

基于对比学习的动态图序列推荐方法

崔昱,陈佳伟*,王灿

(浙江大学 计算机科学与技术学院,浙江 杭州 310027)

摘要:为了缓解目前基于动态图表征的序列推荐研究中存在的用户交互数据高稀疏且包含大量噪音、模型训练需要大量有标记样本等问题,本文提出了基于对比学习的动态图序列推荐方法(CDGSR)。具体而言,CDGSR从粗粒度和细粒度不同层面设计了层间对比学习、两次传播对比学习和随机噪声扰动对比学习三种不同角度的对比学习方法。实验表明,CDGSR在Amazon-Beauty、Amazon-Games、Amazon-CDs三个现实数据集上的归一化折损累计增益NDCG@10分别达到了0.363 3、0.587 3、0.522 0,Hit@10分别达到了0.525 8、0.778 6、0.735 9。与基于矩阵分解的BPR-MF、FPMC,基于神经网络的GRU4Rec、Caser、SASRec和基于图神经网络的SR-GNN、HGN、HyperRec、DGSR等方法相比,CDGSR均取得了最好的结果。其中,与性能最好的DGSR相比,CDGSR在Amazon-CDs数据集上的NDCG@10提升了1.97%,Hit@10提升了1.60%。这些结果表明,本文提出的CDGSR能够有效利用对比学习方法提升动态图序列推荐方法的推荐性能。

关键词:序列推荐;图神经网络;动态图表征;对比学习

中图分类号:TP391 文献标志码:A 文章编号:0253-2395(2024)03-0506-12

Contrastive Dynamic Graph for Sequential Recommendation

CUI Yu, CHEN Jiawei*, WANG Can

(College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China)

Abstract: To alleviate the problems in dynamic graph sequential recommendation, such as sparse and noisy user-item interaction data, and the requirement for a large number of labels, this paper proposes a new dynamic graph sequential recommendation method based on contrastive learning, which is called CDGSR (Contrastive Dynamic Graph for Sequential Recommendation). Specifically, CDGSR designed three different contrastive learning methods from coarse-grained to fine-grained: inter layer contrastive learning, twice propagation contrastive learning and random noise perturbation contrastive learning. The experimental results demonstrate that CDGSR achieves NDCG@10 scores of 0.363 3, 0.587 3, and 0.522 0 on the real-world datasets of Amazon-Beauty, Amazon-Games, and Amazon-CDs, respectively. Additionally, the corresponding Hit@10 scores are 0.525 8, 0.778 6, and 0.735 9. Compared to matrix factorization-based methods like BPR-MF and FPMC, neural network-based methods like GRU4Rec, Caser, SASRec, and graph neural network-based methods like SR-GNN, HGN, HyperRec, and DGSR, CDGSR consistently achieves the best results. Specifically, compared to the best-performing method DGSR, CDGSR improves NDCG@10 by 1.97% and Hit@10 by 1.60% on the Amazon-CDs dataset. These results indicate that CDGSR can effectively utilize contrastive learning to improve the performance of dynamic graph sequential recommendation method.

Key words: sequential recommendation; graph neural networks; dynamic graph representation; contrastive learning

收稿日期:2023-12-15;接受日期:2024-01-12

基金项目:浙江大学上海高等研究院繁星科学基金(SN-ZJU-SIAS-001);国家自然科学基金(62372399)

作者简介:崔昱(2000-),男,博士研究生,研究方向为推荐系统。E-mail:12321228@zju.edu.cn

通信作者:陈佳伟(CHEN Jiawei),E-mail:sleepyhunt@zju.edu.cn

引文格式:崔昱,陈佳伟,王灿.基于对比学习的动态图序列推荐方法[J].山西大学学报(自然科学版),2024,47(3):506-517. DOI:10.13451/j.sxu.ns.2024006

0 引言

序列推荐旨在通过对用户与物品的历史交互序列建模,预测用户接下来可能感兴趣的物品或服务,以此向用户进行推荐来提供个性化服务。随着互联网和移动设备的普及,序列推荐被广泛应用于当今各种在线购物平台、外卖平台、短视频平台等,是如今推荐系统中的热门问题。

早期的序列推荐工作采用的是矩阵分解^[1]、马尔可夫链^[2]等方法。随着神经网络的发展,一些工作提出了基于深度学习的序列推荐方法,例如基于循环卷积网络的序列推荐方法^[3-4]、基于卷积神经网络的序列推荐方法^[5]和基于注意力机制的序列推荐方法^[6]等。尽管这些方法已经取得了一定成功,但是它们将用户交互数据建模为序列信息,这导致模型难以捕捉交互数据中的用户的高阶交互信息和复杂的物品转换关系。

随着图神经网络(Graph Neural Network, GNN)^[7-8]的相关研究逐渐成熟,越来越多的序列推荐研究用GNN来建模复杂的用户与物品交互数据,并发现采用基于GNN的序列推荐方法能获得更好的推荐性能。在现有的采用GNN的序列推荐研究中,基于静态图表征的模型已经取得了不少成果^[9-12],而基于动态图表征的模型发展还未成熟。动态图中节点的时间演化特性有助于识别局部邻域相似但结构不同的节点,真实的用户-物品的历史交互序列应该包含时序信息,利用时序信息可以更好地为用户做出推荐。

尽管动态图序列推荐的方法的确有效,但当前基于动态图的序列推荐模型仍然存在一些局限性:

- 交互数据具有稀疏性。大多数的推荐模型都采用的是监督学习范式^[13-15],其中监督信号来自观察到的用户和物品交互数据。然而,数据集中所观察到的交互数据与实际的整个交互空间相比非常稀疏^[16-17],不足以学习高质量的用户和物品表征,这个问题对于包含时间依赖信息的动态交互数据更为突出。

- 交互数据中存在噪音。用户提供的大多数反馈都是隐式的,观察到的交互数据通常包

含噪音^[18]。例如用户可能被误导而点击或购买某物品,然而事实上并不对该物品感兴趣。与此同时,GNN的各种节点聚合方法扩大了交互噪音对表征学习的影响,这导致用户偏好的学习更容易受到交互噪音的影响。

- 物品存在“长尾效应”。用户和物品的交互数据通常表现出幂律分布^[19-20],其中的“长尾”由缺乏监督信号的冷门物品组成。流行的物品在GNN的邻域聚合和监督损失中出现的频率更高,这导致GNN的各种节点聚合方法很容易偏向于关注流行的物品^[21-22],而忽略了冷门的物品。

自监督学习(Self-Supervised Learning, SSL)^[23]已经在计算机视觉^[24-25]和自然语言处理^[26-28]等领域中得到了应用,它通过增强输入数据来利用未标记的数据空间,来在模型的下游任务中实现性能的改进^[29]。对比学习^[30]作为一种自监督学习方法,通过数据增强建立对比视图来获取正样本和负样本,然后通过最大化正样本之间的相似性和最小化负样本之间的相似性,使得模型能够有效挖掘数据中的自监督信息,利用未标记的数据学习到更具有区分度的特征表示。

为了缓解目前动态图的序列推荐模型用户交互数据的稀疏性和噪音带来的影响、提高模型的泛化能力和推荐的准确性,本文提出了基于对比学习的动态图序列推荐模型称为CDGSR(Contrastive Dynamic Graph for Sequential Recommendation)。CDGSR在基于动态图表征的序列推荐模型^[31]中设计了三种图对比学习方法。首先,受到先前工作简化对比学习模型框架^[32]的启发,CDGSR将动态图推荐网络学习中间层的节点表征和最后得到的节点表征进行对比学习。其次按照序列推荐中图对比学习方法的一般范式^[33],CDGSR简单地将动态图推荐网络进行两次传播,对两次传播得到的节点表征进行对比学习。最后,针对先前工作中动态图推荐网络通过注意力机制聚合节点的长短期兴趣基于动态图表征的序列推荐模型^[31],CDGSR对注意力机制中的权重添加随机噪声进行对比学习。

本文的主要创新点与贡献总结如下:

•提出了一种新的基于对比学习的动态图序列推荐框架 CDGSR, 将对比学习作为辅助任务来提高推荐任务的准确性。

•从不同的角度提出了层间对比学习、两次传播对比学习和随机噪声扰动对比学习三种不同的对比学习方案, 从粗粒度和细粒度提升动态图序列推荐模型的推荐性能。

•在三个真实世界的数据集进行的大量实验表明, CDGSR 的表现优于最先进的序列推荐方法, 证明了 CDGSR 在解决序列推荐问题中的有效性。

本文的结构安排如下: 第 1 章介绍与本文相关的研究工作, 包括基于 GNN 的序列推荐和基于图对比学习的序列推荐。第 2 章介绍本文提出的 CDGSR 模型, 重点阐述三种不同的动态图对比学习方法的设计思路和实现原理。第 3 章介绍实验数据集、实验设置等实验细节, 对比了 CDGSR 和其他序列推荐方法的表现, 并对结果进行了讨论和分析。第 4 章对本文的工作进行了归纳总结, 并对 CDGSR 的未来工作进行了展望。

1 相关工作

1.1 基于图神经网络的序列推荐

近年来, GNN 在处理图结构数据方面表现出了最先进的性能^[7-8], 已经在推荐系统中被广泛使用。基于 GNN 的序列推荐方法通常遵循以下范式^[9-10, 12, 34]: 首先根据用户的历史交互数据进行图结构建模, 然后通过图神经网络方法进行节点聚合和信息传播以学习节点的特征, 最后结合用户的长期和短期偏好表示出用户的最终表征, 进而计算用户与物品交互的得分, 选取其中得分排名 top K 的物品进行推荐。

SR-GNN^[9]是第一个基于 GNN 的序列推荐模型, 它将用户交互序列中的物品数据建模为物品-物品图, 在此之上设计 GNN 算法捕捉物品间的转移关系。之后的方法在 SR-GNN 的基础上从不同角度进行改进。例如结合个性化 GNN 和注意力机制, 在表征中融入用户的个性化信息^[10]; 使用 GNN 对短期内的物品上下文信息进行建模, 并利用共享记忆网络捕获物品间的长期依赖关系^[11]等等。

有一些工作指出, 上述这些基于 GNN 的序列推荐方法在建模用户偏好时, 只关注了用户交互序列内的物品关系, 却忽略了序列间的物品关系。因此一些工作考虑将交互数据建模为更加复杂的图结构, 或者显式利用序列间的相似信息, 以此来捕捉序列之间物品的关系。例如 HyperRec^[34]采用超图来模拟序列内或序列间物品的高阶相关性; CSR^[35]通过计算与当前会话之间的相似度来考虑邻域会话; 还有的工作通过显式关联用户的社会关系属性利用不同的用户序列^[36]等。

然而, 以上基于图结构的序列推荐方法都是基于静态图的方法, 没有考虑用户与物品之间交互的时间信息。真实世界中用户的偏好往往是随着时间变化的, 考虑交互序列的时间信息对于提高推荐的准确率有重要意义。与本文研究最相近的是 DGSR^[31], 该方法通过动态图结构连接不同的用户序列, 探索用户与物品在时间和顺序信息下的交互行为。它的缺点在于没有针对用户和物品稀疏的动态交互数据以及物品的长尾效应等问题设计解决方案。本文将基于这项工作设计对比学习方案进行改进, 与上述基于 GNN 的序列推荐方法的对比分析将在第 3 章阐述。

1.2 基于图对比学习的序列推荐

由于序列推荐经常因为数据稀疏性问题, 导致很难学习到高质量的用户偏好表示, 所以一些研究采用对比学习的方法来缓解序列推荐中数据稀疏性问题^[36-37]。在采用对比学习的序列推荐工作中, 有许多在序列数据上设计对比学习的方法^[38-41]。同时, 随着基于 GNN 的序列推荐方法的发展, 越来越多的工作开始关注基于图对比学习的序列推荐^[42-46]。在这些方法中, 有一部分工作从图本身结构的角度出发, 通过设计不同的图增强方法来增强模型性能。例如, SGL^[33]采用了基于 dropout 的方法, 进行删除节点、删除边或随机游走来实现数据增强; GCL4SR^[42]采用 GraphSAGE 的方法来进行子图采样, 将采样后的子图作为对比视图进行对比学习; SimGCL^[43]和 XSimGCL^[32]通过向节点表征的特征空间中加入随机噪声使表征分布更加均匀, 来增强推荐的准确性。另一部分的

工作从序列推荐问题本身的特点出发,设计解决对应问题的对比学习方案。例如,ICL^[44]为了捕捉用户交互序列中包含的隐藏意图,设计了基于意图的对比学习方案;MCLSR^[45]通过跨视图对比学习范式从兴趣级别和特征级别的四个特定视图学习用户和物品的表征。DCN^[46]同时对用户序列和物品序列进行建模,设计了双重表征对比学习和双重兴趣对比学习。

尽管这些基于图对比学习的序列推荐方法取得了一定成果,但是它们都是为静态图序列推荐设计的。由于动态图中的节点和边包含了用户交互的时间依赖信息,许多静态图的对比学习方法直接运用在动态图序列推荐方法中。本文将从不同角度设计动态图对比学习方案,将其运用在动态图序列推荐模型中,探索它们能否提升动态图序列推荐模型的准确性。

2 基于对比学习的动态图序列推荐模型

本文提出的基于对比学习的动态图序列推荐模型(CDGSR)整体框架参照了DGSR^[31]的设计,不同的是,CDGSR向其中加入了图对比学习模块。具体而言,CDGSR的框架如下:首先,根据所有用户-物品动态交互数据构建出整体的动态图,并对整体动态图按照设计好的采用策略进行子图采样;然后,采样后的子图将通过基于动态图注意力机制的动态图推荐网络,得到用户和物品节点的长期偏好和短期偏好;接着,按照CDGSR设计的三种不同的图对比学习方案生成各自的对比视图进行对比学习;最后,通过多任务学习的方式结合对比学习任务 and 推荐任务进行推荐。

2.1 动态图构建与子图采样

对于一个给定的用户交互序列 $S_u = (i_1, i_2, i_3, \dots, i_k)$, 对应的时间戳序列是 $T_u = (t_1, t_2, t_3, \dots, t_k)$, 其中 u 表示用户, i 表示物品, t 表示交互产生的时间。当用户 u 在时刻 t 和物品 i 产生交互时,会在 u 和 i 之间建立一条边 e , 这次交互可以用五元组 (u, i, t, o_u^i, o_i^u) 表示,其中 o_u^i 是物品 i 在与用户 u 交互的所有物品中的相对位置, o_i^u 是用户 u 在与物品 i 交互的所有用户中的相对位置。所有交互的五元组构建了动

态图 $G = \{(u, i, t, o_u^i, o_i^u) \mid u \in U, i \in V\}$ 。这种方式构建的动态图 G 除了记录用户与物品的交互信息,还记录了用户与物品的交互的时间信息和顺序信息^[31]。预测序列 S_u 的下一个物品就相当于预测在图 G^t 中最有可能与用户节点 u 产生链接的下一个物品 i 。

随着用户交互序列 S_u 的扩展,其邻居序列的数量也在不断增加,由所有用户组成的动态图的规模也在逐渐扩大,这会增加计算量并引入噪声,因此需要子图采样。子图采样策略如下:首先以用户节点 u 为锚点节点,在图 G^t 中选择与锚点节点最近的 n 个邻居节点,记为 N_u , 其中 n 是采样过程中用户和物品的最大序列长度。接下来,对于每一项物品 $i \in N_u$, 依次使用它们作为锚点节点,对最近与它们产生交互的 n 个用户进行采样,写成 N_i 。在采样过程中记录被用作锚点节点的用户节点和物品节点防止重复采样。最后,可以形成用户节点 u 的 m 阶子图 $G_u^m(t_k)$, 其中 m 是用来控制子图大小的超参数

2.2 动态图表征网络

2.2.1 消息传播

动态图表征网络的消息传播过程包括了从物品到用户的传播和从用户到物品的传播,这两个传播过程都需要对长期偏好和短期偏好进行编码。用户的长期偏好反映了用户的固有特征和总体偏好,而用户的短期偏好反映了用户最近的兴趣。与用户类似,物品的长期偏好可以反映物品的总体特征,而短期偏好能体现物品的最新属性。

长期偏好通过动态注意力模块(DAT)编码^[31]。DAT将图注意力机制与序列信息编码相结合,能有效区分不同物品对于用户的影响并利用交互的顺序信息。以用户的长期偏好编码为例,DAT将 r_u^i 定义为物品 i 与用户 u 的邻居中最后一个物品的相对顺序,即 $r_u^i = |N_u| - o_i^u$ 。对于每一个相对顺序的离散值 r , 可以映射到一个一维的参数向量 $\mathbf{p}_r^k \in \mathbb{R}^d$, 将它作为相对位置的嵌入向量来编码位置信息。第 $l-1$ 层中用户表征 h_u^{l-1} 与其邻居节点的表征 h_i^{l-1} 之间的注意力系数会受到 \mathbf{p}_r^k 的影响。用户 u 与其邻居物品 i 之间的注意力分数 e_{ui} 计算如下:

$$e_{ui} = \frac{(\mathbf{W}_2^{l-1} h_u^{l-1})^\top (\mathbf{W}_1^{l-1} h_i^{l-1} + \mathbf{p}_{r_u^i}^K)}{\sqrt{d}}, \quad (1)$$

其中, \mathbf{W}_1^{l-1} 和 \mathbf{W}_2^{l-1} 是可训练的权重参数, d 是嵌入向量的维度, 比例因子 \sqrt{d} 是为了避免点积过大, 加快收敛速度。用户 u 与其邻居物品 i 之间的加权注意力分数 α_{ui} 计算如下:

$$\alpha_{ui} = \text{softmax}(e_{ui}). \quad (2)$$

通过 DAT 聚合所有邻居的信息, 最终可以得到用户的长期偏好 h_u^l :

$$h_u^l = \sum_{i \in N_u} \alpha_{ui} (\mathbf{W}_2^{l-1} h_u^{l-1} + \mathbf{p}_{r_u^i}^V), \quad (3)$$

其中 $\mathbf{p}_{r_u^i}^V \in \mathbb{R}^d$ 为用户 u 的交互序列中物品 i 相对顺序的嵌入向量, 用于捕获消息聚合中物品的顺序信息。

同理, 可以通过 DAT 对物品的邻居节点进行消息传播, 得到物品的长期偏好 h_i^l :

$$e_{iu} = \frac{(\mathbf{W}_2^{l-1} h_u^{l-1})^\top (\mathbf{W}_1^{l-1} h_i^{l-1} + \mathbf{p}_{r_u^i}^V)}{\sqrt{d}}, \quad (4)$$

$$\beta_{iu} = \text{softmax}(e_{iu}), \quad (5)$$

$$h_i^l = \sum_{u \in N_i} \beta_{iu} (\mathbf{W}_1^{l-1} h_i^{l-1} + \mathbf{p}_{r_u^i}^V), \quad (6)$$

其中 β_{iu} 表示加权注意力分数, $\mathbf{p}_{r_u^i}^V \in \mathbb{R}^d$ 为物品 i 的被交互序列中用户 u 相对顺序的嵌入向量, 用于捕获消息聚合中用户的顺序信息。

短期兴趣通常可以通过用户或物品最近的交互行为来反映。许多早期的推荐方法使用最后一个交互物品作为用户的短期兴趣^[9], 然而这一方法忽略了历史交互信息的重要性。CDGSR 使用注意力机制来捕捉最后一次交互的物品或用户节点与每个历史交互的物品或用户之间的关系, 通过这种方式学习到的短期兴趣表示如下:

$$h_u^S = \sum_{i \in N_u} \hat{\alpha}_{ui} h_i^{l-1}, \quad (7)$$

$$h_i^S = \sum_{u \in N_i} \hat{\beta}_{iu} h_u^{l-1}, \quad (8)$$

其中 $\hat{\alpha}_{ui}$ 和 $\hat{\beta}_{iu}$ 分别表示用户节点和物品节点与它们的邻居节点之间的注意力分数, 它们的计算公式如下:

$$\hat{\alpha}_{ui} = \text{softmax} \left(\frac{(\mathbf{W}_3^{l-1} h_{i_{N_u}}^{l-1})^\top (\mathbf{W}_2^{l-1} h_i^{l-1})}{\sqrt{d}} \right), \quad (9)$$

$$\hat{\beta}_{iu} = \text{softmax} \left(\frac{(\mathbf{W}_4^{l-1} h_{u_{N_i}}^{l-1})^\top (\mathbf{W}_1^{l-1} h_u^{l-1})}{\sqrt{d}} \right) \quad (10)$$

其中, $h_{i_{N_u}}^{l-1}$ 表示用户交互序列的最后一个物品表征, $h_{u_{N_i}}^{l-1}$ 表示最后一个与该物品交互的用户表征, \mathbf{W}_3^{l-1} 和 \mathbf{W}_4^{l-1} 为控制最后一次交互的权值矩阵。

2.2.2 节点更新

为了保留不同层次的信息, CDGSR 将消息传播过程所学习到的节点长期偏好表征、短期偏好表征和节点的前一层表征共同拼接在一起, 更新节点的表示。

对于用户节点, 从第 $l-1$ 层到第 l 层的用户节点更新规则可表示为:

$$h_u^l = \tanh(\mathbf{W}_3^l [h_u^l \| h_u^S \| h_u^{l-1}]), \quad (11)$$

其中 $\mathbf{W}_3^l \in \mathbb{R}^{d \times 3d}$ 为用户更新矩阵, 用于控制用户节点长期偏好、短偏好和前一层表征的权重。

对于物品节点, 从第 $l-1$ 层到第 l 层的物品节点更新规则可表示为:

$$h_i^l = \tanh(\mathbf{W}_4^l [h_i^l \| h_i^S \| h_i^{l-1}]), \quad (12)$$

其中 $\mathbf{W}_4^l \in \mathbb{R}^{d \times 3d}$ 为物品更新矩阵, 用于控制物品节点长期偏好、短偏好和前一层表征的权重。

2.3 图对比学习模块

CDGSR 中的图对比学习模块考虑了三种不同角度的对比学习方法, 如图 1 所示。

2.3.1 层间对比学习

在先前的工作中, XSimGCL^[32] 使用跨层对比学习的方法来优化其先前遵循了图对比学习一般范式的 SimGCL^[43] 模型, 在降低了计算复杂度的同时, 也取得了较好的实验结果。受此启发, CDGSR 模型里的层间对比学习方案设计如下: 在层数为 l 的动态图表征网络中, 我们将第 l 层动态图表征网络学习到的包含时序信息的用户和物品表征视为一个对比视图, 同时, 将第 $l-1$ 层动态图表征网络学习到的包含时序信息的用户和物品节点表征视为另一个对比视图。将同一个用户节点或物品节点的对比视图作为正样本, 不同用户节点或物品节点的对比视图视为负样本, 采用信息噪声对比估计 (Information Noise-Contrastive Estimation, InfoNCE) 损失函数^[25] 计算用户节点和物品节点的对比损失:

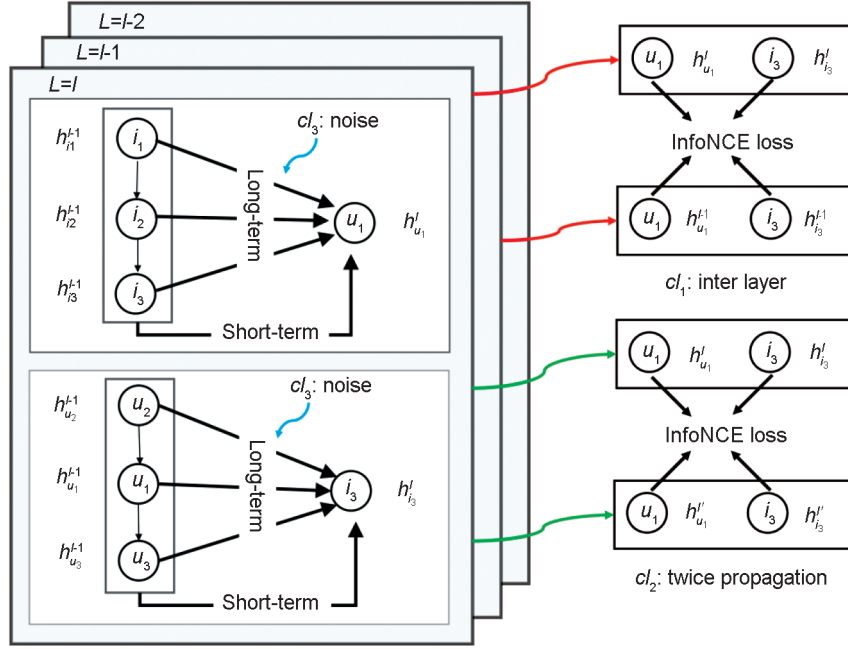


图1 CDGSR中三种不同对比学习方法

注:红色的 cl_1 表示层间对比学习方案,绿色的 cl_2 表示两次传播对比学习方案,蓝色的 cl_3 表示向动态图注意力权重加入随机噪声对比学习的方案。

Fig. 1 Three different contrastive learning methods in CDGSR

$$L_{cl}^{(u)} = -\log \frac{\exp(s(h_{i,l}^{(u)}, h_{i,l-1}^{(u)})/\tau)}{\sum_{i=0}^I \exp(s(h_{i,l}^{(u)}, h_{i,l-1}^{(u)})/\tau)}, \quad (13)$$

$$L_{cl}^{(i)} = -\log \frac{\exp(s(h_{u,l}^{(i)}, h_{u,l-1}^{(i)})/\tau)}{\sum_{u=0}^U \exp(s(h_{u,l}^{(i)}, h_{u,l-1}^{(i)})/\tau)}, \quad (14)$$

其中 $h_{i,l}^{(u)}$ 和 $h_{u,l}^{(i)}$ 表示第 l 层动态图表征网络学习到的节点表征, $h_{i,l-1}^{(u)}$ 和 $h_{u,l-1}^{(i)}$ 表示第 $l-1$ 层动态图表征网络学习到的节点表征, $s(\cdot)$ 表示度量两个用户/物品节点表征相似度函数。 τ 代表温度参数,控制模型对负样本的区分度。

2.3.2 两次传播对比学习

序列推荐中图对比学习方法的一般需要先进行数据增强建立对比视图,然后通过对比两个增强后的视图来实现对比学习^[33]。在本文提出的两次传播对比学习方案中,CDGSR不进行复杂的数据增强,而是简单地将动态图表征网络进行两次传播,把第一次传播和第二次传播通过动态图表征网络学习到的用户和物品节点表征视为两个不同的对比视图。同一个用户节点或物品节点的两个对比视图作为正样本,不同用户节点或物品节点的这两个视图视为负样本,采用InfoNCE损失函数计算用户节点和

物品节点的对比损失,计算公式如下:

$$L_{cl}^{(u)} = -\log \frac{\exp(s(h_{i,l}^{(u)}, h_{i,l}^{(u)})/\tau)}{\sum_{i=0}^I \exp(s(h_{i,l}^{(u)}, h_{i,l}^{(u)})/\tau)}, \quad (15)$$

$$L_{cl}^{(i)} = -\log \frac{\exp(s(h_{u,l}^{(i)}, h_{u,l}^{(i)})/\tau)}{\sum_{u=0}^U \exp(s(h_{u,l}^{(i)}, h_{u,l}^{(i)})/\tau)}, \quad (16)$$

其中 $h_{i,l}^{(u)}$ 和 $h_{u,l}^{(i)}$ 表示第一次传播动态图表征网络学习到的节点表征, $h_{i,l}^{(u)}$ 和 $h_{u,l}^{(i)}$ 表示第二次传播动态图表征网络学习到的节点表征, $s(\cdot)$ 表示度量两个用户/物品节点表征相似度函数。 τ 代表温度参数,控制模型对负样本的区分度。

2.3.3 向权重加入随机噪声对比学习

本文动态表征网络中的注意力机制会计算用户节点与所有邻居物品节点之间的注意力分数。注意力分数通常代表着该用户对于每一个邻居物品的喜好程度,热门的物品往往注意力分数会更高,而冷门的物品注意力分数会很小,并且在网络多次迭代的过程中越来越小。这将导致冷门物品受到推荐的概率更小,从而加剧“长尾效应”,降低推荐的性能。因此,向权重加入随机噪声对比学习就是指向动态表征网络的注意力权重加入很小的随机噪声,以防

止冷门物品的注意力分数过小。我们将原动态表征网络学习到的用户和物品节点表征视为一个对比视图,将加入随机噪声后的方法学习到的节点表征视为另一个对比视图。相同一个用户节点或物品节点的两个视图作为正样本,不同用户节点或物品节点的两个视图视为负样本,对比损失函数仍然使用 InfoNCE 计算,如下所示:

$$L_{cl}^{(u)} = -\log \frac{\exp(s(h_{i,l}^{(u)}, h_{i,l}^{(u)})/\tau)}{\sum_{i'=0}^I \exp(s(h_{i,l}^{(u)}, h_{i',l}^{(u)})/\tau)}, \quad (17)$$

$$L_{cl}^{(i)} = -\log \frac{\exp(s(h_{u,l}^{(i)}, h_{u,l}^{(i)})/\tau)}{\sum_{u'=0}^U \exp(s(h_{u,l}^{(i)}, h_{u',l}^{(i)})/\tau)}, \quad (18)$$

其中 $h_{i,l}^{(u)}$ 和 $h_{u,l}^{(i)}$ 表示原始动态图表征网络学习到的节点表征, $h_{u,l}^{(i)}$ 和 $h_{i,l}^{(u)}$ 表示向注意力权重加入随机噪声后学习到的节点表征, $s(\cdot)$ 表示度量两个用户/物品节点表征相似度函数。 τ 代表温度参数,控制模型对负样本的区分度。

2.4 目标函数

序列推荐的目标是为用户推荐下一个最符合用户偏好的物品,即预测用户交互序列 $S_u = (i_1, i_2, i_3, \dots, i_k)$ 的下一物品交互 i_{k+1} , 在图上表现为预测子图 $G_u^m(t_k)$ 中与用户节点 u 的最有可能链接的物品 i 。

将动态表征网络作用在 $G_u^m(t_k)$ 上,经过消息传播和节点更新,最后可以得到用户节点 u 的多层表示: $\{h_u^{(0)}, h_u^{(1)}, \dots, h_u^{(L)}\}$, 在不同层中的用户表征强调了不同的用户偏好信息。因此,可以将多个表征进行拼接,得到用户节点 u 的最终表示:

$$h_u = h_u^{(0)} \parallel h_u^{(1)} \parallel \dots \parallel h_u^{(L)}. \quad (19)$$

对于给定的候选物品 $i \in I$, 用户 u 与该物品交互的分数计算方式如下:

$$s_{ui} = h_u^T \mathbf{W}_p e_i, \quad (20)$$

其中向量 $s_u = \{s_{u1}, s_{u2}, \dots, s_{u|I|}\}$ 表示每个候选物品在用户 u 的得分向量, $\mathbf{W}_p \in \mathbb{R}^{(L+1)d \times d}$ 为可训练变换矩阵, e_i 表示物品的表征。

用户 u 对候选物品的归一化得分 \hat{y}_u 表示为:

$$\hat{y}_u = \text{softmax}(s_u). \quad (21)$$

推荐任务的目标函数为交叉熵损失如下:

$$L_{rec} = -\sum_S \sum_{i=1}^{|I|} y_{ui} \log(\hat{y}_{ui}) + (1-y_{ui}) \log(1-\hat{y}_{ui}) + \lambda \|\Theta\|_2, \quad (22)$$

其中 y_u 表示在用户 u 的交互序列 S_u 中真实的下一个物品样本, Θ 表示模型中所有的变量, $\|\cdot\|_2$ 表示 L_2 正则化, λ 用于控制正则化的强度。

CDGSR 采用多任务学习的方式进行模型的训练,以对比学习作为辅助任务,推荐任务作为主要任务。最终任务的目标函数表示如下:

$$\text{Loss} = L_{rec} + \lambda_{cl} L_{cl}, \quad (23)$$

其中, λ_{cl} 用于控制对比学习任务的权重。

3 实验与分析

3.1 实验数据集

为了评估 CDGSR 模型的有效性,本文选择了三个不同的公开现实数据集进行了实验。这三个数据集分别是:亚马逊 CD 数据集 (Amazon-CDs)、亚马逊游戏数据集 (Amazon-Games) 和亚马逊美妆数据集 (Amazon-Beauty), 它们的特征如表 1 所示。

表 1 Beauty, Games 和 CDs 数据集的特征

Table 1 Statistics of Beauty, Games and CDs datasets

数据集	Beauty	Games	CDs
用户数量	52 024	31 013	17 052
物品数量	57 289	23 715	35 118
交互数量	394 908	287 107	472 265
序列平均长度	7.6	9.3	27.6
数据密集度	0.01%	0.04%	0.08%

3.2 实验设置

为了证明 CDGSR 模型的有效性,本文将所提出方法与以下方法进行了比较:

- BPR-MF^[1], 一种基于矩阵分解的协同过滤推荐模型,通过学习用户的隐式反馈行为进行推荐。

- FPMC^[2], 一种结合了矩阵分解和一阶马尔可夫链的推荐模型,旨在捕捉用户的长期偏好和物品之间的转换关系。

- GRU4Rec+^[4], 一种基于循环神经网络 (Recurrent Neural Network, RNN) 改进的推荐模型,采用了不同的损失函数和采样策略,旨在提高 Top-K 物品的推荐性能。

• Caser^[5], 一个基于 CNN 的推荐模型, 旨在通过对 L 个最近交互物品应用卷积运算来捕获高阶马尔可夫链。

• SASRec^[6], 一种基于自注意力的模型, 用于通过自注意力机制捕捉物品的相关性以预测下一个物品。

• SR-GNN^[9], 第一个基于 GNN 的序列推荐模型, 旨在建模基于会话的推荐物品的复杂过渡关系。

• HGN^[47], 一个序列模型, 包含特征门控、实例门控和实例门控模块, 用于选择重要特征并明确捕获物品关系。

• TiSASRec^[48], 一个基于间隔感知的自注意力模型, 该模型对序列中的绝对位置以及它们之间的时间间隔进行建模。

• HyperRec^[34], 一个基于超图的模型, 它采用超图来捕捉物品之间的高阶连接、对下一个物品进行推荐。

• DGSR^[31], 一个基于动态图序列推荐的模型, 它同时考虑了用户和物品长期和短期信息, 采用包含时间信息的动态图注意力机制对下一个物品进行推荐。

为了实验的公平性, 本文在实现参与不同方法时采用原方法默认的超参数。CDGSR 中的用户和物品的节点维度 d 设置为 50, 中心节点的最大交互序列长度 n 设置为 50, 对于每个用户序列, 使用最后一个物品用于测试, 倒数第二个物品用于验证, 剩余的所有物品都用于模型训练。动态图子图采样的跳数 k 设置为 3, 动态图推荐网络的层数 L 设置为 3。采用 Adam 优化器, 学习率设置为 0.01, 训练的 batch 大小设置为 50。按照经验, 本文将控制随机噪声的强度的超参数 ϵ 设置为 0.2。对比学习 InfoNCE 损失函数中温度参数 τ 的值从 [0.1, 0.15, 0.2] 中选择表现最好 τ 的作为最终的 τ 值。所有实验都是在装有八个 NVIDIA GeForce RTX2080Ti (11GB) 和两个 Intel Xeon Gold 6254 CPU 的计算机服务器上进行的。

3.3 实验结果分析

表 2 展示了不同方法在三个实验数据集上归一化折损累计增益 (Normalized Discounted Cumulative Gain, NDCG) @10 和 Hit@10 的表

现, 每个评价指标的值均为百分数。CDGSR 取三种对比学习方法中表现最好的实验结果。表中加下划线的数字代表着所有模型评价指标中排行第二好的结果, 而加粗的数字表示最好的结果。最后一行的 Gain 表示最好的结果比第二好的结果效果提升的百分比。

表 2 不同方法的比较结果

Table 2 Comparison results of different methods

Methods	Beauty		Games		CDs	
	NDCG	Hit	NDCG	Hit	NDCG	Hit
	@10	@10	@10	@10	@10	@10
BPR-MF	21.83	37.75	28.75	37.75	36.26	56.27
FPMC	28.91	43.10	46.80	68.02	33.55	51.22
GRU4Rec+	26.42	43.98	45.64	67.15	44.52	67.84
Caser	25.47	42.64	45.93	68.83	45.85	68.65
SASRec	32.19	48.54	53.60	73.98	49.23	71.32
SR-GNN	32.33	48.62	53.25	73.49	48.95	69.63
HGN	32.47	48.63	49.34	71.42	49.34	71.42
TiSASRec	30.45	46.87	50.19	71.85	48.97	71.00
HyperRec	23.26	34.71	48.96	71.24	47.16	71.02
DGSR	<u>35.90</u>	<u>52.40</u>	<u>58.52</u>	<u>77.64</u>	<u>51.22</u>	<u>72.43</u>
CDGSR	36.33	52.58	58.73	77.86	52.20	73.59
Gain	1.19%	0.34%	0.36%	0.28%	1.97%	1.60%

可以看出, CDGSR 模型在三个数据集的不同评价指标上都实现了最佳性能。与没有加入对比学习方案的动态图序列推荐模型 DGSR 相比, CDGSR 的 NDCG@10 值和 Hit@10 值在三个数据集上都得到了提高。这说明了通过向动态图序列推荐模型加入 CDGSR 设计的图对比学习方法, 可以更加有效地利用稀疏和具有噪音的用户交互数据, 挖掘交互数据中的自监督信号来辅助推荐任务提高推荐的准确率。

DGSR 和 CDGSR 的表现明显优于前面几种推荐方法。这是因为 BPR-MF 和 FPMC 是早期的没有采用深度学习的推荐方法, 推荐的性能受到一定局限; 与基于序列方法的 GRU4Rec+、Caser 和 SASRec 相比, DGSR 和 CDGSR 能通过图神经网络捕捉交互数据中的高阶依赖信息; 与基于静态图方法的 SR-GNN、HGN、TiSASRec 和 HyperRec 相比, DGSR 和 CDGSR 模型能够明确地利用交互数据中的时间信息和顺序信息, 从而获得更好的推荐性能。

根据对数据集数据特征的统计, 亚马逊美妆数据集作为稀疏性最高、用户交互序列最短

的数据集,其所构建的动态图也是最稀疏的。而本文的 CDGSR 模型在亚马逊美妆数据集的 NDCG@10 指标上有了显著的提升,这表明 CDGSR 模型可以更好地适应稀疏的数据集,且具有更好的泛化能力和预测准确度,这对于解决实际推荐系统中的数据稀疏问题具有重要意义。

表 3 展示了 CDGSR 中三种不同的对比学习方法在亚马逊游戏和美妆数据集上的实验结果。表中 cl_1 表示层间对比学习的方案,用 cl_2 表示两次传播对比学习的方案,用 cl_3 表示向动态图注意力权重加入随机噪声的对比学习的方案,最好的结果用加粗表示。

表 3 三种对比学习方法比较的结果

Table 3 Comparison of three contrastive learning methods

数据集	评价指标	cl_1	cl_2	cl_3
Games	NDCG@10	0.586 9	0.585 5	0.587 3
	Hit@10	0.780 3	0.780 6	0.778 6
Beauty	NDCG@10	0.363 3	0.360 9	0.360 4
	Hit@10	0.525 8	0.521 8	0.519 7

可以看出,三种不同的对比学习方法均在某一指标上取得了最优表现。第一种层间对比学习方法和第二种两次传播对比学习方法都通过在动态图推荐网络层上建立对比视图来提高模型性能,是粗粒度的对比学习设计方案。相比之下,第三种使用随机噪声向归一化注意力分数权重中引入随机噪声的方法是一种更细粒度的对比学习设计方案。实验数据表明,在 CDGSR 模型中无论是采用粗粒度的对比学习设计方案还是使用更细粒度的对比学习设计方案,都可以为 CDGSR 模型的性能提升做出贡献,这说明粗粒度的对比学习设计方案和细粒度的对比学习设计方案都有它们各自独特的优点。在后续的工作中,可以尝试探索更多不同粗细粒度的对比学习设计方案,并且可以尝试将各种对比方案一起使用,发挥它们各自的优点,共同提升模型的推荐性能。

除此之外,在三种对比学习方案中,层间对比学习方法所花费的时间成本和计算成本最少,却在亚马逊美妆数据集的两个评价指标上均取得了最优结果。这说明即使是简单的对比学习方案都能够有效地提升动态图序列推荐模

型的效果,进一步说明了在序列推荐模型中引入对比学习方案的重要意义。

3.4 敏感性分析

为了探索对比学习强度控制超参数对 CDGSR 模型性能的影响,本文研究了控制对比学习权重的超参数 λ_1 、 λ_2 和 λ_3 对实验性能的影响。不同的 λ_1 、 λ_2 和 λ_3 的值在亚马逊游戏数据集上的实验结果如图 2 所示。

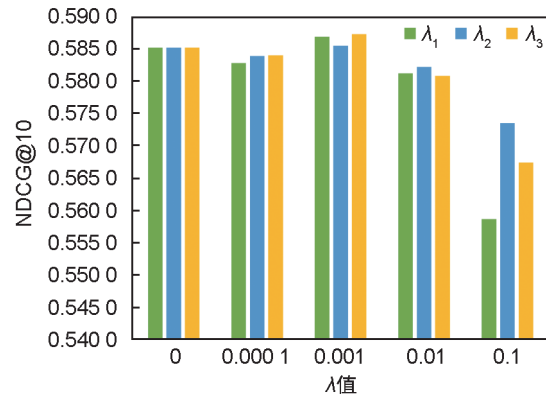


图 2 不同的 λ 对亚马逊游戏数据集的影响

注: λ_1 表示层间对比学习方案, λ_2 表示两次传播对比学习方案, λ_3 表示向动态图注意力权重加入随机噪声对比学习的方案。

Fig. 2 The impact of different λ on the Amazon-Games dataset

在亚马逊游戏数据集上,随着的 λ_1 、 λ_2 和 λ_3 从 0.0001 到 0.1 增加,评估指标 NDCG@10 呈现先提高后下降的变化趋势。这说明了三种对比学习模块在亚马逊游戏数据集上的效果相近,过小和过大的 λ_{cl} 权重都不能提升模型的表现。过小的 λ_{cl} 权重没有体现出对比学习方案的优势,反而影响了推荐任务的学习;而过大的 λ_{cl} 权重导致对比学习方案作为辅助任务占据了太多比重,同样也影响了推荐任务的学习。因此,设计合适的超参数 λ_{cl} 对提升对比学习的效果十分重要。

为了探索不同对比学习方法在更稀疏数据集上的表现,本文在亚马逊美妆数据集上也进行了实验。不同的 λ_1 、 λ_2 和 λ_3 的值在亚马逊美妆数据集上的实验结果如图 3 所示。

可以看出,加入对比学习后推荐性能得到提升的实验结果数量明显增加,这说明对比学习在亚马逊美妆数据集上带来的提升比在亚马逊游戏数据集上带来的提升更加明显。一种可能的原因是:亚马逊美妆数据集比亚马逊游戏

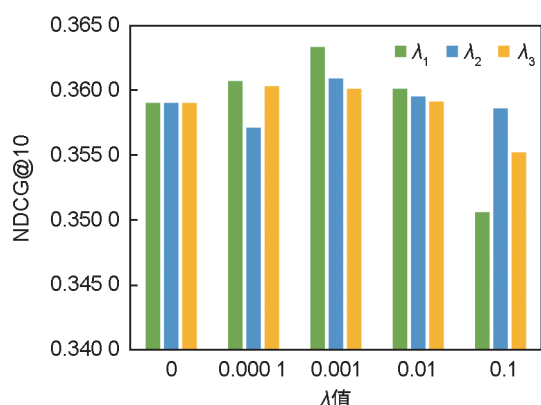


图3 不同的 λ 对亚马逊美妆数据集的影响

注: λ_1 表示层间对比学习方案, λ_2 表示两次传播对比学习方案, λ_3 表示向动态图注意力权重加入随机噪声对比学习的方案。

Fig. 3 The impact of different λ on the Amazon-Beauty dataset

数据集更加稀疏,在亚马逊美妆数据集上应用对比学习,能够更好地缓解用户和物品交互数据稀疏的问题,这也体现了对比学习方法的重要优势。

层间对比学习方法和两次传播对比学习方法都是在动态图推荐网络层的角度设计对比方案,可以称为粗粒度的对比学习方法;向注意力分数增加随机噪声的对比学习方案是在节点消息传播的角度设计对比方案,可以称为细粒度的对比学习方法。整体上看,在两个数据集上三种对比学习方法都对实验结果有一定的提升,这个结果说明了粗粒度对比方案和细粒度对比方案都能够对CDGSR模型的性能带来提升,并且有各自的优点。在后续的工作中,可以将三种不同粒度的对比学习方法同时使用,发挥各自的优点以共同提升模型的推荐性能。

4 总结与展望

本文以基于动态图序列推荐方法为背景,提出了基于对比学习的序列推荐模型CDGSR,从不同角度设计了层间对比学习、两次传播对比学习和向权重加入随机噪声对比学习这三种对比学习方案,将对比学习融入动态图序列推荐模型的框架中,以缓解用户交互数据中存在的稀疏、包含噪音等问题,从而提升动态图序列推荐模型的推荐性能。在三个真实世界的数据集上进行的大量实验表明,本文的CDGSR模型表现优于目前最好的动态图序列推荐

模型,能够有效地提升基于动态图序列推荐方法的推荐准确性。

在本文研究工作的基础上,仍有以下方向值得进一步研究:由于粗粒度和细粒度对比方案都能够对动态图序列推荐模型的性能带来提升,所以可以通过将不同的对比学习方案都加入最终的目标函数中,融合粗粒度和细粒度对比方案的优点,进一步提升推荐模型的准确性。除此之外,目前动态图对比学习的相关研究较少,许多已有的动态图对比学习方法并不能直接用于序列推荐中。因此,后续的工作可以在研究新的图对比学习方法时,关注最新的动态图对比学习方法和序列推荐中其他的图对比学习方法,尝试提出更有效的动态图对比学习方案。

参考文献:

- [1] RENDLE S, FREUDENTHALER C, GANTNER Z, *et al.* BPR: Bayesian Personalized Ranking from Implicit Feedback[C]//Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence. New York: ACM, 2009: 452-461. DOI: 10.5555/1795114.1795167.
- [2] 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. New York: ACM, 2010: 811-820. DOI: 10.1145/1772690.1772773.
- [3] HIDASI B, KARATZOGLOU A, BALTRUNAS L, *et al.* Session-based Recommendations With Recurrent Neural Networks[J]. *Comput Ence*, 2016. DOI: 10.48550/arXiv.1511.06939.
- [4] 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. New York: ACM, 2018: 843-852. DOI: 10.1145/3269206.3271761.
- [5] TANG J X, 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. New York: ACM, 2018: 565-573. DOI: 10.1145/3159652.3159656.
- [6] KANG W C, MCAULEY J. Self-attentive Sequential Recommendation[C]//2018 IEEE International Conference on Data Mining (ICDM). Singapore: IEEE, 2018: 197-206. DOI: 10.1109/ICDM.2018.00035.
- [7] KIPF T N, WELING M. Semi-supervised Classification with Graph Convolutional Networks[EB/OL]. arXiv

- Preprint: 1609.02907, 2016. <https://doi.org/10.48550/arXiv.1609.02907>.
- [8] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, *et al.* Graph Attention Networks[EB/OL]. arXiv Preprint: 1710.10903, 2017. <https://doi.org/10.48550/arXiv.1710.10903>.
- [9] WU S, TANG Y Y, ZHU Y Q, *et al.* Session-based Recommendation with Graph Neural Networks[J]. *Proc AAAI Conf Artif Intell*, 2019, **33**(1): 346–353. DOI: 10.1609/aaai.v33i01.3301346.
- [10] ZHANG M Q, WU S, GAO M, *et al.* Personalized Graph Neural Networks with Attention Mechanism for Session-aware Recommendation[J]. *IEEE Trans Knowl Data Eng*, 2022, **34**(8): 3946–3957. DOI: 10.1109/TKDE.2020.3031329.
- [11] MA C, MA L H, ZHANG Y X, *et al.* Memory Augmented Graph Neural Networks for Sequential Recommendation[J]. *Proc AAAI Conf Artif Intell*, 2020, **34**(4): 5045–5052. DOI: 10.1609/aaai.v34i04.5945.
- [12] FAN W Q, MA Y, LI Q, *et al.* Graph Neural Networks for Social Recommendation[C]//WWW '19: The World Wide Web Conference. New York: ACM, 2019: 417–426. DOI: 10.1145/3308558.3313488.
- [13] HE X N, DENG K, WANG X, *et al.* LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation[C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2020: 639–648. DOI: 10.1145/3397271.3401063.
- [14] HE X N, LIAO L Z, ZHANG H W, *et al.* Neural Collaborative Filtering[C]//Proceedings of the 26th International Conference on World Wide Web. New York: ACM, 2017: 173–182. DOI: 10.1145/3038912.3052569.
- [15] RENDLE S, FREUDENTHALER C, GANTNER Z, *et al.* BPR: Bayesian Personalized Ranking from Implicit Feedback[EB/OL]. arXiv Preprint: 1205.2618, 2012. <https://doi.org/10.48550/arXiv.1205.2618>.
- [16] BAYER I, HE X N, KANAGAL B, *et al.* A Generic Coordinate Descent Framework for Learning from Implicit Feedback[C]//Proceedings of the 26th International Conference on World Wide Web. New York: ACM, 2017: 1341–1350. DOI: 10.1145/3038912.3052694.
- [17] HE R N, MCAULEY J. Ups and Downs: Modeling the Visual Evolution of Fashion Trends with One-class Collaborative Filtering[C]//Proceedings of the 25th International Conference on World Wide Web. New York: ACM, 2016: 507–517. DOI: 10.1145/2872427.2883037.
- [18] WANG W J, FENG F L, HE X N, *et al.* Denoising Implicit Feedback for Recommendation[C]//Proceedings of the 14th ACM International Conference on Web Search and Data Mining. New York: ACM, 2021: 373–381. DOI: 10.1145/3437963.3441800.
- [19] CLAUSET A, SHALIZI C R, NEWMAN M E J. Power-law Distributions in Empirical Data[J]. *SIAM Rev*, 2009, **51**(4): 661–703. DOI: 10.1137/070710111.
- [20] MILOJEVIĆ S. Power Law Distributions in Information Science: Making the Case for Logarithmic Binning [J]. *J Am Soc Inf Sci Technol*, 2010, **61**(12): 2417–2425. DOI: 10.1002/asi.21426
- [21] CHEN J W, DONG H D, WANG X, *et al.* Bias and Debias in Recommender System: a Survey and Future Directions[J]. *ACM Trans Inf Syst*, **41**(3): 67. DOI: 10.1145/3564284.
- [22] TANG X F, YAO H X, SUN Y W, *et al.* Investigating and Mitigating Degree-related Biases in Graph Convolutional Networks[C]//Proceedings of the 29th ACM International Conference on Information & Knowledge Management. New York: ACM, 2020: 1435–1444. DOI: 10.1145/3340531.3411872.
- [23] LIU X, ZHANG F J, HOU Z Y, *et al.* Self-supervised Learning: Generative or Contrastive[J]. *IEEE Trans Knowl Data Eng*, 2023, **35**(1): 857–876. DOI: 10.1109/TKDE.2021.3090866.
- [24] GIDARIS S, SINGH P, KOMODAKIS N. Unsupervised Representation Learning by Predicting Image Rotations[EB/OL]. arXiv Preprint: 1803.07728, 2018. <https://doi.org/10.48550/arXiv.1803.07728>.
- [25] VAN DEN OORD A, LI Y Z, VINYALS O. Representation Learning with Contrastive Predictive Coding[EB/OL]. arXiv Preprint:1807.03748, 2018. <https://doi.org/10.48550/arXiv.1807.03748>.
- [26] DEVLIN J, CHANG M W, LEE K, *et al.* BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[EB/OL]. arXiv Preprint: 1810.04805, 2018. <https://doi.org/10.48550/arXiv.1810.04805>.
- [27] LAN Z Z, CHEN M D, GOODMAN S, *et al.* ALBERT: A Lite BERT for Self-supervised Learning of Language Representations[EB/OL]. arXiv Preprint: 1909.11942, 2019. <https://doi.org/10.48550/arXiv.1909.11942>.
- [28] LOGESWARAN L, LEE H. An Efficient Framework for Learning Sentence Representations[EB/OL]. arXiv Preprint: 1803.02893, 2018. <https://doi.org/10.48550/arXiv.1803.02893>.
- [29] CHEN T, KORNBLITH S, NOROUZI M, *et al.* A Simple Framework for Contrastive Learning of Visual Representations[C]//Proceedings of the 37th International Conference on Machine Learning. New York: ACM, 2020: 1597–1607. DOI: 10.5555/3524938.3525087.
- [30] ARORA S, KHANDEPARKAR H, KHODAK M, *et al.* A Theoretical Analysis of Contrastive Unsupervised Rep-

- resentation Learning[EB/OL]. arXiv Preprint: 1902.09229, 2018. <https://doi.org/10.48550/arXiv.1902.09229>.
- [31] ZHANG M Q, WU S, YU X L, *et al.* Dynamic Graph Neural Networks for Sequential Recommendation[J]. *IEEE Trans Knowl Data Eng*, 2023, **35**(5): 4741–4753. DOI: 10.1109/TKDE.2022.3151618.
- [32] YU J L, XIA X, CHEN T, *et al.* XSimGCL: Towards Extremely Simple Graph Contrastive Learning for Recommendation[J]. *IEEE Trans Knowl Data Eng*, 2024, **36**(2): 913–926. DOI: 10.1109/TKDE.2023.3288135.
- [33] WU J C, WANG X, FENG F L, *et al.* Self-supervised Graph Learning for Recommendation[C]//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2021: 726–735. DOI: 10.1145/3404835.3462862.
- [34] WANG J L, DING K Z, HONG L J, *et al.* Next-item Recommendation with Sequential Hypergraphs[C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2020: 1101–1110. DOI: 10.1145/3397271.3401133.
- [35] WANG M R, REN P J, MEI L, *et al.* A Collaborative Session-based Recommendation Approach with Parallel Memory Modules[C]//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2019: 345–354. DOI: 10.1145/3331184.3331210.
- [36] SONG W P, XIAO Z P, WANG Y F, *et al.* Session-based Social Recommendation via Dynamic Graph Attention Networks[C]//Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining. New York: ACM, 2019: 555–563. DOI: 10.1145/3289600.3290989.
- [37] 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. New York: ACM, 2020: 1893–1902. DOI: 10.1145/3340531.3411954.
- [38] XIE X, SUN F, LIU Z Y, *et al.* Contrastive Learning for Sequential Recommendation[C]//2022 IEEE 38th International Conference on Data Engineering (ICDE). Kuala Lumpur: IEEE, 2022: 1259–1273. DOI: 10.1109/ICDE53745.2022.00099.
- [39] XIA X, YIN H Z, YU J L, *et al.* Self-supervised Hypergraph Convolutional Networks for Session-based Recommendation[J]. *Proc AAAI Conf Artif Intell*, 2021, **35**(5): 4503–4511. DOI: 10.1609/aaai.v35i5.16578.
- [40] LIN Z H, TIAN C X, HOU Y P, *et al.* Improving Graph Collaborative Filtering with Neighborhood-enriched Contrastive Learning[C]//Proceedings of the ACM Web Conference 2022. New York: ACM, 2022: 2320–2329. DOI: 10.1145/3485447.3512104.
- [41] QIU R H, HUANG Z, YIN H Z, *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. New York: ACM, 2022: 813–823. DOI: 10.1145/3488560.3498433.
- [42] ZHANG Y X, LIU Y, XU Y H, *et al.* Enhancing Sequential Recommendation with Graph Contrastive Learning[C]//Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2022: 2398–2405. DOI: 10.24963/ijcai.2022/333.
- [43] YU J L, YIN H Z, XIA X, *et al.* Are Graph Augmentations Necessary? : Simple Graph Contrastive Learning for Recommendation[C]//Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2022: 1294–1303. DOI: 10.1145/3477495.3531937.
- [44] CHEN Y J, LIU Z W, LI J, *et al.* Intent Contrastive Learning for Sequential Recommendation[C]//Proceedings of the ACM Web Conference 2022. New York: ACM, 2022: 2172–2182. DOI: 10.1145/3485447.3512090.
- [45] WANG Z Y, LIU H Y, WEI W, *et al.* Multi-level Contrastive Learning Framework for Sequential Recommendation[C]//Proceedings of the 31st ACM International Conference on Information & Knowledge Management. New York: ACM, 2022: 2098–2107. DOI: 10.1145/3511808.3557404.
- [46] LIN G Y, GAO C, LI Y F, *et al.* Dual Contrastive Network for Sequential Recommendation[C]//Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2022: 2686–2691. DOI: 10.1145/3477495.3531918.
- [47] MA C, KANG P, LIU X. Hierarchical Gating Networks for Sequential Recommendation[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2019: 825–833. DOI: 10.1145/3292500.3330984.
- [48] LI J C, WANG Y J, MCAULEY J. Time Interval Aware Self-attention for Sequential Recommendation[C]//Proceedings of the 13th International Conference on Web Search and Data Mining. New York: ACM, 2020: 322–330. DOI: 10.1145/3336191.3371786.