat{x}_j^t / \tau)} \quad (7)$$

其中, \hat{x}_i^t 是序列位置 i 在扩散步骤 t 中条件去噪解码器的输出。

交叉视角的 InfoNCE 确保了序列编码器生成合理的序列表示。通过最小化具有轻微噪声插值的相同输入序列之间的距离, InfoNCE 给出了交叉视角的函数表示：

$$L_{\text{cross}}^t = \frac{1}{N} \sum_{i=1}^N \log \frac{\exp(\hat{x}_i^{tT} \tilde{x}_i^t / \tau)}{\sum_j \exp(\hat{x}_i^{tT} \tilde{x}_j^t / \tau) + \sum_j \exp(\hat{x}_i^{tT} \hat{x}_j^t / \tau)} \quad (8)$$

其中, \hat{x}_i^t 是序列位置 i 在扩散步骤 t 中条件去噪解码器的输出。

随着扩散过程的进行, 在更高扩散步骤中丢失的信息会更多。为了避免过多关注添加的噪声, 通过为每个扩散步骤的损失添加比例系数来控制其占比。最终的优化目标表示为：

$$L = \sum_{t=0}^T \frac{1}{t+1} (L_{\text{cd}}^t + \lambda(L_{\text{in}}^t + L_{\text{cross}}^t)) \quad (9)$$

表 2 整体结果比较

Table 2 Comparison of overall results

Datasets	Metric	SVAE	ConstrastVAE	CL4Rec	DouRec	GRU4Rec	SASRec	CDDRec	GDRec
Office	R@1	0.0088	0.0194	0.0094	0.0120	0.0051	0.0198	<u>0.02710</u>	0.0296
	R@5	0.0316	0.0642	0.0294	0.0330	0.0241	0.0656	<u>0.07651</u>	0.0805
	R@10	0.0597	0.1052	0.0430	0.0559	0.0510	0.0989	<u>0.10911</u>	0.1164
	N@5	0.0202	0.0411	0.0194	0.0223	0.0149	0.0428	<u>0.05211</u>	0.0557
	N@10	0.0292	0.0544	0.0237	0.0296	0.0234	0.0534	<u>0.06271</u>	0.0672
	MRR	0.0249	0.0463	0.0207	0.0264	0.0204	0.0457	<u>0.05481</u>	0.0585
Tools	R@1	0.0055	0.0090	0.0060	0.0058	0.0047	0.0103	<u>0.01271</u>	0.0141
	R@5	0.0118	0.0242	0.0189	0.0182	0.0154	0.0284	<u>0.03591</u>	0.0378
	R@10	0.0204	0.0364	0.0293	0.0361	0.0242	0.0427	<u>0.05221</u>	0.0545
	N@5	0.0086	0.0166	0.0123	0.0120	0.0102	0.0194	<u>0.02441</u>	0.0261
	N@10	0.0114	0.0206	0.0156	0.0148	0.0129	0.0240	<u>0.02971</u>	0.0311
	MRR	0.0098	0.0178	0.0132	0.0128	0.0113	0.0207	<u>0.02531</u>	0.0267

注: 加粗表示最佳结果, 下划线表示次优结果。

5.3 结果分析

在序列推荐任务中, Recall, NDCG 和 MRR 是 3 类常用的模型性能评估指标。Recall 主要反映模型的覆盖能力, 值越高表示模型能够检索出更多相关项目。NDCG 则用于评估推荐系统和信息检索任务中的排序质量。MRR 用于评估模型对前几项结果的排序精度。

GDRec 模型在 Office 和 Tools 两个数据集上的表现均优于其他模型, 充分展示了其优越性。在 Recall@1 方面, GDRec 相比第二名的模型在两个数据集上的效果分别提升

5 实验

5.1 数据集

实验选用了两个 Amazon 数据集^[26] Office 和 Tools, 作为研究的数据源。对于每位用户, 实验按照交互时间对数据进行排序, 将倒数第二次和最后一次交互记录分别作为验证集和测试集, 其余的交互记录用于训练模型。数据集的具体信息如表 1 所列。

表 1 数据及信息统计

Table 1 Data and information statistics

数据集	# 用户数	# 物品数	# 交互数	# 每个物品 平均交互次数	平均序列 长度
Office	4905	2420	53258	22.00	10.8
Tools	16638	10217	134476	13.16	8.1

5.2 实验对比

实验将 GDRec 与最新的基线算法进行对比, 以验证 GDRec 的优越性, 表 2 展示了不同方法在数据集上的表现。

SVAE^[27]: 首次将变分自编码器引入到顺序推荐任务。

ContrastVAE: 引入名为 ContrastELBO 的目, 以最大化潜在变量之间的互信息。

CL4Rec: 引入数据增强策略, 包括掩码、打乱和剪裁, 并通过 InfoNCE 损失优化模型。

DouRec: 通过考虑与正视图具有相同目标项的序列语义增强来提高性能。

GRU4Rec^[28]: 首次尝试将循环神经网络用于序列推荐任务。

SASRec: 首次将基于 Transformer 的编码器引入到序列推荐任务。

CDDRec: 首次在条件自回归生成范式中条件去噪扩散模型应用到序列推荐任务中。

了 9.2% 和 11.0%。此外, 在 MRR 指标上, GDRec 相比第二名的模型在两个数据集上的效果分别提升了 6.3% 和 5.5%。这些结果归因于模型在扩散过程中的有效引导机制, 从而显著提升了推荐精度。

5.4 消融实验

为了验证 GDRec 模型在扩散过程中嵌入目标项目表示对扩散机制的引导作用, 本文设计了对比实验, 将包含目标项目表示的 GDRec 与去除目标项目表示嵌入的变体 DRec 进行性能评估, 在 Office 和 Tools 数据集上的实验结果如图 2 所示。

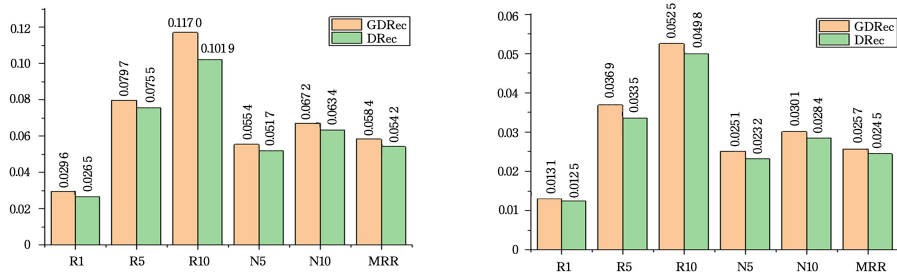


图 2 消融实验对比

Fig. 2 Comparison of ablation experiments

在推荐列表的首项推荐性能指标(即 $R@1$)上,GDRec 相较于 DRec,在 Office 数据集上推荐项目被用户感兴趣的比例提升了 11.69%,在 Tools 数据集上提升了 4.8%。这一结果表明,目标项目表示的嵌入对最优推荐项目的准确性具有显著促进作用。

进一步分析整体推荐性能,在推荐列表的前 10 个项目中至少包含一个用户感兴趣项目的比例(即 $R@10$)上,GDRec 相较于 DRec,在 Office 数据集和 Tools 数据集上的提升分别达到了 14.8%和 5.4%。这一结果表明,嵌入目标项目表示的扩散机制能够显著提高推荐列表整体的覆盖率和准确性。

实验结果表明,GDRec 通过目标项目表示的嵌入,在扩散过程中能够有效提高推荐模型的性能,尤其是在单一最优推荐和整体推荐准确性两方面均具有显著的优势。

5.5 参数分析

在本文的超参数调优中,重点讨论了最大扩散步数、最大噪声水平及目标项目嵌入系数对模型性能的影响。如图 3 所示,在 Office 和 Tools 数据集上,最优的最大扩散步长分别是 25 和 20,最佳最大噪声水平分别是 0.04 和 0.06,最佳的目标项目嵌入系数分别是 0.09 和 0.07。造成这种系数不同的原因可能是较长的序列本身就包含嘈杂的交互记录,需要更少的噪声。

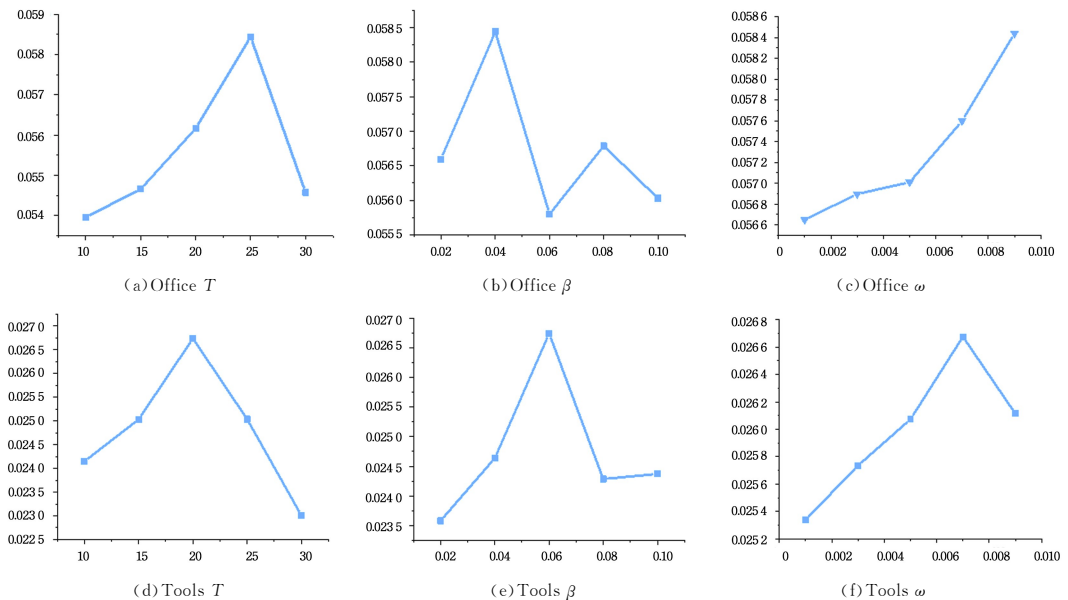


图 3 超参数分析结果

Fig. 3 Results of hyperparameter analysis

结束语 本文提出了一种基于引导扩散的序列推荐方法 GDRec,其核心思想是将目标项目表示嵌入到扩散模型中,从而在扩散过程中有效引导用户的当前意图,实现在序列推荐中的更高精度。实验结果表明,GDRec 在序列推荐任务中的表现优于现有扩散模型,并具有更好的可解释性。未来研究将继续探索扩散模型在序列推荐中的应用,并尝试融入对比学习方法,以进一步提升模型对用户行为模式的捕捉能力和序列学习的效果。

参考文献

[1] LIU Z,FAN Z,WANG Y,et al. Augmenting sequential recommendation with pseudo-prior items via reversely pre-training

transformer [C]// Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2021.

[2] WANG Y,LIU Z,ZHANG J,et al. DRDT:Dynamic reflection with divergent thinking for llm-based sequential recommendation [J]. arXiv:2312.1136,2023.

[3] WANG Y,ZHANG H,LIU Z,et al. Contrastvae: Contrastive variational autoencoder for sequential recommendation [C] // Proceedings of the 31st ACM International Conference on Information & Knowledge Management. 2022.

[4] XIE Z,LIU C,ZHANG Y,et al. Adversarial and contrastive variational autoencoder for sequential recommendation [C] // Proceedings of the Web Conference 2021. 2021.

- [5] KANG W C, MCAULEY J. Self-attentive sequential recommendation[C]// Proceedings of the 2018 IEEE International Conference on Data Mining(ICDM). IEEE, 2018.
- [6] SUN F, LIU J, WU J, et al. BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer [C]// Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019.
- [7] JOZEFOWICZ R, ZAREMBA W, SUTSKEVER I. An empirical exploration of recurrent network architectures [C] // Proceedings of the International Conference on Machine Learning. PMLR, 2015.
- [8] ZHOU Y, HUANG C, HU Q, et al. Personalized learning full-path recommendation model based on LSTM neural networks [J]. Information Sciences, 2018, 444: 135-152.
- [9] HUANG X, QIAN S, FANG Q, et al. Csan: Contextual self-attention network for user sequential recommendation[C]// Proceedings of the 26th ACM International Conference on Multimedia. 2018.
- [10] CAO Y, ZHANG W, SONG B, et al. Position-aware context attention for session-based recommendation [J]. Neurocomputing, 2020, 376: 65-72.
- [11] QIU R, HUANG Z, YIN H, et al. Contrastive learning for representation degeneration problem in sequential recommendation [C]// Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining. 2022.
- [12] XIE X, SUN F, LIU Z, et al. Contrastive learning for sequential recommendation[C]// Proceedings of the 2022 IEEE 38th International Conference on Data Engineering(ICDE). IEEE, 2022.
- [13] GUO X, WANG Y, DU T, et al. Contranorm: A contrastive learning perspective on oversmoothing and beyond [J]. arXiv: 2302. 06562, 2023.
- [14] ZHOU K, WANG H, ZHAO W X, et al. S3-rec: Self-supervised learning for sequential recommendation with mutual information maximization[C]// Proceedings of the 29th ACM International Conference on Information & Knowledge Management. 2020.
- [15] WANG Y, LIU Z, YANG L, et al. Conditional denoising diffusion for sequential recommendation[C]// Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, 2024.
- [16] RENDLE S, FREUDENTHALER C, SCHMIDT-THIEME L. Factorizing personalized markov chains for next-basket recommendation[C]// Proceedings of the 19th International Conference on World Wide Web, 2010.
- [17] SHANI G, HECKERMAN D, BRAFMAN R I, et al. An MDP-based recommender system [J]. The Journal of Machine Learning Research, 2005, 6(9): 1265-1295.
- [18] ZIMDARS A, CHICKERING D M, MEEK C J A P A. Using temporal data for making recommendations [C]// Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence. 2013: 580-588.
- [19] TANG J, WANG K. Personalized top-n sequential recommendation via convolutional sequence embedding[C]// Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. 2018.
- [20] DING Y, MA Y, WONG W K, et al. Leveraging two types of global graph for sequential fashion recommendation[C]// Proceedings of the 2021 International Conference on Multimedia Retrieval. 2021.
- [21] HO J, JAIN A, ABBEEL P J A I N I P S. Denoising diffusion probabilistic models [J]. arXiv: 2006. 11239, 2020.
- [22] GONG S, LI M, FENG J, et al. Diffuseq: Sequence to sequence text generation with diffusion models [J]. arXiv: 2210. 08933, 2022.
- [23] DU H, YUAN H, HUANG Z, et al. Sequential recommendation with diffusion models [J]. arXiv: 2304. 04541, 2023.
- [24] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017: 6000-6010.
- [25] OORD A V D, LI Y, VINYALS O J A P A. Representation learning with contrastive predictive coding [J]. arXiv: 1807. 03748, 2018.
- [26] MCAULEY J, TARGETT C, SHI Q, et al. Image-based recommendations on styles and substitutes[C]// Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2015.
- [27] SACHDEVA N, MANCO G, RITACCO E, et al. Sequential variational autoencoders for collaborative filtering[C]// Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining. 2019.
- [28] HIDASI B, KARATZOGLOU A. Recurrent neural networks with top-k gains for session-based recommendations[C]// Proceedings of the 27th ACM International Conference on Information and Knowledge Management. 2018.



LI Bo, born in 2003, undergraduate, is a member of CCF (No. T9206G). His main research interests is recommender system.



MO Xian, born in 1990, Ph.D, associate professor, master supervisor, is a member of CCF (No. R6178M). His main research interests include data mining, graph neural network, and recommender system.