

融合多源图信息的图神经网络会话推荐算法

林 炜, 吴开军

(上海海洋大学信息学院, 上海 201306)

摘 要: 现有的基于图神经网络的会话推荐算法通过将会话序列构筑为图形结构捕捉项目转换关系, 能够有效提高推荐性能. 然而多数图神经网络及其改进模型在建模会话时仅考虑会话序列中项目的单次转换关系, 忽略了会话中包含的大量有效信息, 同时缺乏对项目间隐藏关联性的分析. 因此提出融合多源图信息的图神经网络会话推荐算法. 将用户重复行为信息, 项目内容关联信息纳入到会话图建模过程当中, 有效提取项目更深层次的复杂转换关系, 并通过线性转换进行聚合. 此外采用外部注意力机制辅助获取会话序列项目隐藏关联信息, 使得生成的会话向量在推荐过程中更加精确. 在真实数据集 Yoochoose 和 Diginetica 上进行实验, 实验结果表明该模型优于基准模型, 特别地, 相较于 SR-GNN 模型在 $MRR@20$ 指标上提高了 12.50%, 能更好地预测用户的下一次点击项目.

关键词: 序列信息; 图信息; 图神经网络; 注意力机制; 会话推荐

中图分类号: TP391 **文献标识码:** A **DOI:** 10.19907/j.0490-6756.2022.062003

Graph neural networks combined with multi-source graph information for session-based recommendation

LIN Wei, WU Kai-Jun

(College of Information, Shanghai Ocean University, Shanghai 201306, China)

Abstract: Existing session-based recommendations with graph neural networks could capture the item's transition relationship by constructing graph structures from sessions. However, most graph neural networks and their improved models only consider the single transition relationship of items in the session when modeling sessions. As a result, a large amount of effective information is ignored, and the analysis of hidden correlations between items is lacking. Therefore, a session-based recommendation algorithm with graph neural network and multi-source graph information is proposed. In the algorithm, the users' repeat behavior information and item content related information are incorporated into the session graph modeling process, which effectively extracts the deeper complex transformation relationship of items, and aggregates it through linear transformations. In addition, an external attention mechanism is used to obtain the hidden association information of the session sequence items, making the generated session vectors more accurate. The experiments were performed on the real datasets: Yoochoose and Diginetica, and the results showed that the model outperformed the benchmark model. In particularly, it outperforms the state-of-the-art benchmark model GC-SAN, on average by 12.50% in terms of the $MRR@20$ evaluation metric, and can better predict user's next click items.

收稿日期: 2022-02-07

基金项目: 上海市科技创新计划项目(20dz1203800)

作者简介: 林炜(1996—), 男, 广东潮州人, 硕士研究生, 研究方向为推荐系统. E-mail: linweicz@sina.com

通讯作者: 吴开军. E-mail: kjwu@shou.edu.cn

Keywords: Session information; Graph information; Graph neural network; Attention mechanism; Session-based recommendation

1 引言

随着互联网和移动设备的发展和广泛普及,推荐系统在现代社会中也扮演着越来越重要的角色.推荐系统可以有效降低因互联网而产生的信息过载,帮助用户在众多选项中精确地选择用户最感兴趣的商品或信息,提高用户体验.基于内容的推荐系统^[1]和基于协同过滤算法^[2,3]的推荐系统,是近几年来应用最为广泛的推荐方法,这是因为他们能够快速估算出项目间的相似性,同时具备简单、高效的特点.然而,这类推荐算法过于依赖用户的个人信息和长期历史交互数据.由于游客和匿名用户的出现,用户的长期交互数据在实际应用中并非一定能得到.因此出现了基于会话(Session-based)的推荐^[4],该推荐可以通过短期内的匿名交互序列预测用户的下一次点击行为,不用依赖用户的个人信息和长期历史交互数据.

由于其高度的实用价值,大量集中于会话推荐领域的算法不断被提出.早期会话推荐模型算法大多采用模式识别^[5]和马尔科夫链^[6]对会话序列进行处理.近年来,随着深度学习,特别是图神经网络(Graph Neural Network, GNN)的兴起,大量研究开始将其应用于会话推荐领域,实现了强大的模型表现力.但它们仍存在问题:(1) 仅采用会话中连续项目的单次转换关系构建会话图,忽视了用户的重复点击行为信息,掩盖了不同项目对用户的重要程度差异;(2) 忽视了丰富的项目内容关联信息,如项目属性特征交互性和局部图结构交互性等,而这类信息有助于捕捉项目间更深层次的复杂转换关系;(3) 在生成会话表示时,忽略了长距离项目间的联系,同时也缺乏对项目隐含关系的分析.针对这些问题,本文提出了融合多源图信息的图神经网络会话推荐模型.在项目表示向量生成阶段,将用户重复点击行为信息,项目内容关联信息纳入到会话序列的建模过程当中,对会话图信息进行多样性分析与提取.此外,在会话向量表示生成阶段,采用外部注意力机制捕捉不同项目间的隐藏关联性,并与软注意力机制相结合,生成更加精确的会话向量表示.最后利用获得的会话向量表示计算出所有候选项目的点击概率作为输出.

2 相关工作

2.1 基于传统方法

会话推荐是序列推荐的一个子类.早期的会话推荐主要基于数据挖掘技术,尤其是基于规则和共现模式的无模型推荐算法. Han 等^[5]将关联规则挖掘与协同过滤相结合,从而避免挖掘过量信息,有效提高了推荐算法的性能.这类推荐算法具有高效简单、易于实现的特性,然而容易造成数据丢失,且无法处理诸如长尾数据等复杂数据.由于马尔科夫链具有捕捉用户与项目的交互序列模式,基于马尔科夫链的算法^[6]也被大量提出.然而马尔科夫链虽然可以有效减少信息丢失问题,但是忽略了不连续项目之间的关系,难以捕捉到项目的高阶依赖关系,从而导致用户偏好与序列顺序的失衡. FPMC 算法^[7]将马尔科夫链与因式分解机算法结合到一起,利用因式分解机将观察到的项目转换关系分解为用户或项目的潜在表示,从而弥补了马尔科夫链的不足之处.

2.2 基于深度学习方法

近年来,深度学习算法由于其强大的性能,大量的研究开始将其应用于视觉图像处理与自然语言处理等诸多领域,皆取得了不错的成绩.其中,循环神经网络(Recurrent Neural Network, RNN)因其强大的序列数据处理能力,成为会话推荐领域的重要组成部分. Hidasi 等率先提出了用于处理会话序列的 RNN 模型 GRU4Rec^[8],被认为是深度学习模型在会话序列推荐领域的首次应用. Li 等提出了一种神经注意力推荐机(NARM)^[9],采用注意力机制捕捉项目特征. STAMP 模型^[10]从会话的上下文信息中获取用户的长期偏好信息,同时将用户的最后一次点击作为用户当前兴趣表示.

随着图神经网络 GNN 在深度学习领域中的兴起, Wu 等将其应用于会话推荐领域,提出了 SR-GNN 模型^[11],通过将序列数据建模为图形结构,最大限度地保留了项目间的转换关系,在推荐性能上相较于 RNN 模型有着明显的提高. GCRNN 模型^[12]通过图卷积神经网络层(Graph Convolutional Network, GCN)捕捉会话图结构信息,以此获取更丰富的项目转换关系.受注意力机制在各种任务中的成功启发,研究人员应用注意力

机制学习图中节点表示,产生了图注意力网络(Graph Attention Network, GAT)^[13]. Li 等^[14]将图注意力模型应用于序列推荐领域,从而把不同领域项目之间的内容关联性纳入考虑. 为了弥补 GAT 模型无法有效捕提高阶图信息的特性, Zhang 等提出了自适应结构指纹(ADSF)模型^[15],在节点属性特征外,引入了以节点为中心的图形结构信息作为辅助,丰富了对节点内容信息的分析.

Liang 等提出完全基于注意力机制的 Transformer 模型^[16],在机器翻译任务上取得了最先进的性能表现. GC-SAN 模型^[17]在 SR-GNN 基础上结合了 Transformer 模型,保留 GNN 对会话项目建模能力的同时,采用自注意力机制捕捉项目的全局依赖关系,并与局部偏好信息加权结合,生成更加精确的会话表达. Tay 等^[18]通过大量对比实验,验证自注意力机制中的自我值计算的非必要性,特别是在对话生成等领域,随机生成矩阵相比标准自注意力机制具有更好的表现力,同时运算速度更快. Guo 等提出了外部注意力模型(External Attention, EA)^[19],针对自注意力机制的约束条件,采用基于两个外部的小型可学习记忆单元进行替换,从而获得比自注意力机制更加优秀的模型表现能力.

3 本文模型

3.1 概述

会话推荐是根据当前匿名用户与项目的交互序列进行建模学习,从而生成用户可能选择的下一

项推荐. 使用 $V = \{v_1, v_2, \dots, v_m\}$ 表示数据集的所有项目集合,其中 m 表示所有项目的数量. $S = \{S_1, S_2, \dots, S_n\}$ 表示用户交互项目所构成的全部会话序列的集合. $S_s = [v_{s,1}, v_{s,2}, \dots, v_{s,n}]$ 表示用户在会话序列 s 中的交互项目的集合,其中 $S_s \in V$. 会话推荐模型的工作任务,即是根据给定的会话序列 S_s ,预测出用户下一个可能的交互项目 $v_{s,n+1}$. 具体来说,模型的输出即是当前会话序列的下一时刻,用户可能产生交互的项目所构成的候选项目各自所代表的概率列表 $Y = \{y_1, y_2, \dots, y_m\}$. 其中 y_i 表示 $t+1$ 时刻项目 v_i 与用户交互的可能性大小, y_i 的值越高,则对应项目 v_i 交互的可能性就越大. 将预测的项目概率进行降序排列,从中选择概率最大的 N 个项目作为推荐列表输出.

模型的总体框架如图 1 所示,主要包含:图神经网络模块、会话表示生成模块和预测模块三部分. 在图神经网络模块中,首先将匿名用户对应的当前会话序列 S_s 转换为会话图 G_s . 设计用户行为信息模块和项目内容信息模块提取图中信息,从而捕捉更深层次的项目转换关系,生成更加准确的项目隐藏向量. 在会话表示生成模块中,利用外部注意力机制辅助提取出会话序列中的隐藏项目关联信息,并与软注意力机制提取结果相结合,作为用户的长期偏好. 将会话序列中最后一个项目作为用户短期偏好,并与获取的长期偏好相结合,借此在预测模块中求出所有候选项目在当前会话的下一点击中出现概率.

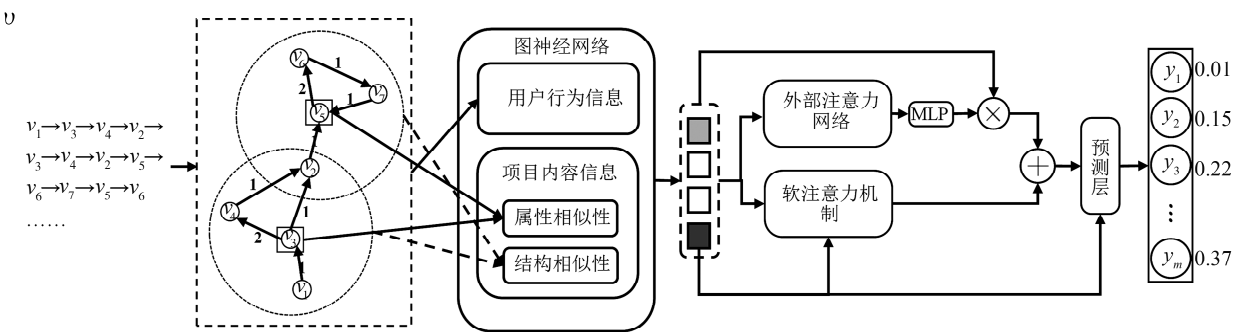


图 1 模型总体框架
Fig. 1 Framework of the proposed model

3.2 图神经网络

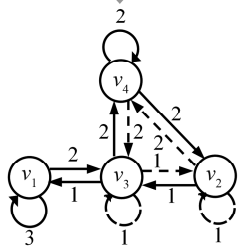
3.2.1 会话图生成 首先将会话序列 S_s 转换为对应的会话图 $G_s = (V_s, E_s)$, 其中 V_s 代表节点集, 会话图 G_s 中的每个节点 $v_i \in V$ 都代表会话序列 S_s 中唯一的项目. E_s 代表边集, 其中边 $(v_i, v_j) \in E_s$ 表示用户在当前会话序列 S_s 中点击项目 v_i 后, 再

点击项目 v_j 的情况, 其初始权重定义为会话中 (v_i, v_j) 出现的频率, 将该初始权重存储于会话图 G_s 对应的初始邻接矩阵 A_s 中. 把图 G_s 中节点 v_i 投入到 d 维嵌入向量空间 E 中, 生成对应的 d 维项目隐藏向量 e_i . 本文在图神经网络模块中设计两个子属模块: 用户行为信息模块和项目内容信息模

块,以丰富对于会话图中节点的复杂转换关系的分析。

3.2.2 用户行为信息模块 用户行为信息模块用于生成两个邻接矩阵 A_s^O 和 A_s^I 。传统基于 GNN 的会话推荐模型,如文献[11,17]将会话序列中项目的单次转换关系作为邻接矩阵权重系数。而用户在点击项目的过程中,总是倾向或无意识地对喜欢的项目进行重复性点击行为,这在一定程度上体现了用户偏好。如图 2 所示,会话序列 S_1 中的各节点项目 v_1 出现了 5 次, v_2 出现了 2 次, v_3 出现了 3 次,然而依照传统 GNN 方法构筑邻接矩阵,如图 2a 所示, v_1, v_2, v_3, v_4 之间的初始链接权重系数皆为 1。这样会忽略掉用户大量的重复点击信息,掩盖不同项目对于用户的重要程度差异,使得推荐的精度下降。

$S_1 = \{v_1, v_3, v_4, v_4, v_2, v_3, v_1, v_1, v_1, v_3, v_4, v_4, v_2\}$ target: v_3
 $S_2 = \{v_1, v_3, v_4, v_4, v_2, v_3, v_1, v_1, v_1, v_3, v_4, v_4, v_2\}$ target: v_2



	1	2	3	4		A_s^O				A_s^I			
1	1	0	1	0	1	1/2	0	1/2	0	1/2	0	1/2	0
2	0	0	1	0	2	0	0	1	0	0	0	0	1
3	1	0	0	1	3	1/2	0	0	1/2	1/2	1/2	0	0
4	0	1	0	1	4	0	1/2	0	1/2	0	0	1/2	1/2

(a) 传统 GNN 邻接矩阵构建

	1	2	3	4		A_s^O				A_s^I			
1	3	0	2	0	1	3	0	2	0	3	0	1	0
2	0	0	1	0	2	0	1	1	0	0	1	0	2
3	1	0	0	2	3	1	0	1	2	2	1	1	0
4	0	2	0	2	4	0	2	0	2	0	0	2	2

(b) 用户行为信息纳入邻接矩阵

图 2 邻接矩阵生成方法

Fig. 2 Adjacency matrix generation method

用户的重复点击行为可分为存在的可观测重复点击行为与可能存在的隐藏重复点击行为。用户的可观测重复点击行为可以直接从会话序列中观测得到。而用户的隐藏重复点击行为往往隐藏在会话序列的最后一次点击项目与实际的目标项目之间。当前会话的目标项目与会话中的最后点击项目

通常是邻近项目甚至是相同项目,如会话 S_2 的目标项目与 S_2 的最后一个项目完全相同,这类情况在会话数据中经常出现。采用传统 GNN 构造 S_2 的邻接矩阵时,若此前会话中不存在最后项目的自我转换关系,则会忽略掉可能存在的隐藏重复点击行为,故参考文献[20]的做法,将自环效应也纳入考虑。将两种用户行为信息纳入考虑后生成的新邻接矩阵如图 2b 所示。

根据以上分析,用户行为信息模块将节点间的链接数作为邻接矩阵 A_s^O 和 A_s^I 的权重系数,即采用初始邻接矩阵 A_s 构筑会话图对应的出入度矩阵 A_s^O 和 A_s^I 。因此,生成的新邻接矩阵可以表示如下。

$$A_s^O = \text{count}(v_i, v_j) \quad (1)$$

$$A_s^I = \text{count}(v_j, v_i) \quad (2)$$

其中, $\text{count}(x, y)$ 表示项目 y 在用户点击项目 x 后出现的次数。同时,将自环 $\text{count}(x, x)$ 也纳入考虑,有助于提高模型表现力。若 $\text{count}(x, x)$ 的值为 0,则将其设置为 1。如果两个节点之间不存在链接关系,则 A_s^O 和 A_s^I 中的相应位置用 0 填充。

3.2.3 项目内容信息模块 项目内容信息模块通过提取会话图 G_s 中的节点内容信息,构筑邻接矩阵 A_c^O 和 A_c^I 。不同于用户行为信息模块,项目内容信息模块主要关注图形结构中节点项目内容上的关联性。为了推断不同节点项目之间属性特征上的交互性关系,首先采用注意力机制学习节点间的权重系数,特别地,从节点 v_i 到节点 v_j 的权重是根据其初始节点状态进行计算的。具体公式如下。

$$\alpha_{ij} = \frac{\exp(\text{LeakyRelu}(W_{att} [W_w e_i \parallel W_w e_j]))}{\sum_l \exp(\text{LeakyRelu}(W_{att} [W_w e_i \parallel W_w e_j]))} \quad (3)$$

其中, $W_w \in \mathbb{R}^{d \times d}$, $W_{att} \in \mathbb{R}^d$ 是可学习参数; \parallel 表示拼接操作; LeakyRelu 函数赋予所有负数值一个非零的斜率,防止节点 v_i 信息遗失。使用 softmax 函数将获得系数转换为概率形式使得不同节点间的交互性可以进行相互比较。

单纯将节点内容相似性作为权重系数,容易忽略掉会话图中丰富的图结构信息,如局部边连接的拓扑或形状结构等信息。如图 3 所示,中间项目与右边项目之间属性特征上的相似性强,并不代表结构上的关联性强。而中间项目与左边项目间的属性特征上相似性低,在整体结构上的关联性可能更强。单纯考虑项目特征上的相似性,会在一定程度上掩盖用户短时间内的主要意图。

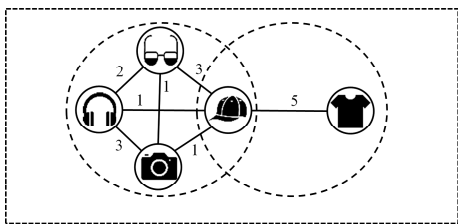


图 3 展示项目的内容关联性与结构关联性的例子

Fig. 3 An example showing the content and structural relevance of items

为了充分利用会话图中的结构信息,参考 ADSF 模型^[15]的构造方式,利用重启随机游走算法(Random Walk with Restart, RWR)^[21]计算会话图中节点 v_i 与会话图中其余节点的相关程度。在基于会话的推荐领域当中,由于会话序列的长度相对较短,生成的会话图规模并不大,所以可以在不进行任何采样的情况下,考虑单个节点的整个邻域。以节点 v_i 为初始节点,以初始邻接矩阵 \mathbf{A}_s 中对应边值为正比的概率转移到其邻接节点,在每一步中,仍旧有概率 c 回到初始节点 v_i 。直接采用 RWR 的迭代收敛公式,具体如下式所示。

$$d_i = (I - c\tilde{\mathbf{A}}_s)^{-1} x_i^T \quad (4)$$

其中, $\tilde{\mathbf{A}}_s$ 表示归一化后的初始邻接矩阵 \mathbf{A}_s ; $c \in [0, 1]$ 为超参数,表示随机游走和返回原节点的概率; $x_i \in \mathbb{R}^n$ 是除节点 v_i 对应位置为 1 外全为 0 的向量。 $d_i = [d_{ij}] \in \mathbb{R}^n$ 量化了节点 v_i 与其邻接节点 v_l 之间的相关性,同时也反映出会话图的局部结构信息。

将 d_i 和 d_j 的范围扩展到会话图 G_s 的全部节点上,计算出以每个节点为中心的图形结构之间的 jaccard 相似性,同时采用 softmax 函数实现图结构间的相似性可比,具体公式如下。

$$\beta_{ij} = \text{softmax}\left(\frac{\sum_{v_l \in G_s} \min(d_{il}, d_{jl})}{\sum_{v_l \in G_s} \max(d_{il}, d_{jl})}\right) \quad (5)$$

其中, max 和 min 函数采用平滑形式:

$$\max(x, y) = \lim_t \frac{\log(e^{t \cdot x} + e^{t \cdot y})}{t} \quad (6)$$

$$\min(x, y) = \lim_t -\frac{\log(e^{t \cdot (-x)} + e^{t \cdot (-y)})}{t} \quad (7)$$

将获得的基于节点属性特征权重 α_{ij} 以及基于局部图结构权重 β_{ij} 相结合,获得最终权重表示。

$$s_{ij} = \frac{\sigma(\alpha_{ij})\alpha_{ij} + \sigma(\beta_{ij})\beta_{ij}}{\sigma(\alpha_{ij}) + \sigma(\beta_{ij})} \quad (8)$$

其中, $\sigma(\cdot)$ 表示 sigmoid 函数,用以调整 α_{ij} 和 β_{ij} 的比例。因此会话图 G_s 可以采用以上计算结果,根据

节点 v_i 与节点 v_j 的链接情况,构筑存储项目内容关联信息的邻接矩阵 \mathbf{A}_c^O 和 \mathbf{A}_c^I 。具体如下表示。

$$\mathbf{A}_c^O[i, j] = s_{ij} \quad (9)$$

$$\mathbf{A}_c^I[i, j] = s_{ji} \quad (10)$$

同样考虑到自环效应,若 s_{ii} 对应位置为 0,则将其置为 1。

3.2.4 节点状态更新 在获得了 \mathbf{A}_s^O 、 \mathbf{A}_s^I 和 \mathbf{A}_c^O 、 \mathbf{A}_c^I 之后,根据上述的四个邻接矩阵,为目标结点 v_i 聚合相关的邻域信息,信息聚合的 t 时刻公式如下表示。

$$\begin{aligned} a_i^{s,t}(O) &= \mathbf{A}_{s,i}^O[e_1^{t-1}, \dots, e_n^{t-1}]W_s^O + b_s^O \\ a_i^{s,t}(I) &= \mathbf{A}_{s,i}^I[e_1^{t-1}, \dots, e_n^{t-1}]W_s^I + b_s^I \\ a_i^{c,t}(O) &= \mathbf{A}_{c,i}^O[e_1^{t-1}, \dots, e_n^{t-1}]W_c^O + b_c^O \\ a_i^{c,t}(I) &= \mathbf{A}_{c,i}^I[e_1^{t-1}, \dots, e_n^{t-1}]W_c^I + b_c^I \end{aligned} \quad (11)$$

其中, $\mathbf{A}_{s,i}^O$ 、 $\mathbf{A}_{s,i}^I$ 、 $\mathbf{A}_{c,i}^O$ 和 $\mathbf{A}_{c,i}^I$ 分别对应节点 v_i 的四个邻接矩阵中的第 i 行元素。 $[e_1^{t-1}, \dots, e_n^{t-1}]$ 表示在 $t-1$ 时刻会话序列包含的节点表示向量组成的列表。值得注意的是, e_i^0 则表示节点投入到嵌入空间 E 后的初始向量值,当 $t=1$ 时,表示节点第一次从邻域收集信息后更新的向量状态。提取节点的邻域信息之后,将它们拼接到一起获得节点的完整邻域信息聚合表示,具体操作如下。

$$a_i^t = [a_i^{s,t}(O) \parallel a_i^{s,t}(I) \parallel a_i^{c,t}(O) \parallel a_i^{c,t}(I)] \quad (12)$$

接着,采用门控循环单元(GRU)更新每一个节点的表示,如下式所示。

$$\begin{aligned} z_i^t &= \sigma(W_z a_i^t + U_z e_i^{t-1}) \\ r_i^t &= \sigma(W_r a_i^t + U_r e_i^{t-1}) \\ \tilde{e}_i^t &= \tanh(W_o a_i^t + U_o (e_i^{t-1} \odot r_i^t)) \\ e_i^t &= (1 - z_i^t) \odot e_i^{t-1} + z_i^t \odot \tilde{e}_i^t \end{aligned} \quad (13)$$

其中, $W_z, W_r, W_o \in \mathbb{R}^{4d \times d}$; $U_z, U_r, U_o \in \mathbb{R}^{d \times d}$ 属于模型可学习参数; \odot 表示同位元素对应相乘。经过 GRU 单元计算后,得到节点 v_i 更新后的节点向量表示 e_i^t 。因此式(13)可以简写为

$$e_i^t = \text{GRU}(e_i^{t-1}, a_i^t) \quad (14)$$

先前的研究^[22]证明将低阶和高阶表示融合到一起模型的表现能力更好。引入额外的残差连接更新节点状态,有利于低阶特征重用和梯度反向传播。因此式(14)可以被重写为

$$e_i^t = \text{GRU}(e_i^{t-1}, a_i^t) + e_i^{t-1} \quad (15)$$

经过 t 步骤之后,得到了会话图 G_s 中每个节点的最终表示 e_i^t 。为了方便表示,使用 e_i 替代 e_i^t 。

3.3 会话向量表示生成

外部注意力模型是由文献[19]所提出的一种

新型注意力机制模型,该模型基于两个外部、小型、可学习且共享的内存进行训练. 外部注意力机制具有线性复杂性,可以考虑当前会话序列中所有会话项目之间的隐藏关联性. 将会话序列 S_s 输入到图神经网络后,可以获得对应会话图 G_s 中所有节点隐藏向量,即 $E=[e_1, e_2, \dots, e_n]$. 接下来,利用多头外部注意力机制捕捉会话序列的项目隐含关系,具体公式如下.

$$L=(L)_{ij}=\text{Norm}(EM_i^k) \quad (16)$$

$$H_i=LM_i^v \quad (17)$$

其中,矩阵 $M_i^k \in \mathbb{R}^{d_i \times s}$, $M_i^v \in \mathbb{R}^{s \times d_i}$ 分别是第 i 个注意力头部对应的两个存储记忆矩阵, d_i 表示第 i 头的维度大小, s 表示记忆空间的维度大小. 与自注意力机制采用 softmax 函数对乘积矩阵进行归一化不同,外部注意力机制采用双重归一化方式对矩阵进行处理. Norm 的具体过程如下.

$$\begin{aligned} \tilde{l}_{ij} &= EM_i^k \\ \hat{l}_{ij} &= \frac{\exp(\tilde{l}_{ij})}{\sum_l \exp(\tilde{l}_{lj})} \\ l_{ij} &= \frac{\hat{l}_{ij}}{\sum_k \hat{l}_{ik}} \end{aligned} \quad (18)$$

接下来,将每个头部学习得到的向量拼接起来,并采用线性转换获得外部注意力机制的最终结果.

$$H=W_h(H_1 \parallel H_2 \parallel \dots \parallel H_h) \quad (19)$$

其中, $W_h \in \mathbb{R}^{d \times d}$ 是可学习参数; h 代表注意力分部的数目. 最终获得 $H \in \mathbb{R}^{n \times d}$, 其中 $d = \sum_1^h d_i$.

接着,聚合会话序列中所有项目信息,将外部注意力机制获取的隐藏向量 H 作为注意力参数,将原先的项目隐含向量当做图像,两者进行元素对位相乘求和,获取用户在会话序列的隐藏偏好表示 s_f , 具体公式如下.

$$s_f = \sum_{i=1}^n \sigma(h_i) \odot e_i \quad (20)$$

软注意力机制赋予会话序列中最后一个项目 e_n 最大的优先级,求出会话序列的全局表示 s_g , 具体公式如下.

$$\alpha_i = q^T \sigma(W_1 e_n + W_2 e_i + b) \quad (21)$$

$$s_g = \sum_{i=1}^n \alpha_i e_i \quad (22)$$

其中, $q \in \mathbb{R}^d$ 和 $W_1, W_2 \in \mathbb{R}^{d \times d}$ 是可学习参数. 将外部注意力机制与软注意力机制获取的向量结合,作为新的会话全局表示.

$$s_o = s_f + s_g \quad (23)$$

通常用户的最后点击行为往往受到前一次点击行为的影响,故将会话序列中的最后一个项目作为用户的局部偏好表示,即 $s_l = e_n$.

3.4 预测层

将会话序列的局部偏好表示和全局偏好表示结合起来,并通过一个线性转换获得会话序列 S_s 对应的会话序列向量表示,具体过程如下.

$$s_h = W_3 [s_l \parallel s_o] \quad (24)$$

其中, $W_3 \in \mathbb{R}^{d \times 2d}$ 是可学习参数,负责将全局表示向量和局部表示向量拼接后的结果映射到一个 d 维向量空间,由此得到会话序列的向量表示 s_h .

在获得了每一个会话序列对应的向量表示 s_h 后,将每一个候选项目的嵌入向量 e_i 与其相乘,获得每个候选项目的分数值.

$$\hat{z}_i = s_h^T e_i \quad (25)$$

$\hat{z}=[\hat{z}_1, \hat{z}_2, \dots, \hat{z}_m]$ 代表每一个候选项目的推荐分数. 接着通过 softmax 函数获得模型的输出概率向量.

$$\hat{y}_i = \text{softmax}(\hat{z}_i) \quad (26)$$

$\hat{y}=[\hat{y}_1, \hat{y}_2, \dots, \hat{y}_m]$ 表示当前会话序列中所有候选项目在下一时刻被用户点击的概率.

采用交叉熵作为优化目标来学习参数.

$$\begin{aligned} L(\hat{y}) = & - \sum_{i=1}^n y_i \log(\hat{y}_i) + \\ & (1 - y_i) \log(1 - \hat{y}_i) + \frac{\lambda \|\theta\|^2}{2} \end{aligned} \quad (27)$$

其中, $y_i \in y$ 为 one-hot 向量,代表着真实的概率分布情况. 最后使用反向传播算法来计算模型.

4 实验与分析

4.1 实验数据集

采用两个具有代表性的数据集 Yoochoose 和 Diginetica 来测试本文提出模型的性能. Yoochoose 数据集来自 2015 年的 RecSys 挑战赛,其内包含了用户在电子商务网站 6 个月内的点击数据; Diginetica 数据集是由 2016 年的 CIKM 杯比赛提供的比赛数据,包含适用于会话推荐系统的用户交易数据. 出于公平方面的考虑,参考文献[11, 12, 17]的做法,过滤掉长度为 1 的会话序列,以及出现次数少于 5 次的项目. 同时,将每一个会话的倒数第一个点击项目作为该会话对应的标签,此前的点击项目作为模型学习用数据. 将 Yoochoose 数据集中最后几天的数据以及 Diginetica 数据集中的最后几周数据划分为测试集,此前部分划分为训练集. 由于 Yoochoose 数据集中包含 7 966 257

条会话,其数据量过于庞大,将训练集中最新部分按照 1/64 的比例划分成新的训练集. 经过预处理之后,数据集的情况如表 1 所示.

表 1 实验数据集
Tab. 1 Experimental dataset

数据集	Yoochoose(1/64)	Diginetica
点击次数	557 248	982 961
训练会话数目	369 859	719 470
测试会话数目	55 898	60 858
项目总数	16 766	43 097
平均长度	6. 16	5. 12

4.2 评价指标

本文采用了会话推荐系统常用的两个评价指标来测试模型的实际效果:精确率(P@K, Precision)和平均倒数排名(MRR@K, Mean Reciprocal Rank).

P@20:被广泛用于预测准确度的评价指标,它表示前 N 个项目中推荐正确的项目占比,这里 N 取 20,具体计算公式为

$$Precision = \frac{TP}{TP + FP} \tag{28}$$

其中, TP 表示推荐正确的项目; FP 代表推荐错误的项目.

MRR@20:指正确推荐项目的倒数排名的平均值. 当排名超过 20 时,倒数排名设置为 0. MRR 方法充分考虑到排名的顺序性,当列表中推荐的正确项目越靠前时, MRR 值越大. 这里 N 取 20,具体计算公式为

$$MRR = \frac{1}{N} \sum_{i=1}^N \frac{1}{rank_i} \tag{29}$$

其中, N 表示项目查询个数; $rank_i$ 表示第 i 个项目的正确排名.

4.3 参数设置

模型的单次训练批量大小设置为 100. 采用均值为 0, 标准差为 0. 1 的高斯分布初始化模型的所有可学习参数. 利用小批量 Adam 优化器对参数进行优化. 初始率学习率设置为 0. 001, 每经过 3 个训练周期衰减 0. 1. L_2 参数正则化项设置为 0. 000 05, 用以减轻过拟合. 在两个数据集上的隐藏向量维度均设为 100. 经验证集测试后, 超参数 c 选取 0. 6 为最适值. 外部注意力机制固定分布头部数量为 4, 记忆空间维度大小选取 32.

4.4 算法比较

为了验证本文提出的模型性能, 将其与下列基

准算法进行比较: (1) POP: 直接推荐整个训练集中最流行的项目, 尽管机制简单, 但在某些情况下可以作为参考; (2) S-POP: 直接推荐当前会话中最流行的项目; (3) Item-KNN^[23]: 最近邻算法之一, 向用户推荐与会话中最后点击项目最相似的项目, 相似值的大小通过计算两者的余弦距离获得; (4) FPMC^[7]: 基于马尔科夫链和矩阵分解的混合模型算法, 在考虑到项目的顺序性同时, 可以捕捉序列的长期依赖关系; (5) GRU4REC^[8]: 将 RNN 模型用于会话推荐领域, 利用小批量的并行会话序列进行训练, 同时引入使用排名的损失函数; (6) NARM^[9]: 在 RNN 的基础上引入了注意力机制, 将用户的长期偏好纳入模型考虑; (7) STAMP^[10]: 同样采用注意力机制捕捉会话中用户的长期偏好信息, 同时将最后一次点击项目作为用户的短期偏好; (8) SR-GNN^[11]: 首次将图神经网络引入会话推荐领域, 可以捕捉项目间的复杂转换关系; (9) GC-SAN^[17]: 使用自注意力机制学习会话项目间的全局依赖关系, 从而平衡会话的长期与短期表达.

在 Yoochoose 和 Diginetica 数据集上进行测试, 通过与上述基准推荐算法进行比较, 比较结果如表 2 所示, 通过记录, 可以看到在所有的推荐算法中, POP 由于只根据全数据中项目出现的频率做出推荐, 因而推荐性能最差. S-POP 考虑到了会话信息的影响, 推荐性能相较于 POP 有了显著的改善. 这两者都是基于数据挖掘的无模型算法. FPMC 是结合了因式分解技术和马尔科夫链的混合算法, 其推荐性能同样证明了上下文信息的重要性, 同时考虑到了会话的序列特征. Item-KNN 根据项目间的相似性进行推荐, 取得了比 FPMC 更好的推荐效果, 同时, Item-KNN 并未考虑到会话的序列信息, 在数据集扩大的时候, 可能无法取得更好的泛化能力.

GRU4REC 采用 RNN 结构对会话进行编码, 相较于传统方法在推荐性能上有着显著的提高. NARM 在 GRU4REC 的基础上融合了注意力机制捕获项目, 使得该模型的推荐性能有较大提升, 这证明了对不同的输入项目分配不同的注意力能有效提高推荐性能.

SR-GNN 将会话序列建模为图形格式, 并使用类 RNN 的门控单元 GRU 对项目进行编码学习, 同时保留 STAMP 中结合会话全局和局部的用户偏好的软注意力机制预测会话表达, 相较于

RNN 系列的模型有着明显的提高. 然而模型更多地集中在会话序列中项目的单次转换关系上, 忽视了用户的偏好和项目的内容关联性. GC-SAN 则是在 SR-GNN 的基础上, 采用了自注意力机制平衡长期和短期的关系.

表 2 各类推荐模型在数据集上的性能比较
Tab. 2 Performance comparison of various recommendation models on datasets

方法	Yoochoose(1/64)		Diginetica	
	P@20	MRR@20	P@20	MRR@20
POP	6. 70	1. 66	0. 86	0. 22
S-POP	30. 44	18. 36	35. 75	11. 57
ItemKNN	51. 60	21. 81	5. 24	1. 98
FPMC	45. 62	15. 02	26. 53	6. 95
GRU4Rec	64. 64	22. 89	29. 46	8. 33
NARM	68. 32	28. 64	49. 70	16. 67
STAMP	69. 83	31. 94	55. 32	14. 42
SR-GNN	70. 36	30. 64	51. 68	17. 67
GC-SAN	70. 06	29. 35	50. 45	17. 43
Ours	69. 98	32. 68	56. 12	19. 88

本文提出的模型延续了 SR-GNN 的主体框架,但是在会话图信息聚合方面将用户重复点击行为信息,包含节点属性特征和图结构信息等项目关联内容都纳入考虑,能更深层次地捕捉项目间的复杂转换关系. 同时采用外部注意力机制收集会话中项目的隐藏依赖关系,并与软注意力机制生成的全局表示相结合,生成更精确的会话表示向量. 相较于各项评分指标最高的 SR-GNN 模型,在 Yoochoose 数据集上的 MRR@20 上提高了 6. 65%; 在 Diginetica 数据集上 P@20 提高了 8. 59%, 在 MRR@20 上提高了 12. 50%. 而在 Yoochoose 数据集上的 P@20 指标反而下降了 0. 54%. 通过实验结果分析,本文模型适用于用户点击次数较多,同时包含更多项目的会话数据集. 同时对于 MRR@20 评分指标的提升要明显优于 P@20.

4.5 消融实验

为了验证本文提出模型构成模块各方面的有效性,进一步推测出不同部分模块对于模型整体的贡献比例,设置了消融实验,实验结果如表 3 所示. (1) GNN-0:去除用户行为信息模块和项目内容信息模块,以及外部注意力机制,仅采用带自环的无权图聚合邻域信息;(2) GNN-1:去除项目内容信

息模块和外部注意力机制,采用用户行为信息模块,在建模会话项目时将用户重复点击行为纳入考虑;(3) GNN-2:去除外部注意力机制,采用用户行为信息模块与项目内容信息模块辅助节点聚合邻域信息;(4) GNN-3:去除项目内容信息模块,采用用户行为信息模块聚合会话图信息,并采用外部注意力机制提取项目隐含关联信息.

表 3 消融实验中各模块性能比较
Tab. 3 Performance comparison of each module in ablation experiments

方法	Yoochoose(1/64)		Diginetica	
	P@20	MRR@20	P@20	MRR@20
GNN-0	70. 13	30. 84	51. 68	17. 88
GNN-1	69. 73	31. 94	55. 32	19. 39
GNN-2	69. 89	32. 57	55. 64	19. 80
GNN-3	69. 96	32. 16	55. 73	19. 52
Ours	69. 98	32. 68	56. 12	19. 88

根据表 3 可知,GNN-1 将用户重复点击行为纳入考虑,可以极大提高模型表现力,特别是在 Diginetica 数据集上相较于采用无权图的 GNN-0,在 P@20 和 MRR@20 上都有着极大的提高. 依据实验结果与数据分析,原因在于 Diginetica 数据集中包含有大量的用户重复点击行为. 而在 Yoochoose(1/64)数据集中用户点击次数较少,包含的用户重复点击次数同样减少,将其纳入信息聚合当中,容易发生过拟合现象,造成推荐项目数量减少,使得 P@20 上反而下降了 0. 21%. GNN-2 在用户行为信息模块的基础上,添加项目内容信息模块提取的节点项目内容关联信息作为辅助,能有效提高评分指标. GNN-3 在 GNN-1 的基础上采用外部注意力机制作为辅助,挖掘会话序列中项目间存在的隐含关联性,可以少量提高推荐性能. 通过消融实验,表明了模型各模块的设计合理性.

4.6 实验参数的影响

4.6.1 外部注意力记忆空间维度 使用本文模型在 Yoochoose 和 Diginetica 数据集上进行测试,横向对比不同记忆空间维度对于评分指标 MRR@20 产生的影响. 将记忆空间维度划分为 8、16、24、32、40、48、56、64. 测试结果如图 4 所示. 在图 4a 中,当外部注意力机制的记忆空间维度处于[32, 48]区间时,MRR@20 指标在 Yoochoose1/64 数据集上达到最佳表现. 在图 4b 中,当记忆空间维度

处于 $[24, 40]$ 时, $MRR@20$ 指标在 Diginetica 数据集上达到最佳表现. 因此, 本文选取记忆空间维度为 32, 综合上最利于模型准确率提高.

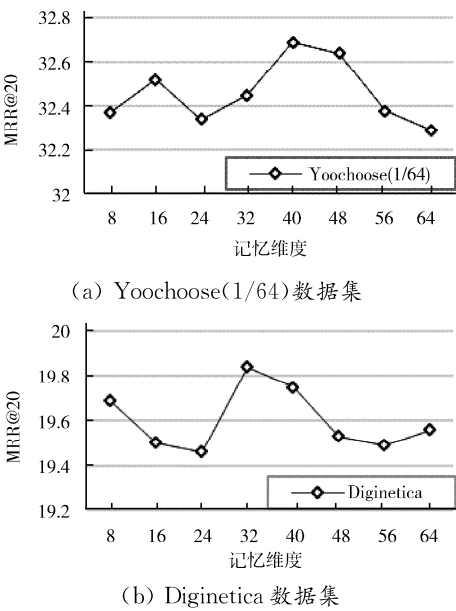


图 4 外部注意力记忆维度对于 $MRR@20$ 的影响
Fig. 4 The influence of memory dimension of external attention on $MRR@20$

4.6.2 超参数 c 值的影响 在 Yoochoose(1/64) 和 Diginetica 数据集上探索不同参数 c 值对于 $MRR@20$ 产生的影响. 从图 5 可以看出, 当 c 值位于 $[0.4, 0.6]$ 区间时, 对模型的提升程度最大. 本文选取 0.6 作为 c 的参数值.

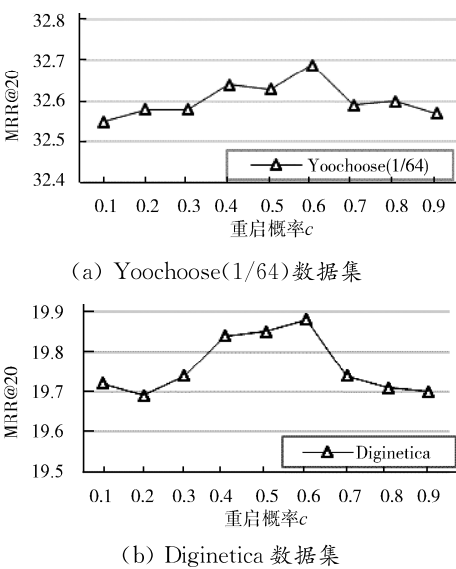


图 5 超参数 c 值对于 $MRR@20$ 的影响
Fig. 5 The influence of hyperparameter c value on $MRR@20$

4.6.3 嵌入维度大小的影响 通过对比实验测试

不同项目嵌入维度大小在两个数据集上对于 $MRR@20$ 产生的影响, 同时与 SR-GNN 基准模型做对比, 结果如图 6 所示. 在 $MRR@20$ 评分指标下, 本文模型在所有嵌入维度 d 上均优于 SR-GNN. 从图 6a 中可以看出, 在 Yoochoose(1/64)数据集上, SR-GNN 与本文模型的 $MRR@20$ 均随着嵌入维度 d 的不断增大而不断提升, 直到进入 $[120, 140]$ 区间, SR-GNN 进入稳定状态, 而本文模型表现力开始呈现下滑趋势, 这可能是由于本文模型收集过多信息, 因 d 增大而发生了过拟合现象导致. 从图 6b 中看到, 在 Diginetica 数据集中, 超过 $[80, 100]$ 区间之后, SR-GNN 模型开始呈现下滑趋势, 而本文模型仍在增长, 一直到进入 $[120, 140]$ 区间才开始下滑. 根据实验结果分析, 本文模型在 $MRR@20$ 指标上的最适嵌入维度 d 为 $[100, 120]$ 之间.

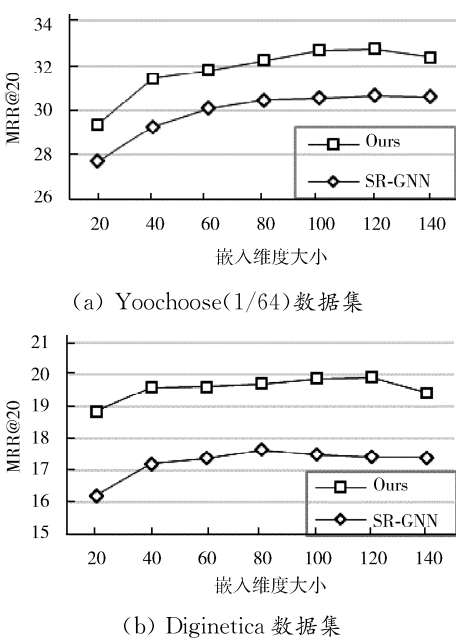


图 6 项目嵌入维度对于 $MRR@20$ 的影响
Fig. 6 The influence of item embedding dimension on $MRR@20$

5 结 论

为了解决当前基于图神经网络的会话推荐算法未能充分利用会话图信息的问题. 本文提出融合多源图信息的图神经网络会话推荐算法. 该模型将用户重复点击行为信息, 项目内容关联信息纳入会话序列建模过程之中, 以此捕捉项目间更深层的复杂转换关系. 同时采用外部注意力机制捕捉会话序列中项目的隐含依赖关系, 实现对当前会话序列中

项目间关联信息的提取,生成更利于推荐的会话向量表示.模型在两个真实数据集上进行实验,证明了所提出的算法在会话推荐领域的有效性和可行性.在后续的工作中,将研究如何进一步地挖掘会话图信息的潜能,同时寻找更好的生成会话向量表达方式.

参考文献:

- [1] 王光,张杰民,董帅含,等.基于内容的加权粒度序列推荐算法[J].计算机工程与科学,2018,40:7.
- [2] 郎亚坤,王国中.融入用户隐式信任的协同过滤算法[J].小型微型计算机系统,2021,42:6.
- [3] 陈珏伊,朱颖琪,周刚,等.基于迁移的联合矩阵分解的协同过滤算法[J].四川大学学报:自然科学版,2020,57:7.
- [4] Wang S, Cao L, Wang Y, *et al.* A survey on session-based recommender systems [J]. ACM Comput Surv (CSUR), 2021, 54: 1.
- [5] Han J, Cheng H, Xin D, *et al.* Frequent pattern mining: current status and future directions [J]. Data Min Knowl Disc, 2007, 15: 55.
- [6] Zhang Z Y, Nasraoui O. Efficient hybrid web recommendations based on markov clickstream models and implicit search [C]//Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence (WI07). Fremont, CA, USA: IEEE, 2007.
- [7] Rendie 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. Raleigh, NC, USA: ACM, 2010.
- [8] Hidasi B, Karatzoglou A, Baltrunas L, *et al.* Session-based recommendations with recurrent neural networks [C]// Proceedings of the 4th International Conference on Learning Representations (ICLR). San Juan, Puerto Rico: ICLR, 2016.
- [9] Li J, Ren P J, Chen Z M, *et al.* Neural attentive session-based recommendation [C]// Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. Singapore: ACM, 2017.
- [10] Liu Q, Zeng Y, 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 & Data Mining. London, UK: ACM, 2018.
- [11] Wu S, Tang Y, Zhu Y, *et al.* Session-based recommendation with graph neural networks [C]// Proceedings of the 33th AAAI Conference on Artificial Intelligence. Honolulu, Hawaii, USA: AAAI Press, 2019.
- [12] 曹万平,周刚,陈黎,等.基于会话的图卷积递归神经网络推荐模型[J].四川大学学报:自然科学版,2021,58:022002.
- [13] Velickovi P, Cucurull G, Casanova A, *et al.* Graph Attention Networks [C]// Proceedings of the 6th International Conference on Learning Representations (ICLR). Vancouver, BC, Canada: ICLR, 2018.
- [14] Li Z, Cui Z, Wu S, *et al.* Fi-gnn: Modeling feature interactions via graph neural networks for CTR prediction [C]//Proceedings of the 28th ACM International Conference on Information and Knowledge Management. Beijing, China: ACM, 2019.
- [15] Zhang K, Zhu Y, Wang J, *et al.* Adaptive structural fingerprints for graph attention networks [C]// Proceedings of the 7th International Conference on Learning Representations. New Orleans, LA, USA: [s. n.], 2019.
- [16] 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. Long Beach, CA, USA: Curran Associates Inc, 2017.
- [17] Xu C, Zhao P, Liu Y, *et al.* Graph contextualized self-attention network for session-based recommendation [C]// Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao, China: AAAI Press, 2019.
- [18] Tay Y, Bahri D, Metzler D, *et al.* Synthesizer: rethinking self-attention in transformer models [C]// Proceedings of the 38th International Conference on Machine Learning. Vienna, Austria: [s. n.], 2021.
- [19] Guo M H, Liu Z N, Mu T J, *et al.* Beyond self-attention: external attention using two linear layers for visual tasks [C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: IEEE, 2021.
- [20] Sankar A, Liu Y, Yu J, *et al.* Graph neural networks for friend ranking in large-scale social platforms [C]// Proceedings of the Web Conference 2021. Ljubljana, Slovenia: ACM, 2021.
- [21] Tong H, Faloutsos C, Pan J Y. Fast random walk

with restart and its applications [C]//Proceedings of the 6th International Conference on data mining (ICDM). NW Washington, USA; IEEE, 2006.

[22] Cheng H T, Koc L, Harmsen J, *et al.* Wide & deep learning for recommender systems [C]//Proceedings of the 1st workshop on deep learning for recommender systems. Boston, MA, USA; ACM, 2016.

[23] Sarwar B. Item-based collaborative filtering recommendation algorithms [C]//Proceedings of the 10th International Conference on World Wide Web. Hong Kong, China; ACM, 2001.

引用本文格式:

中 文: 林炜, 吴开军. 融合多源图信息的图神经网络会话推荐算法[J]. 四川大学学报: 自然科学版, 2022, 59: 062003.

英 文: Lin W, Wu K J. Graph neural networks combined with multi-source graph information for session-based recommendation [J]. J Sichuan Univ; Nat Sci Ed, 2022, 59: 062003.