

面向会话型推荐的特征保留图学习方法

陈洪辉¹, 潘志强¹, 田双喜^{1, 2}

¹ (国防科技大学 信息系统工程重点实验室, 长沙 410073)

² (湖南财政经济学院, 长沙 410074)

E-mail: tsx_0726@nudt.edu.cn

摘要: 本文针对基于会话的推荐提出了基于软归一化约束的特征保留图学习方法。首先将每个会话转化为会话图, 然后提出特征保留的图神经网络来在进行信息传播的同时保留物品各自的原始特征, 从而学习到准确的物品表示。另外, 在利用交叉熵作为主要损失函数的基础上, 设计了软归一化约束方法来缓解由物品流行度不同导致的长尾问题。在两个基准数据集, 即 Diginetica 和 Gowalla 上进行的实验证明, 所提方法能够有效地提高基于会话的推荐在 Recall@20 和 MRR@20 指标上的表现。

关键词: 推荐系统; 基于会话的推荐; 图神经网络; 软归一化约束

中图分类号: TP391.3

文献标识码: A

Feature Preserving Graph Learning for Session-based Recommendation

CHEN Hong-hui¹, PAN Zhi-qiang¹, TIAN Shuang-xi^{1,2}

¹ (Science and Technology on Information Systems Engineering Laboratory, National University of Defense Technology, Changsha 410073, China)

² (Hunan University of Finance and Economics, Changsha 410074, China)

Abstract: This paper proposes a feature preserving graph learning method with soft normalization constraint. First, the session is transformed into a session graph, and a feature preserving graph neural network is proposed to preserve the original item features while conducting information propagation, so as to learn accurate item representations. In addition, on the basis of utilizing the cross-entropy as the main loss function, a soft normalization constraint is designed to alleviate the long-tail problem caused by the various item popularity. Experiments conducted on two benchmark datasets, i.e., Diginetica and Gowalla, demonstrate that the proposal can effectively improve the performance in terms of Recall@20 and MRR@20 on the session-based recommendation task.

Key words: recommender systems; session-based recommendation; graph neural networks; soft normalization constraint

1 引言

随着网络上信息的快速增长, 用户在获取个性化需求时经常会面临信息冗余的问题^[1-3]。作为一种有效过滤无用信息的方法, 推荐系统能够根据用户的历史交互信息探测用户兴趣, 从而为用户产生推荐来满足其信息需求^[4]。然而, 在一些实际场景下, 如用户刚注册或者用户未登录时, 用户的长期历史交互并不可用^[5,6]。因此, 基于会话的推荐被提出, 旨在通过用户近期有限数目的交互行为来探测用户意图并预测其在下一时刻的行为^[5]。

当前的方法主要通过循环神经网络 (Recurrent Neural Network, RNN) 或图神经网络 (Graph Neural Network, GNN) 来建模会话内物品序列中的时序关系或者物品间的传递关系^[5,7-10]。例如, Hidasi 等人提出 GRU4REC, 利用门控循环单

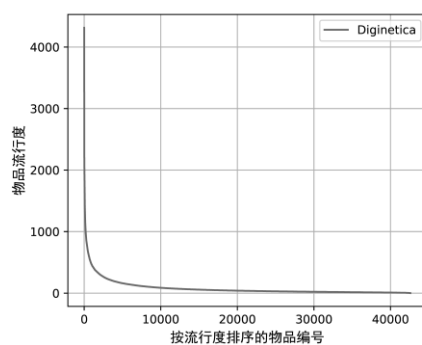


图 1 长尾问题, 其中物品编号按照其流行度进行排序
Fig. 1 The long-tail problem, where the item IDs are ranked according to their popularity

元 (Gated Recurrent Unit, GRU) 来捕获物品间的时序依赖关系^[5]。另外, Wu 等人首次使用门控图神经网络 (Gated Graph Neural Network, GGNN), 通过将物品序列转化为会话图来建模物品间的传递关系^[8]。另外, 注意力机制也被广泛地应用, 可以单独使用^[11]也可以和其他方法如 RNN、GNN 等共同使用, 通过为不同物品分配不同的重要性来捕获用户的主要意图^[12, 13]。

尽管上面提到的方法已经实现了不错的效果, 但仍然存在一些问题: (1) 基于时序的方法无法建模到物品之间复杂的传递关系; (2) 基于图网络的方法在进行消息传播后会导致过度平滑问题, 使得同一会话内的物品学习到的嵌入向量高度相似, 从而失去物品本身独有的特征; (3) 如图 1 所示, 不同物品的流行度差别较大, 会在模型训练时引入偏差信息, 使得物品嵌入向量的 L2 范数随着流行度的升高而增大, 对物品推荐的生成产生干扰^[14]。

因此, 为了解决上述的问题, 本文提出了基于软归一化约束的特征保留图学习 (Feature Preserving Graph Learning with Soft Normalization Constraint, FPGL-SMC) 方法。具体来说, 本文首先将会话中的物品序列转化为会话图, 使用输出和输入矩阵来表示。然后, 本文设计了特征保留的图神经网络 (Feature Preserving Graph Neural Network, FPGNN) 来学习物品的准确表示, 并根据学到的表示获得用户在当前时刻的偏好。最后, 进行物品预测, 采用交叉熵 (Cross Entropy, CE) 函数作为主损失函数进行训练, 并设计了软归一化约束 (Soft Normalization Constraint, SMC) 来解决物品流行度不同导致的长尾问题。

总的来说, 本文中的贡献可以总结如下: (1) 本文设计了特征保留的图神经网络, 能够有效缓解传统神经网络中存在的过度平滑问题; (2) 本文在交叉熵损失函数的基础上引入软归一化约束进行联合训练, 能够有效解决长尾问题; (3) 本文在两个公开可用的数据集, 即 Diginetica 和 Gowalla 上进行了大量的实验, 实验结果证明所提出的 FPGL-SMC 方法能够在 Recall@20 和 MRR@20 两个指标上实现当前最优的效果。

2 基于软归一化约束的特征保留图学习方法

首先, 对基于会话的推荐任务进行数学化的描述。用 $V = \{v_1, v_2, \dots, v_{|V|}\}$ 来表示所有会话中独特的物品, 其中 $|V|$ 是所有独特物品的数目。给定包含 n 个时序物品的会话 $S = \{v_1, v_2, \dots, v_n\}$, 其中 $v_t \in V$ 是会话中的第 t 个物品。基于会话的推荐目的是预测用户在下一个时刻将会点击的物品, 即 v_{n+1} 。具体地说, 通过输出在所有物品上的概率 $\hat{\mathbf{y}} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_{|V|}\}$, 其中 $\hat{y}_i \in \hat{\mathbf{y}}$ 表示点击物品 v_i 的概率。最后, 预测分数最高的 K 个物品将会被推荐给用户。图 2 展示了本文提出的 FPGL-SMC 方法框架, 主要包括会话图构造、特征保

留的物品表示学习以及模型训练三个模块。

2.1 会话图构造

对于每个会话 $S = \{v_1, v_2, \dots, v_n\}$, 本文建立一个会话图来表示其中物品之间的传递关系。具体来说, 每个会话被表示为 $G_s = \{\mathcal{V}, \mathcal{E}\}$, 其中 $\mathcal{V} = \{x_1, x_2, \dots, x_m\}$ 表示会话中所有独特的节点。值得注意的是, 由于用户在一个会话中可能会与部分物品重复进行交互, 因此独特节点的数目 m 小于等于会话中交互物品的总数 n 。 \mathcal{E} 是会话图中的边, 每条边 $(x_i, x_j) \in \mathcal{E}$ 表示用户在会话中点击 x_i 之后点击了 x_j 。然后, 本文使用邻接矩阵 $\mathbf{A}_s = \{\mathbf{A}_s^{out}, \mathbf{A}_s^{in}\}$ 来表示会话中物品之间的连接关系, 包括输出矩阵 \mathbf{A}_s^{out} 和输入矩阵 \mathbf{A}_s^{in} , 分别表示会话图中加权的输出和输入连接。例如, 给定会话 $S = \{x_3, x_2, x_4, x_5, x_4, x_6\}$, 可以如图 3 中构造输出和输入矩阵。

2.2 特征保留的物品表示学习

在构造完会话图后, 对于每个在第 l 层图网络的节点 x_i , 首先使用输出矩阵和输入矩阵获取来自邻居节点的信息如下:

$$\begin{aligned} \mathbf{a}_i^l &= \text{Concat}(\mathbf{A}_i^l[\mathbf{x}_1^{l-1}, \mathbf{x}_2^{l-1}, \dots, \mathbf{x}_m^{l-1}]^T \mathbf{W}^l + \mathbf{b}^l), \\ \mathbf{A}_i^l[\mathbf{x}_1^{l-1}, \mathbf{x}_2^{l-1}, \dots, \mathbf{x}_m^{l-1}]^T \mathbf{W}^o + \mathbf{b}^o) \end{aligned} \quad (1)$$

其中 $\mathbf{x}_i^0 \in R^d$ 是物品 x_i 经过嵌入层产生的物品表示向量, $\mathbf{A}_i^o, \mathbf{A}_i^l \in R^{1 \times m}$ 是对应于节点 x_i 的输出和输入权重, 即输出矩阵和输入矩阵的第 i 行。 $\mathbf{W}^o, \mathbf{W}^l \in R^{d \times d}$ 分别是输出和输入边的可训练参数, $\mathbf{b}^o, \mathbf{b}^l \in R^d$ 是偏差向量。因此, 可以获得 $\mathbf{a}_i^l \in R^{1 \times 2d}$ 来表示传播到节点 x_i 的邻居信息。

在获取到邻居信息 $\mathbf{a}_i^l \in R^{1 \times 2d}$ 后, 本文设计了特征保留的图神经网络来生成物品表示, 其结构如图 4 所示。具体来说, 给定节点 x_i 在上一层图网络的隐藏向量 \mathbf{h}_i^{l-1} 、初始嵌入向量 \mathbf{h}_i^0 以及来自第 l 层的邻居信息 $\mathbf{a}_i^l \in R^{1 \times 2d}$, 本文按照如下方式更新节点 x_i 的表示:

$$\begin{aligned} \mathbf{z}_i^l &= \sigma(\mathbf{W}_z \mathbf{a}_i^l + \mathbf{V}_z \mathbf{h}_i^0 + \mathbf{U}_z \mathbf{h}_i^{l-1}) \\ \mathbf{r}_i^l &= \sigma(\mathbf{W}_r \mathbf{a}_i^l + \mathbf{V}_r \mathbf{h}_i^0 + \mathbf{U}_r \mathbf{h}_i^{l-1}) \\ \tilde{\mathbf{h}}_i^l &= \tanh(\mathbf{W}_h \mathbf{a}_i^l + \mathbf{V}_h \mathbf{h}_i^0 + \mathbf{U}_h (\mathbf{r}_i^l \odot \mathbf{h}_i^{l-1})) \\ \tilde{\mathbf{h}}_i^l &= (1 - \mathbf{z}_i^l) \odot \mathbf{h}_i^{l-1} + \mathbf{z}_i^l \odot \tilde{\mathbf{h}}_i^l \\ \mathbf{h}_i^l &= \mathbf{W}_c[\mathbf{h}_i^0, \tilde{\mathbf{h}}_i^l] \end{aligned} \quad (2)$$

其中, $\mathbf{W}_z, \mathbf{W}_r, \mathbf{W}_h \in R^{d \times 2d}$ 和 $\mathbf{V}_z, \mathbf{V}_r, \mathbf{V}_h, \mathbf{U}_z, \mathbf{U}_r, \mathbf{U}_h \in R^{d \times d}$ 是网络中的可训练参数, σ 表示激活函数 sigmoid, \odot 是元素级别的乘法。 \mathbf{z}_i^l 和 \mathbf{r}_i^l 是更新门和重置门, 分别控制前置状态 \mathbf{h}_i^{l-1} 中的多少信息应该被保留以及多少信息应该被写入到候选激活状态 $\tilde{\mathbf{h}}_i^l$ 中。通过这种方式, 来自邻居的协同信息就可以在 FPGNN 中进行传播。同时, 在计算更新门、重置门以及生成最终表示时加入了节点初始状态 \mathbf{h}_i^0 , 可以保证在进行信息传播的同时, 保存每个节点自身的独特

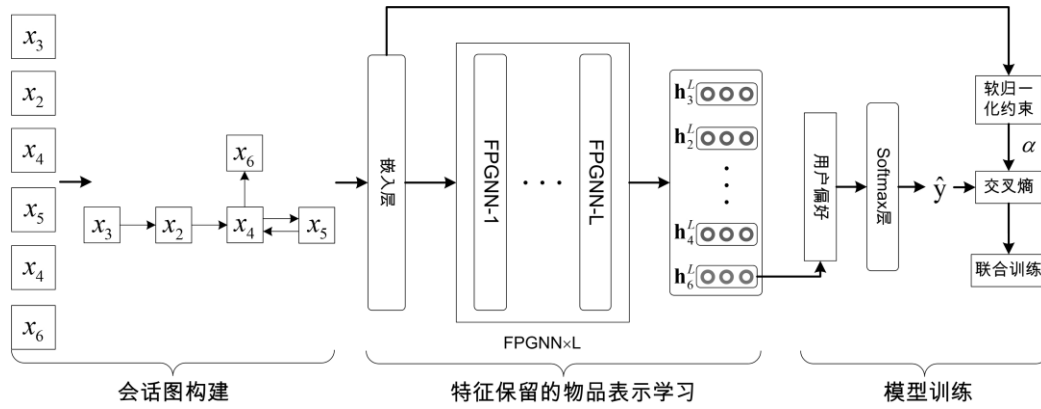


图 2 基于软归一化约束的特征保留图学习方法框架

Fig.2 The framework of the feature preserving graph learning method with soft normalization constraint

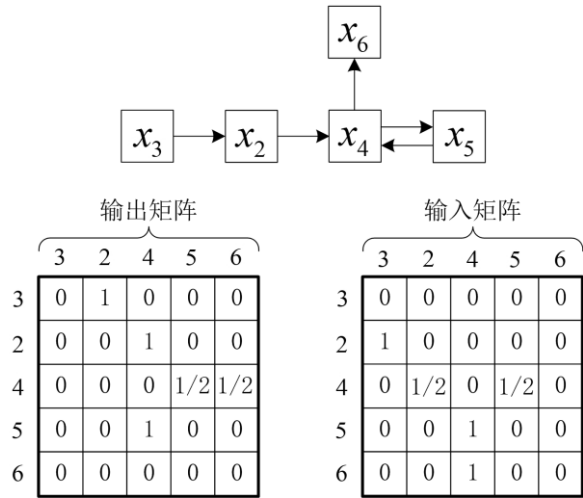


图 3 输出矩阵和输入矩阵构建案例

Fig. 3 An example for constructing the outgoing and incoming matrices

特征,有效避免了传统图网络^[15]中会面临的过度平滑问题。

2.3 模型训练

在使用 L 层 FPGNN 学习到会话中物品的表示之后,采用会话中最后交互的物品表示作为用户偏好表示,即 $\mathbf{z} = \mathbf{h}_n^L$, 这样可以捕获到用户的实时兴趣偏好。

接下来,使用获取的用户偏好来计算 V 中所有候选物品被点击的概率,从而产生推荐。具体地说,通过将用户偏好表示和每个物品的嵌入向量相乘,来计算物品集 V 中每个物品的预测分数:

$$\tilde{\mathbf{y}} = \mathbf{z}^T \mathbf{v}_i \quad (3)$$

然后,使用 Softmax 函数来对计算得到的分数进行归一化:

$$\hat{\mathbf{y}} = \text{Softmax}(\tilde{\mathbf{y}}) \quad (4)$$

最后, $\hat{\mathbf{y}}$ 中分数最高的 K 个物品将会被推荐给用户。为了训练本文所提出的模型,本文采用交叉熵作为主要的优化函数来学习模型参数:

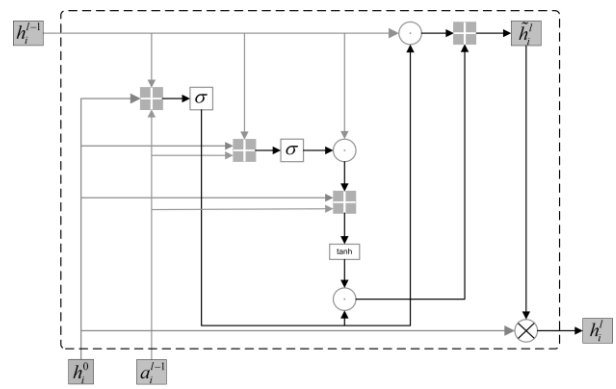


图 4 特征保留的图神经网络结构图

Fig. 4 The structure of the feature preserving graph neural network

$$L_{CE} = -\sum_{i=1}^{|V|} \mathbf{y}_i \log(\hat{\mathbf{y}}_i) + (1 - \mathbf{y}_i) \log(1 - \hat{\mathbf{y}}_i) \quad (5)$$

其中, \mathbf{y}_i 和 $\hat{\mathbf{y}}_i$ 分别是用户的真实偏好分布和预测偏好分布中的第 i 个元素。也就是说,如果物品 v_i 是当前会话对应的目标物品,那么 $\mathbf{y}_i = 1$; 否则 $\mathbf{y}_i = 0$ 。

另外,本文设计了软归一化约束来避免物品流行度不同导致的长尾问题,如下所示:

$$L_{SMC} = \frac{1}{2} \sum_{i=1}^{|V|} (\|\mathbf{v}_i\|_2 - \beta)^2 \quad (6)$$

其中, β 是可以调节的超参数。最后,将交叉熵损失和软归一化约束损失使用相加进行联合训练:

$$L = L_{CE} + \alpha L_{SMC} \quad (7)$$

其中, α 也是可调节的超参数。最后,使用反向传播算法^[16]来训练所提出的 FPGL-SMC 模型。

3 实验设计

3.1 研究问题

本文通过解决以下研究问题来检验所提出的 FPGL-SMC 方法的有效性:(1) FPGL-SMC 能否在基于会话的推荐任务上实现比当前最优的基准模型更好的效果?(2) FPGL-SMC 方法中的各个模块对推荐效果的贡献是怎样的?(3) 本文

设计的软归一化约束方法能否缓解基于会话的推荐中存在的长尾问题?

3.2 数据集

参照[9, 17], 本文采用了两个基准数据集来进行实验, 即 Diginetica 和 Gowalla. Diginetica 是一个由 CIKM Cup 2016 发布的电子商务数据集. 参照[9], 本文采用其中的交易数据. Gowalla 是一个时间跨度由 2009 年 2 月到 2010 年 10 月的签到数据集. 本文保留了最流行的 30,000 个位置进行实验. Diginetica 中每个用户交互的物品和 Gowalla 中每个用户在 24 小时内的交互记录被视为一个会话. 另外, 两个数据集中, 出现次数少于 5 的物品被过滤掉, 长度为 1 的会话也被移除. Diginetica 中最后一周的会话和 Gowalla 中最后 20% 的会话分别被分离出来作为测试集. 另外, 本文采用训练集中最后 20% 的部分作为验证集来调节 FPGL-SMC 的超参数. 预处理之后的数据统计信息见表 1.

表 1 数据集统计信息

统计信息	Diginetica	Gowalla
点击数目	981,620	1,122,788
训练会话数目	716,835	675,561
测试会话数目	60,194	155,332
物品数目	42,596	29,510
平均会话长度	4.80	3.85

3.3 模型总结

为了验证本文所提出方法的有效性, 将 FPGL-SMC 与下列有竞争力的竞争方法进行比较: (1) Item-KNN^[18] 依赖于余弦相似度来推荐与当前会话内物品相似的候选物品; (2) FPMC^[19] 是一个混合方法, 采用马尔可夫链来捕获时序信号; (3) NextItNet^[20] 是一种基于卷积神经网络 (Convolutional Neural Networks, CNN) 的方法, 使用扩张的卷积而不是有损的池化操作来增加接收域; (4) NARM^[12] 是一种基于 RNN 的方法, 采用 GRU 来捕获时序信号, 并利用注意力机制强调用户的主要意图; (5) FGNN^[13] 设计了一种加权的图注意力网络来计算会话图中的信息流, 并使用读出函数来生成用户偏好; (6) SR-GNN^[8] 提出将每个会话转化为会话图, 并且采用 GGNN 来捕获物品间的传递关系; (7) GC-SAN^[21] 在 SR-GNN 的基础上采用自注意力机制来捕获物品之间的远距离依赖关系; (8) LESSR^[9] 使用无损边顺序保存聚合和快捷图注意力方法来解决传统图神经网络中存在的信息损失问题; (9) GCE-GNN^[22] 设计了一个统一的模型来同时利用全局图和会话图中的全局级和会话级的物品传递关系。

3.4 实验设置

参照[8, 12], 本文采用一种序列分割方法来进行训练样本和测试样本的数据增强, 即为 $S = \{v_1, v_2, \dots, v_n\}$ 会话生成 $([v_1], v_2), ([v_1, v_2], v_3), \dots, ([v_1, v_2, \dots, v_{n-1}], v_n)$. 然后, 单独建模每个增强后的序列. FPGL-SMC 和基准模型的超参数通过验证集进行调节. 具体来说, 采用网格搜索来寻找每个模型的最优参数, 其中嵌入向量在 $\{50, 100, 150, 200\}$ 中进行搜索, 学习率和 L2 正则化分别在 $\{1e^{-5}, 1e^{-4}, 1e^{-3}, 1e^{-2}\}$ 和

$\{1e^{-6}, 1e^{-5}, 1e^{-4}, 1e^{-3}, 1e^{-2}\}$ 中进行调节. 考虑到深度图网络会导致严重的过拟合问题, 参照[8], FPGL-SMC 的图网络层数设置为 1. 另外, 本文采用 Adam 优化器来对模型内的参数进行训练, 学习率初始化为 $1e^{-3}$ 并每 2 个迭代次数衰减 0.1, 另外批大小设为 100. 对于 FPGL-SMC, α 和 β 分别在 $\{1, 5, 10, 20, 50\}$ 和 $\{0.5, 1.0, 1.5, 2.0, 2.5\}$ 内进行调节. 所有的参数均使用高斯分布进行初始化, 其中均值为 0 且标准差为 0.1.

3.5 评价指标

本文采用 Recall@20 和 MRR@20 来衡量推荐效果.

Recall@20: Recall@20 分数衡量目标物品是否被包括在物品推荐列表的前 20 个位置中:

$$\text{Recall} @ K = \frac{n_{hit}}{N} \quad (8)$$

其中, N 是数据集中测试序列的数目, n_{hit} 是目标物品被包含在排序列表前 20 个物品中的测试样本数目.

MRR@20 (Mean Reciprocal Rank): MRR@20 考虑了目标物品在物品推荐列表中的位置. 如果目标物品不在推荐列表的前 20 个位置中, MRR@20 被设为 0, 否则按照如下方式进行计算:

$$\text{MRR} @ 20 = \frac{1}{N} \sum_{v_{target} \in S_{test}} \frac{1}{\text{Rank}(v_{target})} \quad (9)$$

其中, v_{target} 是目标物品, $\text{Rank}(v_{target})$ 是目标物品在推荐列表中的位置. 相较于 Recall@20, MRR@20 是一个将目标物品位置考虑进来的归一化指标.

4 实验结果和讨论

4.1 与对比模型比较

为了回答研究问题 1, 本文将 FPGL-SMC 的 Recall@20 和 MRR@20 得分和基准模型进行比较. 实验结果见表 2. 首先, 如表 2 所示, 基于神经网络的模型, 如 NARM 和 NextItNet, 普遍优于传统方法, 即 Item-KNN 和 FPMC. 另外, Item-KNN 在两个数据集上的所有情况下均表现优于 FPMC. 这可能是由于 FPMC 中的马尔科夫链只能捕获邻近物品之间的时序依赖, 而忽略了整个物品序列中的时序关系.

另外, 神经网络模型中, 基于 RNN 的方法, 即 NARM, 比基于 CNN 的方法, 即 NextItNext, 展示出了更好的效果. 这可能是由于相较于 CNN, RNN 更适用于建模会话中物品间的时序关联. 另外, 基于 GNN 的方法实现了在基准模型中最优的效果, 表示建模序列中物品之间的传递关系能够帮助提高基于会话的推荐效果. 另外, 可以发现, 在 SR-GNN 的基础上, 通过建模物品之间长距离的依赖关系, GC-SAN 实现了更好的效果. 另外, 通过解决有损会话编码问题, LESSR 在 Diginetica 和 Gowalla 两个数据集上的大多数情况下都实现了最好的效果. 对于基准模型 GCE-GNN, 其同时考虑了全局和局部的物品传递关系, 在 Gowalla 数据集上的 Recall@20 指标上实现了最好的效果, 但在其他情况下表现

表 2 模型表现

Table 2 Model Performance

方法	Diginetica		Gowalla	
	Recall@20	MRR@20	Recall@20	MRR@20
Item-KNN	39.51	11.22	38.60	16.66
FPMC	28.50	7.67	29.91	11.45
NextIttNet	45.41	15.19	45.15	21.26
NARM	49.80	16.57	50.07	23.92
FGNN	50.03	17.01	50.06	24.12
SR-GNN	50.81	17.31	50.32	24.25
GC-SAN	50.90	17.63	50.68	24.67
LESSR	<u>51.71</u>	<u>18.15</u>	51.34	<u>25.49</u>
GCE-GNN	51.66	17.53	<u>51.53</u>	23.52
FPGL-SMC	52.79	18.72	51.75	25.84
提高比例	2.09	3.14	0.43	1.37

表 3 消融实验

Table 3 Ablation Study

方法	Diginetica		Gowalla	
	Recall@20	MRR@20	Recall@20	MRR@20
GL-SMC	50.16	16.46	50.26	22.86
FPGL	51.93	18.09	50.83	25.08
FPGL-SMC	52.79	18.72	51.65	25.74

不佳。这可能是由于 GCE-GNN 无法区分出全局图和局部图中的无关物品，导致容易引入较多误差。

接下来，关注于本文所提出的 FPGL-SMC 模型。总的来说，FPGL-SMC 在两个数据集上的 Recall@20 和 MRR@20 结果都是最高，说明了所提方法在基于会话的推荐任务上的有效性。另外，可以看到 FPGL-SMC 相较于最好的基准模型，即 LESSR 和 GCE-GNN，在 Diginetica 数据集上的 Recall@20 和 MRR@20 指标上的提升分别为 2.09% 和 3.14%，在 Gowalla 数据集上相对应的提升分别为 0.43% 和 1.37%。可以明显看到，在两个数据集上，FPGL-SMC 在 MRR@20 指标上所实现的提升相较于 Recall@20 指标都更加明显。这说明本文所提出的 FPGL-SMC 方法在进行推荐时能够更有效地将目标物品排在推荐列表中更靠前的位置。

4.2 消融实验

为了验证所提出的 FPGL-SMC 模型中各个模块的有效性，本文将 FPGL-SMC 与以下两个变体模型进行比较：（1）GL-SMC：使用 GGNN 来代替 FPGL-SMC 中的 FPGNN 模块；（2）FGPL：从 FPGL-SMC 模型中移除掉 SMC 模块。实验结果展示在表 3 中。

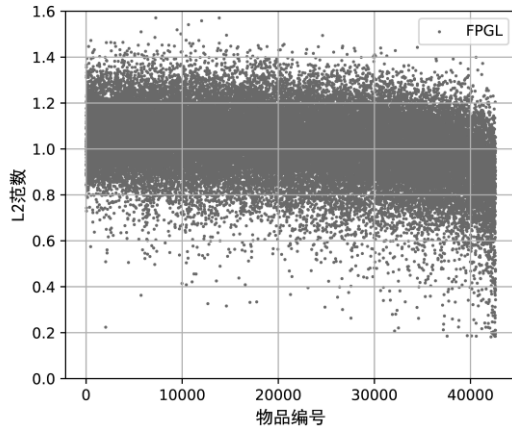
从表 3 中，可以看到，变体模型 GL-SMC 和 FPGL 在两个数据集上的 Recall@20 和 MRR@20 指标上的推荐效果都低于 FPGL-SMC，证明了本文所设计的特征保留的图神经网络和软

归一化约束的有效性。另外，可以观察到，GL-SMC 在两个数据集上所有情况下的效果都低于 FPGL，这说明相较于物品流行度带来的偏差影响，传统图神经网络中存在的过度平滑问题对推荐准确率的影响更大，尤其是在 MRR@20 指标上。

4.3 软归一化约束有效性验证

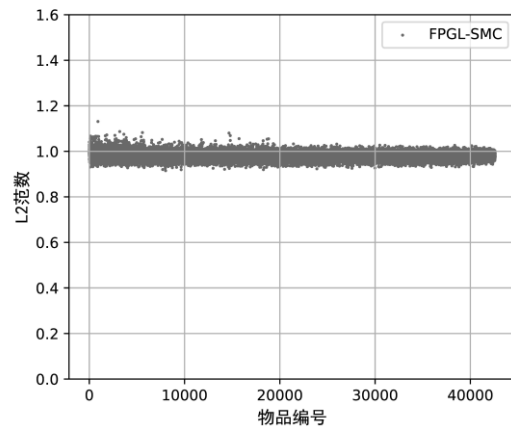
为了验证所提出的软归一化约束的有效性，本文比较了 FPGL-SMC 和其变体模型 FPGL 学习到的物品嵌入表示的 L2 范数，结果见图 5。从图 5(a) 中可以看到对于没有使用软归一化约束的 FPGL 模型，其所学习到的物品嵌入向量的 L2 范数浮动较大。这是由于物品的流行度偏差给模型训练带来的负面影响所导致的，会使得流行度较大的物品其嵌入向量的 L2 范数也更大，相较于流行度较小的物品更容易被推荐给用户。另外，从图 5(b) 中可以看到，通过使用所设计的软归一化约束，FPGL-SMC 模型能够将物品嵌入向量的 L2 范数约束到相近的值，从而避免了物品流行度偏差在生成推荐时带来的不利影响，使得 FPGL-SMC 能够产生比 FPGL 更加精准的推荐。

另外，为了分析参数 α 和 β 对模型效果的影响，本文将分别在 [1, 5, 10, 20, 50] 和 [0.5, 1.0, 1.5, 2.0, 2.5] 范围内进行调节，来观察 FPGL-SMC 模型的效果变化。从表 4 和表 5 可以看到，当 $\alpha = 20$ 且 $\beta = 1.0$ 时，FPGL-SMC 在 Recall@20 和 MRR@20 两个指标上都实现了最好的效果。并



(a) FPGL 模型学习到的物品嵌入 L2 范数

(a) The L2 norm of item embeddings learned by FPGL



(b) FPGL-SMC 模型学习到的物品嵌入 L2 范数

(b) The L2 norm of item embeddings learned by FPGL-SMC

图 5 FPGL 和 FPGL-SMC 学习到的物品嵌入表示 L2 范数比较

Fig. 5 The L2 norm comparison of item embeddings learned by FPGL and FPGL-SMC

表 4 α 和 β 对 FPGL-SMC 在 Recall@20 指标上的影响Table 4 The influence of α and β on FPGL-SMC in terms of Recall@20

参数	1	5	10	20	50
0.5	51.81	52.19	52.26	52.41	51.95
1.0	51.77	52.68	52.34	51.75	51.16
1.5	52.19	52.70	52.23	51.61	50.77
2.0	52.11	52.79	52.06	51.46	50.69
2.5	52.21	52.74	52.14	51.22	50.61

表 5 α 和 β 对 FPGL-SMC 在 MRR@20 指标上的影响Table 5 The influence of α and β on FPGL-SMC in terms of MRR@20

参数	1	5	10	20	50
0.5	17.86	18.27	18.50	18.48	18.25
1.0	17.78	18.60	18.48	18.28	18.17
1.5	17.88	18.60	18.46	18.31	18.03
2.0	17.82	18.72	18.60	18.24	18.09
2.5	18.00	18.68	18.47	18.29	18.14

且,在固定 α 时,大多数情况下随着 β 的增大,模型效果普遍升高,但在部分情况下 β 值较大时,效果会略微下降。另外,在固定 β 时,随着 α 的增大,在所有情况下FPGL-SMC的效果均是先升高再降低。这说明较小的 α 不足以使得SMC模块克服流行度偏差带来的不利影响,然而较大的 α 会干扰模型本身的表示学习能力。因此, $\alpha=20$ 对FPGL-SMC模型来说较为合适。

5 结论

本文针对基于会话的推荐,提出了一种基于软归一化约束的特征保留图学习方法,通过设计特征保留的图神经网络解决传统图神经网络中的过度平滑问题,并提出软归一化约束方法缓解由物品流行度不同导致的长尾问题。实验证明,本文方法能在基于会话的推荐任务上实现目前最优的效果。

References:

- [1] Adomavicius G and Tuzhilin A. 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions [J]. IEEE Trans. Knowl. Data Eng,2005,17(6):734-749.
- [2] WANG J Y, Wu G D, FAN W C, TU L J, LI J X. A survey on personalized recommendation of user behavior sequence[J]. Journal of Chinese Mini-Micro Computer Systems, 2022,43(05):921-935.
- [3] Chen W Y, Cai F, Chen H H, & de Rijke M. A dynamic co-attention network for session-based recommendation [C]//Proceedings of the 28th ACM International Conference on Information and Knowledge Management,2019:1461-1470.

- [4] 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,2017:173–182.
- [5] Hidasi B, Karatzoglou A, Baltrunas L, et al. Session-based recommendations with recurrent neural networks [C]//Proceedings of 4th International Conference on Learning Representations,2016.
- [6] SUN S J, GUO G, QIAN M W. A deep learning model on personalized sequential recommendation with context information fusion[J]. Journal of Chinese Mini-Micro Computer Systems,2021,42(06):1121–1128.
- [7] 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,2019:345–354.
- [8] WU S, TANG Y Y, ZHU Y Q, et al. Session-based recommendation with graph neural networks [C]//Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence,2019:346–353.
- [9] CHEN T W, Wong R C W. Handling information loss of graph neural networks for session-based recommendation [C]//Proceedings of the 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining,2020:1172–1180.
- [10] ZENG Y F, MOU Q L, ZHOU L, et al. Session-based Recommender Systems with Graph Representation Learning [J]. Journal of Computer Research and Development, 2020,57(03):590–603.
- [11] LIU Q, ZENG Y F, Mokhosi R, et al. STAMP: Short-term attention/memory priority model for session-based recommendation [C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,2018:1831–1839.
- [12] LI J, REN P J, CHEN Z M, et al. Neural attentive session-based recommendation [C]//Proceedings of the 26th ACM on Conference on Information and Knowledge Management,2017:1419–1428.
- [13] QIU R H, LI J J, HUANG Z, et al. Rethinking the item order in session-based recommendation with graph neural networks [C]//Proceedings of the 28th ACM on Conference on Information and Knowledge Management,2019:579–588.
- [14] Gupta, P., Garg, D., Malhotra, P., Vig, L., & Shroff, G. (2019). NISER: Normalized item and session representations to handle popularity bias [EB/OL], <https://arxiv.org/abs/1909.04276>,2019.
- [15] LI Y J, Tarlow D, Brockschmidt M, et al. Gated graph sequence neural networks [C]//Proceedings of the 4th International Conference on Learning Representations,2016.
- [16] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors [J]. Nature, 1986,323(6088):533–536.
- [17] GUO L, YIN H Z, WANG Q Y, et al. Streaming session-based recommendation [C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,2019:1569–1577.
- [18] Davidson J, Liebald B, Liu JN, et al. The YouTube video recommendation system [C]//Proceedings of the 2010 ACM Conference on Recommender Systems,2010:293–296.
- [19] Rendle S, Freudenthaler C, and Thieme L S. Factorizing personalized Markov chains for next-basket recommendation [C]//Proceedings of the 19th International Conference on World Wide Web,2010:811–820.
- [20] YUAN F J, Karatzoglou A, Arapakis I, et al. A simple convolutional generative network for next item recommendation [C]//Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining,2019:582–590.
- [21] XU C F, ZHAO P P, LIU Y C, et al. Graph contextualized self-attention network for session-based recommendation [C]//Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence,2019:3940–3946.
- [22] WANG Z Y, Wei W, CONG G, et al. Global context enhanced graph neural networks for session-based recommendation [C]//Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval,2020:169–178.

附中文参考文献

- [2] 汪菁瑶, 吴国栋, 范维成, 等. 用户行为序列个性化推荐研究综述[J]. 小型微型计算机系统, 2022, 43 (05) :921–935.
- [6] 孙淑娟, 过弋, 钱梦薇. 融合上下文信息的个性化序列推荐深度学习模型 [J]. 小型微型计算机系统, 2021, 42 (06) :1121–1128.
- [10] 曾义夫, 牟其林, 周乐, 等. 基于图表示学习的会话感知推荐模型 [J]. 计算机研究与发展, 2020, 57 (03) :590 – 603.