

基于图注意力和会话偏好识别的会话推荐^①



杨中一¹, 陈平华¹, 陈丽丽², 王文斌², 周宏虹²

¹(广东工业大学 计算机学院, 广州 510006)

²(广东省科技创新监测研究中心, 广州 510033)

通信作者: 陈平华, E-mail: phchen@gdut.edu.cn

摘要: 现有的基于自注意力的会话推荐融合所有高阶邻居信息, 在丰富节点项目表示的同时也带来了节点项目表示同质化及弱相关邻居对项目表示产生干扰的问题. 针对这些问题, 提出一种基于图注意力和会话偏好识别的会话推荐模型. 模型在生成节点项目嵌入时仅选择 μ 阶邻居, 采取图注意力机制聚合 μ 阶邻居特征信息; 同时, 针对不同用户的不同会话偏好不一的情况, 提出会话偏好分类识别及会话嵌入学习方法, 将会话偏好分为偏好集中和偏好发散两种类型, 进行分类计算; 最后, 应用软注意力机制融合会话中的项目嵌入学习得到会话嵌入表示进行预测. 在 Nowplaying 和 Diginetica 两个真实数据集上进行实验, 与基线方法相比, 本文模型均达到了更优的实验效果.

关键词: 会话推荐; 图注意力; 会话偏好识别; 邻居选择矩阵; 注意力机制

引用格式: 杨中一, 陈平华, 陈丽丽, 王文斌, 周宏虹. 基于图注意力和会话偏好识别的会话推荐. 计算机系统应用, 2025, 34(7): 128-139. <http://www.c-s-a.org.cn/1003-3254/9889.html>

Session Recommendation Based on Graph Attention and Session Preference Identification

YANG Zhong-Yi¹, CHEN Ping-Hua¹, CHEN Li-Li², WANG Wen-Bin², ZHOU Hong-Hong²

¹(School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, China)

²(Guangdong Science and Technology Innovation Monitoring and Research Center, Guangzhou 510033, China)

Abstract: Existing attention-based session recommendation models integrate all higher-order neighbor information, enriching node-item representations but also leading to representation homogenization and interference from weakly related neighbors. To address this issue, a session recommendation model based on graph attention and session preference identification is proposed. In this model, when generating node-item embeddings, only μ -neighbor are selected, and a graph attention mechanism is used to aggregate the feature information of these μ -neighbor. In addition, to account for varying session preferences across users, a method for session preference classification and embedding learning is introduced, categorizing session preferences into two types: concentrated preferences and divergent preferences, for classification computation. Finally, a soft attention mechanism is applied to integrate the item embeddings learned from the session, obtaining the session embedding representation for prediction. Experiments on two real datasets, Nowplaying and Diginetica, demonstrate that the proposed model achieves superior performance compared to baseline methods.

Key words: session recommendation; graph attention; session preference identification; neighbor selection matrix; attention mechanism

① 基金项目: 广东省重点领域研发计划 (2023B1111050010)

收稿时间: 2024-12-03; 修改时间: 2025-01-02; 采用时间: 2025-01-21; csa 在线出版时间: 2025-05-29

CNKI 网络首发时间: 2025-05-29

会话指用户在一定时间内与网络平台或网络应用的连续交互,如在电商平台的一次购物过程、在音乐应用中的系列播放行为等。会话推荐(session recommendation)是根据用户已有的网上行为去预测用户下一个可能的行为,如电商平台上的下一个商品推荐、音乐应用中的下一首歌曲推荐等^[1]。与基于用户长期静态偏好建模的传统推荐方法不同,由于没有用户注册登录信息^[2],会话推荐只得尽可能地在当前有限的交互行为中对匿名用户动态短期兴趣(即会话偏好)进行建模,能够为用户提供更加及时、准确和个性化的推荐^[3],在适用匿名用户、捕捉短期偏好、提高推荐实时性、增强用户体验方面具有重要的应用价值,广泛应用于电子商务、智能客服、线上教育、网上社交等实时性强的应用领域。

早期的会话推荐方法包括基于模式/规则方法与基于马尔可夫链方法。基于模式/规则的方法主要依据项目之间的相似度^[4]或共现模式,基于马尔可夫链的方法主要通过估计连续两个项目之间的转移概率,这两类早期方法仅关注一定步长项目之间的转移模式,无法捕捉项目之间的长期依赖^[5]。随着深度学习的兴起,凭借序列建模的优势,循环神经网络(recurrent neural network, RNN)^[6]被广泛应用于会话推荐中,然而RNN过于强调项目之间的顺序依赖,无法对项目之间复杂关系进行建模,未能充分挖掘会话中的隐含信息。后来得益于Transformer^[7]的成功,基于自注意力的会话推荐方法达到了较RNN更好的效果,但是其作为序列模型仍无法充分建模会话中的非序列化特征。最新的图神经网络(graph neural network, GNN)^[8-10]会话推荐方法通过将会话建模为图,利用图挖掘会话中隐含的项目转换模式,通过注意力机制使模型能够关注重要的项目,取得了比之前方法更好的效果,成为研究热点。

然而,已有的基于自注意力的会话推荐方法在学习项目嵌入时聚合了会话内所有项目的特征信息而没有考虑这些项目之间的相关性,即没有考虑聚合过多相关性较弱的高阶邻居信息会对项目嵌入学习造成干扰的问题。图1是一个电子商务领域的例子,表示偏好发散的购物会话序列,会话中最后两个项目都是衣服,最早两个项目都是手机,显然,衣服与手机的相关性较弱。虽然通过自注意力机制可以给弱相关项目分配较小的权重,然而,较小的权重经过堆积也会对会话中项目表示学习造成干扰,使得学习的项目表示趋于同质

化,进而影响最终推荐效果。

此外,在学习会话偏好(嵌入)时,现有的会话推荐方法都没有对会话偏好进行更深层次分析,往往过于关注会话末尾的项目,简单地认为越靠近会话末尾的项目越能代表会话偏好^[11,12]。然而,现实情况是,不同用户间的会话偏好不尽相同,同一个用户不同时刻的会话偏好也不一样。在有些场合,会话中的项目嵌入表示相对比较集中,如图2的购物会话序列所示,会话中的项目均与衣服有关,我们把这种情况称之为会话偏好集中;在另外一些场合,会话中的项目分布比较分散,如图1所示,会话中的项目有衣服,也有手机,还有书籍,我们把这种情况称之为会话偏好发散。在会话偏好集中的情况下,把会话末尾项目作为会话偏好比较合理;反过来,在会话偏好发散的情况下,如果再把会话末尾的项目作为会话偏好就不合理了,此时应综合考虑整个会话的所有项目。



图1 会话偏好发散情况



图2 会话偏好集中情况

自注意力机制能够有效捕捉会话序列中任意位置项目之间的依赖关系,能够对会话中的长距离依赖进行建模,并且相较于循环神经网络(RNN),自注意力机制具有更大的感受野,能够获取更多的上下文信息,使得模型能够更好地理解会话的整体结构和用户的整体兴趣。此外,通过将会话序列建模为图结构,可以更加方便地判断项目之间的相关性,这些在图上距离较近的项目之间往往具有较高相关性。故针对以上问题,本文提出基于图注意力和会话偏好识别的会话推荐模型(session recommendation based on graph attention and session preference recognition, SR-GASPR)。该模型中的图注意力网络结合自注意力和会话的图结构信息,考虑了会话内项目之间的相关性,使模型在学习会话中项目嵌入时仅聚合会话内相关性较高项目的特征信息,从而减少会话中弱相关项目的影响,解决在基于自注意力的会话推荐方法中项目表示同质化和弱相关邻居

对项目嵌入学习造成干扰的问题。其次通过分类识别用户会话偏好类型是发散或集中,来增强会话嵌入学习能力,使模型能够更加精确地刻画用户会话偏好。

综上,本文的主要贡献如下。

(1) 提出一种基于图注意力的 μ 阶邻居节点嵌入学习方法,选取会话中强相关项目节点进行聚合,削弱弱相关项目对节点嵌入学习的干扰。

(2) 提出一种会话偏好分类识别及会话嵌入学习方法,通过计算会话末尾节点项目嵌入与会话嵌入之间的距离,将会话划分为偏好集中和偏好发散两种类型,依据会话偏好类型进行会话嵌入增强学习。

(3) 在 Nowplaying 和 Diginetica 两个公开数据集上进行详细实验,相比于基线模型,所提模型均能取得更优的效果,验证了模型的有效性。

1 相关工作

1.1 传统方法

能够处理会话推荐任务的传统方法主要包括基于模式/规则方法和马尔可夫链的方法。基于模式/规则的方法,如关联规则^[13]、基于物品的协同过滤^[14]等,主要依据项目之间的相似性或共现模式来实现会话推荐。FPMC^[15]使用马尔可夫链,为每个用户生成独特的转移矩阵。Yang 等^[16]提出的 U2CMS 模型是一个混合推荐模型,其结合协同过滤、基于内容的推荐以及马尔可夫链来推荐,主要利用高阶马尔可夫链来模拟用户项目序列中的短期偏好。Mlika 等^[17]提出的 RSMCG 模型利用马尔可夫链模拟用户的短期动态行为,基于用户过去选择的电影列表来预测用户的兴趣。在上述方法中,虽然基于马尔可夫链的方法利用了会话序列中的时序信息,但其侧重于捕捉会话中相邻项目之间的信息,提取远距离依赖信息的能力均不足。

1.2 基于深度学习的方法

相对于传统方法,基于深度学习的会话推荐方法达到了更好的效果。基于深度学习的会话推荐方法主要包括基于循环神经网络(RNN)、基于图神经网络(GNN)以及基于自注意力(SAN)的方法。

GRU4Rec^[6]是第1个利用循环神经网络(RNN)来学习会话序列中项目转换的方法,其通过使用多层门控循环单元(gated recurrent unit)对用户会话序列进行建模,来学习项目表示。Li 等^[18]提出了 NARM 模型,参考 GRU4Rec 利用门控循环单元 GRU 来处理用户点击

序列,并引入注意力机制增强模型对用户会话偏好的捕捉能力。Bach 等^[19]提出的 RecConRec 模型结合 CNN 的局部特征提取能力与 RNN 的长期依赖关系建模能力,构建了一个统一的网络结构来捕捉会话序列中的不同特征信息。Wang 等^[20]提出的 CSRMM 模型通过结合当前会话信息和邻域会话的协同信息来提高推荐效果。虽然基于 RNN 的方法利用了时序信息,但其并未充分利用项目之间复杂的转换信息。

基于图神经网络(GNN)的方法通过将会话序列构建为图,利用消息传递机制捕获项目之间复杂的转换信息,取得了更加出色的推荐效果。Wu 等^[21]提出了 SR-GNN 模型,其将会话序列构建为图,充分利用了项目之间复杂的转换关系,并且使用注意力机制结合用户的全局兴趣和当前兴趣生成会话表示,但其固定把会话末项当作用户的当前兴趣,在用户偏好发散的情况下,模型的预测结果会不准确。Li 等^[22]提出的 Disen-GNN 模型将项目表示分解为多个因素的向量表示,通过捕捉会话中影响用户意图的细粒度因素来提高推荐准确性。Wang 等^[23]提出的 GCE-GNN 模型利用多个会话序列生成全局图,利用单个会话生成会话图,使得模型能够捕获更加复杂的跨会话的全局信息以及会话内的局部信息,并且利用反向位置编码学习会话表示,使得模型更加关注靠近会话末尾的项目。Yang 等^[24]提出 TRHMCI-GNN 模型,其构建了一个包含跨会话项目信息和隐藏主题关系的异构图来挖掘项目间的隐藏关系,并且构建了自环星图来捕捉会话内部复杂的项目转换关系。Qiao 等^[25]提出的 TASI-GNN 模型首先利用单层的 GGNN 融合邻居节点信息来学习项目表示,然后使用自适应的稀疏转换函数来削弱噪声项目对推荐性能的影响,最后利用相似会话的信息来增强会话表示。Zhang 等^[26]提出的 CARES 模型利用不同的会话构建全局图,能够提取跨会话信息,并且在考虑项目之间的转换模式时结合了项目之间的类别信息。Chen 等^[27]提出的 CT-GNN 模型基于图神经网络,通过引入时间信息,能够更准确地捕获用户偏好随时间的变化。Wang 等^[28]提出的 SRGI 模型利用全局物品转换信息来增强会话级别的推荐过程。Zhang 等^[29]提出的 MCGNN 模型认为项目预测和类别预测是密切相关的,并将二者分别作为联合多任务学习(joint multi-task learning)框架内的主任务和辅任务,来提高推荐效果。吴永庆等^[30]提出的 ATSGCN 模型构建了3种视图,利用图卷积和

注意力机制学习不同级别的物品间的转换关系, 并通过引入自监督任务实现数据增强, 缓解了会话推荐系统存在的数据稀疏性问题. 虽然可以聚合高阶的邻居信息, 但这些方法大部分仅在聚合低阶邻居信息时获得了最优的结果, 因此聚合弱相关高阶邻居的信息可能会对推荐效果造成损害.

自从 Transformer 模型被提出以来, 因其自注意力机制对序列数据具有出色的处理能力, 已被应用在诸如机器翻译、语音识别以及图像处理等多个领域并且取得了显著的成就. Wang 等^[31]提出 EAT-SGNN 模型, 通过引入自监督学习解决会话推荐中的数据稀疏问题, 提出增强注意力模块利用全局级图和会话级图的知识, 并使用 Transformer 增强会话意图表示. Xu 等^[32]提出的 GC-SAN 模型通过图神经网络捕捉会话序列中项目之间的局部依赖信息, 然后应用自注意力机制捕捉会话序列中项目之间的长距离依赖信息. Luo 等^[33]提出的 CoSAN 模型利用相似会话的协同信息生成项目表示, 并利用自注意力机制学习项目之间的全局依赖关系, 生成会话表示. Zhu 等^[34]提出的 GCAN 模型利用自注意力机制和协同注意力机制学习项目表示. 文献^[35]利用 Transformer 捕捉长距离依赖信息, 利用 CNN 提取局部特征信息, 以生成更准确的会话表示. 然而, 在学习项目表示时这些使用自注意力的方法大部分都聚

合了会话内所有其他项目的特征信息, 并且也没有对用户会话偏好的类型进行深入探究.

2 SR-GASPR 模型

本文提出的模型如图 3 所示, 分为会话图构建、节点项目嵌入学习、会话偏好识别与嵌入学习、推荐预测这 4 个部分. 首先在会话图构建部分, 根据会话序列中项目之间的相邻关系构建会话图, 并根据会话图构建邻居选择矩阵. 其次在节点项目嵌入学习部分, 相较于已有的基于自注意力的会话推荐方法, 本文提出的图注意力网络创造性地将会话的图结构信息融入自注意力机制中, 在学习会话内项目嵌入时, 根据邻居选择矩阵仅融合相关性较强的图邻居节点的信息, 进而摒除弱相关项目对项目嵌入学习的干扰, 从而提高模型的推荐效果. 然后在会话偏好识别与嵌入学习部分, 首先对会话内项目嵌入求平均以及取末项分别作为 *global* (全局兴趣) 与 *local* (当前兴趣), 然后构建会话偏好识别模块使模型能够更细粒度地识别用户的会话偏好类型, 判断其是倾向于集中还是发散, 并根据不同类型对会话偏好增强, 然后在会话嵌入学习模块中利用软注意力机制使模型更关注会话序列中符合用户偏好的项目来学习会话嵌入. 最终在推荐预测部分预测得到推荐列表.

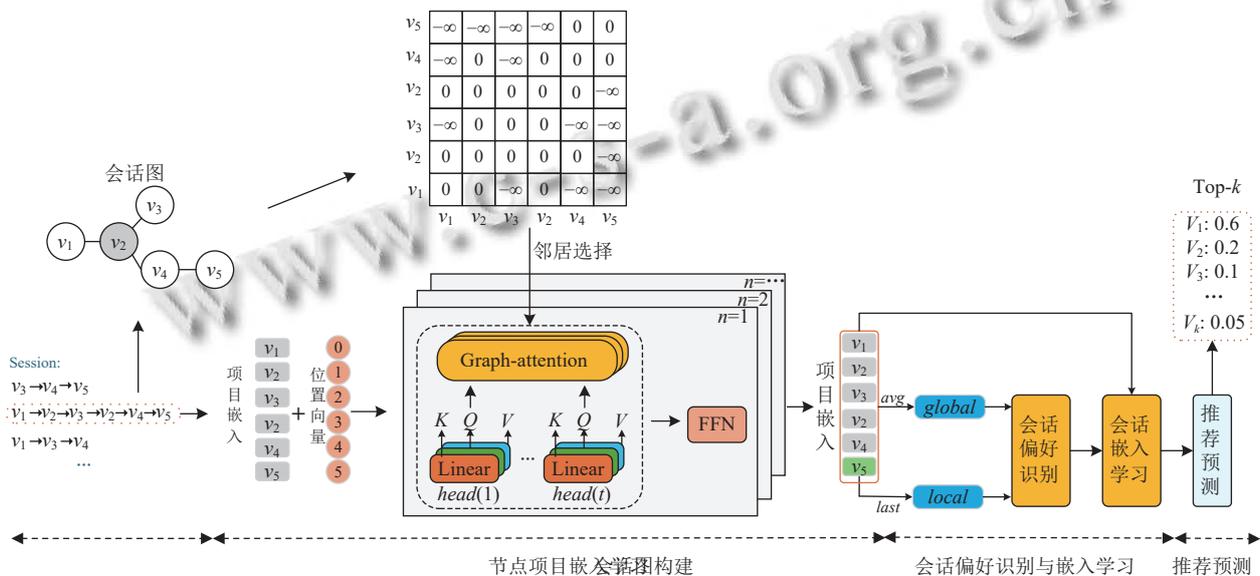


图 3 SR-GASPR 模型示意图

2.1 问题描述

假设在推荐系统中存在 n 种不同的项目, m 个会

话序列. 使用集合 $V=[v_1, v_2, v_3, \dots, v_n]$ 表示项目集, 使用集合 $S=[s_1, s_2, s_3, \dots, s_m]$ 表示推荐系统中包含的所有

会话序列. 会话序列集中包含的第 i 个会话可表示为 $s_i=[v_{si1}, v_{si2}, v_{si3}, \dots, v_{sil}]$, $i \in [1, m]$, $l \in [1, n]$, 其中, l 代表此会话序列的长度.

假设会话 s_i 是用户当前正在交互的会话, s_i 中的项目序列代表用户已点击的项目, 则推荐系统的目标是预测下一个时刻用户可能交互的第 $l+1$ 个项目. 具体地说, 推荐系统的目标就是根据用户已点击的项目序列, 选取模型预测概率最大的前 Top- k ($k \in [1, n]$) 个预测项目作为推荐结果. 表 1 给出了相关符号表示及对应含义说明.

表 1 符号表示及含义说明

符号	含义
V	所有会话包含的项目集
S	所有会话序列
G_s	会话构成的无向无权图
A	μ 阶邻居选择矩阵
X	融合了可学习位置编码的初始项目嵌入
O	聚合多头注意力输出得到的项目嵌入
E	单层图注意力输出的项目嵌入
$E^{(n)}$	n 层图注意力输出的项目嵌入
$local$	会话当前偏好
$global$	会话全局偏好
T	会话当前偏好与全局偏好的相似度判断阈值
I	增强后的用户偏好
z	会话嵌入

2.2 会话图构建

图 3 中包含一个具体的例子. 假设给定一个会话序列 $s=[v_1, v_2, v_3, \dots, v_l]$, 用户在同一个会话中可能会多次点击同一个项目, 所以会话序列中可能存在重复的点击项. 将会话序列构成的无向无权图表示为 $G_s = \langle V_s, E_s \rangle$, 其中 $V_s \in V$ 代表会话序列 s 中独特的项目, E_s 代表会话图的边集. 会话中的点击序列在时间上是有先后顺序的, 相邻的点击项目在点击时间上也是相邻的, 所以对于会话内的点击项, 其相邻的点击项与其自身十分相关. 故针对会话序列内的任意两个项目, 若二者是相邻的, 则二者之间存在无向边.

2.3 节点项目嵌入学习

由于现存基于自注意力的会话推荐方法在学习项目嵌入时融合会话内弱相关邻居信息会损害模型推荐效果, 所以提出图注意力网络, 创新地将会话的图结构信息引入到自注意力计算中, 利用图来为每个项目选择会话内要融合的相关性较强的项目. 对于相关性我们认为会话图中任意两个项目之间的距离越远相关性越弱, 距离越近则相关性越强.

2.3.1 μ -neighbor 选择矩阵构建

图 3 中包含一个具体的例子, 其中 $\mu=1$. 以 v_2 为例, 每个节点都可能不同阶数的邻居. 我们用 A 表示 μ -neighbor 选择矩阵, 每个会话 s 对应一个矩阵 A . 对于矩阵 A 中的任意一个元素 A_{ij} , 其表示对于点击项目 v_{si} 来说, 在编码阶段是否融合点击项目 v_{sj} 的特征信息. 我们利用无向图的广度优先搜索算法思想, 遍历得到与点击项目 v_{si} 的特征信息更为相关的 μ 阶邻居节点, 按照式 (1), 若 v_{sj} 是点击项目 v_{si} 的 μ 阶以内的邻居节点, 则 A_{ij} 为 0, 否则 A_{ij} 为 $-\infty$. $hop(i, j)$ 表示 v_{si} 的邻居 v_{sj} 的阶数. 总之, 在学习项目表示时, 模型只会融合其 μ 阶之内邻居节点的特征信息, 会话序列内其他的项目则不考虑.

$$A_{i,j} = \begin{cases} -\infty, & \text{if } hop(i, j) \in [\mu, \infty] \\ 0, & \text{if } hop(i, j) \in [1, \mu] \end{cases} \quad (1)$$

2.3.2 项目嵌入初始化层

在一个会话中, 用户的兴趣可能是波动的, 不同位置的项目可能反映了用户不同的兴趣, 所以项目的位置信息十分重要^[36]. 通过向初始项目表示添加位置编码, 可以更好地提取会话的时序信息. 本文采用一个可学习的位置编码来更好提取时序信息. 给定一个会话序列 $s=[v_1, v_2, v_3, \dots, v_l]$, 经过编码层编码得到会话内项目的初始表示 $H=[h_1, h_2, \dots, h_l]$, $H \in R^{l \times d}$.

$$X = \text{sum}(H, P_t) \quad (2)$$

其中, $X=[x_1, x_2, \dots, x_l]$, $X \in R^{l \times d}$ 表示融合了可学习位置编码的项目表示, 并将其作为模型的输入. sum 表示向量相加. $P_t \in R^{l \times d}$, $t \in [1, l]$, 表示可学习的位置编码, d 是向量维度, l 是会话长度.

2.3.3 μ -neighbor 注意力计算层

对于输入的项目表示矩阵, 通过对会话序列中的每个项目表示做线性变换, 可分别得到它们的键向量 (Key)、查询向量 (Query) 和值向量 (Value), 并分别用 K 、 Q 、 $V \in R^{l \times d}$ 表示.

$$Q = W^Q \cdot X, K = W^K \cdot X, V = W^V \cdot X \quad (3)$$

其中, W^Q 、 W^K 、 $W^V \in R^{d \times d}$ 是映射矩阵.

在学习项目表示时, 依次将每个项目表示作为查询向量, 将会话内的其他项目表示作为键向量, 通过计算上述二者内积可以判断相似性. 对于会话内的其他项目表示, 若与它们的内积越大则越相似, 内积越小则

越不相似. 为了稳定梯度, 将点积的结果除以 \sqrt{d} , 其中, d 是项目表示的维度. 最终得到会话内所有项目表示与会话内其他项目表示的相似性分数, 并用矩阵 $B \in R^{l \times l}$ 进行表示:

$$B = \frac{Q \cdot K^T}{\sqrt{d}} \quad (4)$$

利用 μ -neighbor 选择矩阵 $A \in R^{l \times l}$, 对会话内任一项目 x_i , 若 x_j 是 x_i 的 μ 阶以内的邻居, A 中则对应零向量, 经式 (5) 计算后不改变二者之间的相似分数, 否则, A 中则对应负无穷小向量, 经式 (5) 计算将二者之间的相似分数设为一个无穷小值, 经 *Softmax* 计算会得到基本为 0 的权重, 以此消除 μ 阶以外的邻居的影响 (μ 的值通过实验确定). 这样在学习每个项目嵌入时, 仅融合了与该项目相关性较强的图邻居项目的特征信息. 从会话序列的角度看, 即仅融合了子序列内的项目的特征信息, 而非会话内的所有项目. 以此来消除弱相关项目的影响. 经过 *Softmax* 函数归一化处理, 得到会话内其他项目的融合权重 β . 最后融合会话内邻居项目信息得到最终的项目表示 $C \in R^{l \times d}$.

$$\beta = \text{Softmax}(A + B) \quad (5)$$

$$C = \beta \cdot V \quad (6)$$

2.3.4 多头自注意力

采用多头注意力机制, 并行使用 t 个具有不同参数的独立注意力头学习更丰富的项目表示. 把这些注意力模型的所有输出连接起来, 生成项目表示 $O \in R^{l \times d}$.

$$O = W_1 \text{Concat}(C_1, C_2, \dots, C_t) \quad (7)$$

$$C_i = \text{Softmax} \left(\frac{(XW_i^Q)(XW_i^K)^T}{\sqrt{d}} + A \right) (XW_i^V) \quad (8)$$

其中, $W_1 \in R^l$ 是转换矩阵, C_i 是第 i 个注意力头的输出.

2.3.5 前馈神经网络 (feedforward neural network, FFN)

FFN 层由两层 MLP 和一个残差连接组成, 其赋予了模型非线性特性. 采用残差连接, 使模型能够充分学习低层信息, 缓解可能在深层网络结构中出现的梯度消失问题. 此外, 采用层归一化操作稳定训练, 采用随机 *Dropout* 缓解堆叠多层网络带来的过拟合问题.

$$\tilde{C} = \text{Dropout}(O) \quad (9)$$

$$D = \text{ReLU}(\tilde{C}W_2 + b_1)W_3 + b_2 \quad (10)$$

$$E = \text{LayerNorm}(D + \tilde{C}) \quad (11)$$

其中, $E \in R^{l \times d}$ 表示最终的项目表示. W_2 、 W_3 是可学习参数, b_1 、 b_2 是偏置.

2.3.6 多层自注意力

不同层可以捕获不同类型特征信息, 故堆叠多层自注意力模块来更充分地提取会话内的项目转换信息^[32]. 简单起见, 将上述自注意力计算步骤表示为 *SAN*:

$$E = \text{SAN}(X) \quad (12)$$

其中, $X \in R^{l \times d}$ 表示融合了可学习位置编码的初始项目表示, E 是式 (11) 的结果, 代表项目表示. 因此, 第 n 层自注意力模块可被表示为:

$$E^{(n)} = \text{SAN}(E^{(n-1)}) \quad (13)$$

其中, $E^{(n)} = [e_1, e_2, \dots, e_l]$, $E^{(n)} \in R^{l \times d}$ 代表会话中的项目嵌入.

2.4 会话偏好识别与嵌入学习

用户会话偏好的类型不是唯一的, 所以提出会话偏好识别与嵌入学习方法, 使模型能够动态识别用户的会话偏好是集中类型还是发散类型, 并在学习会话嵌入时根据会话偏好类型关注会话序列中用户真正感兴趣的部分, 而不像现有方法 (如结合反向位置编码) 简单地认为越靠近会话末尾的项目越能代表用户的兴趣, 使模型能够更加准确地预测用户兴趣.

2.4.1 会话偏好感知层

在会话中, 用户即时偏好会随着时间的推移而发生改变, 最近的点击项目反映了用户的瞬时偏好, 整个会话序列的点击项目则反映了用户的全局偏好. 基于此, 会话偏好感知层首先将多层自注意力模块最后一层输出的会话尾节点嵌入 e_l 视作当前偏好, 而将该网络末层输出的会话平均节点嵌入视作全局偏好.

$$\text{local} = \text{Last}(E^{(n)}) \quad (14)$$

$$\text{global} = \text{Avg}(E^{(n)}) \quad (15)$$

其中, *local* 表示会话当前偏好, *global* 表示会话全局偏好, *Last* 表示取会话末尾的项目嵌入, *Avg* 表示求会话中所有项目嵌入的平均.

2.4.2 会话偏好增强层

利用余弦相似度 \cos 计算当前偏好与全局偏好的相似程度 *DOS* (degree of similarity).

$$\text{DOS}(\text{local}, \text{global}) = \cos(\text{local}, \text{global}) \quad (16)$$

$$\cos(x, y) = \frac{x^T y}{\|x\| \|y\|} \quad (17)$$

阈值 T (threshold) 是准确判断用户会话偏好的关键, 选取合适的阈值对提高模型推荐效果至关重要. 以所有会话当前偏好与全局偏好的相似度平均值作为基准, 判断用户会话的偏好类型.

$$T = \varepsilon \times \frac{\sum_{i=1}^{|\mathcal{S}|} DOS(local, global)}{|\mathcal{S}|} \quad (18)$$

其中, $|\mathcal{S}|$ 是会话个数, ε 是相似度控制系数. 不同数据集中的会话数据特点不同, 因此为每个数据集选取合适的相似度控制系数 ε , 以达到更优的推荐效果.

如果用户在当前会话的全局偏好与当前偏好的相似度 $DOS(local, global)$ 大于阈值, 表示用户会话偏好比较趋近于集中. 由于末尾项目最能代表用户当前偏好, 故将全局偏好与当前偏好相结合使模型更加关注靠近会话末尾的项目; 如果用户在当前会话的全局偏好与当前偏好的相似度 $DOS(local, global)$ 小于阈值, 则表示用户会话偏好比较趋近于发散. 这种情况下不应过多关注当前偏好, 全局偏好更能代表用户的会话偏好.

$$I = \begin{cases} W(local \parallel global) + b, & DOS(local, global) > T \\ global, & DOS(local, global) \leq T \end{cases} \quad (19)$$

其中, $I \in \mathbb{R}^{1 \times d}$ 表示增强后的用户偏好, \parallel 表示拼接操作, T 表示阈值, W 为转换矩阵, b 为偏置.

2.4.3 会话嵌入学习层

会话内的项目对会话表示的学习重要性是不同的, 因此我们采用软注意力机制合理分配权重, 使模型更加关注会话内更符合用户会话偏好的项目. 最终的会话表示由会话内所有项目表示按权重相加得到.

$$\alpha_i = f^T \text{Sigmoid}(W^a I + W^b e_i + b_3) \quad (20)$$

$$z = \sum_{i=1}^n \alpha_i e_i \quad (21)$$

其中, W^a 、 $W^b \in \mathbb{R}^{d \times d}$ 与 $f^T \in \mathbb{R}^d$ 均是可学习参数, $b_3 \in \mathbb{R}^d$ 是偏置向量, e_i 和 I 分别表示会话内的第 i 个项目表示和用户的增强会话偏好, z 表示最终的会话表示, n 代表会话内的项目个数.

2.5 预测模块

利用当前会话表示 z 与候选项目表示 h_i 的点积来计算候选项目的推荐概率, 并采用 *Softmax* 函数进行归

一化以获得输出.

$$\hat{y} = \text{Softmax}(z^T h_i) \quad (22)$$

其中, \hat{y} 表示预测概率.

采用标签平滑, 以缓解模型训练过程中可能出现的梯度爆炸或梯度消失问题, 提升模型的泛化能力并使训练过程更为稳定.

$$\tilde{y} = y(1 - \theta) + \frac{1}{N} \theta \quad (23)$$

其中, θ 是超参数, y 是项目的 one-hot 向量表示, N 表示项目种类数.

最终, 采用交叉熵损失函数训练模型.

$$\ell = - \sum_{i=1}^N \tilde{y}_i \log(\hat{y}_i) + (1 - \tilde{y}_i) \log(1 - \hat{y}_i) \quad (24)$$

3 实验

本节首先介绍实验用到的数据集, 然后介绍要对比的会话推荐领域的基线方法, 再从以下方面探究本文所提模型的效果.

问题 1: SR-GASPR 的整体推荐效果如何?

问题 2: 本文设计的图注意力网络是否有效? 是否能够解决现存使用自注意力的会话推荐方法中融合弱相关项目信息对项目嵌入学习造成干扰的问题?

问题 3: 本文设计的用户会话偏好识别与嵌入学习方法是否有效? 与现存会话嵌入学习方法相比是否更具优势?

问题 4: 相关参数对模型推荐效果有什么影响?

3.1 数据集与数据处理

3.1.1 数据集

Diginetica 数据集出自 CIKM2016 电子商务推荐挑战赛, 它包含用户在一定时间之内的点击日志, 如商品的详细信息、用户的点击和购买记录、会话信息等. Nowplaying 数据集是一个用于音乐推荐和分析的公开数据集, 主要来源于社交媒体 Twitter, 此数据集包含用户在 Twitter 上发布的音乐播放记录. 两个数据集的网址分别为 Diginetica (<https://competitions.codalab.org/competitions/11161>) 和 Nowplaying (<https://dbis.uibk.ac.at/node/263#nowplaying>).

3.1.2 数据处理方法

参照文献[21]中的数据处理方法, 删除数据集中长度为 1 的会话以及在数据集中出现总次数小于 5 的项

目. 取数据集最近一周的数据作为测试集, 余下的为训练集. 为扩充数据集, 对于每个会话序列 $s=[v_1, v_2, v_3, \dots, v_m]$, 将其分为 $([v_1], v_2), ([v_1, v_2], v_3), \dots, ([v_1, v_2, \dots, v_{m-1}], v_m)$, 其中每个序列的最后一项作为标签. 数据集的详细信息如表 2 所示.

表 2 数据统计

属性	Diginetica	Nowplaying
项目数量	43 097	60 417
训练会话数量	719 470	825 304
测试会话数量	60 858	89 824
会话平均长度	5.12	7.42

3.2 基线模型介绍

为了验证本文工作的有效性, 我们选取多种会话推荐模型与本文模型进行比较. 这些基线模型包括传统方法、基于循环神经网络的方法、基于注意力的方法以及基于图神经网络的方法.

POP: 推荐训练集中出现次数较多的流行项目.

Item-KNN^[4]: 利用余弦计算两个项目之间的相似度, 并推荐相似度较高的项目.

FPMC^[15]: 利用矩阵分解和马尔可夫链进行推荐.

GRU4Rec^[6]: 基于 RNN 的模型, 引入门控机制, 处理会话序列中的序列信息.

NARM^[18]: 利用门控循环单元处理会话序列, 使用注意力机制来加强模型对用户会话偏好的捕捉能力.

STAMP^[37]: 利用注意力网络计算长期记忆, 将会话序列的末项当作短期记忆, 将二者结合进行推荐.

SR-GNN^[21]: 将会话序列构建为图, 利用门控图神经网络捕获项目之间的复杂转换信息.

Disen-GNN^[22]: 利用多个因素, 学习项目更加细粒度的向量表示.

TRHMCI-GNN^[24]: 构建主题关系异构跨项目全局图和自环星图, 分别捕获会话间和会话内的跨项目信息.

3.3 评价指标和参数设置

精确率 $P@N$ 广泛应用于衡量会话推荐系统预测结果的准确性, 表示在前 N 个推荐项目中, 正确推荐的项目所占的比例. 计算公式如下:

$$P@N = \frac{n_{\text{hit}}}{N_{\text{test}}} \quad (25)$$

其中, N_{test} 是测试集样本总数, n_{hit} 是前 N 个推荐项目中包含的正确推荐项目的数量.

平均倒数排名 $MRR@N$ 表示所有正确推荐项目在推荐列表中排名的倒数的平均值. 值越大意味着模型

在推荐时能够更快地预测正确结果. 计算公式为:

$$MRR@N = \frac{1}{n} \sum_{i \in N_{\text{hit}}} \frac{1}{\text{Rank}(i)} \quad (26)$$

其中, n 是测试集样本总数, N_{hit} 是前 N 个推荐项目中包含正确项目的样本集, $\text{Rank}(i)$ 是项目 i 在推荐列表中的排名.

通过多次实验确定模型的最优参数如下: 初始学习率为 0.002, 且每经过 3 轮训练学习率衰减 0.1; 隐向量维度设为 100; 采用 Adam 优化器; 将会话长度限定为 20, 即若原始会话长度大于 20 则截取靠近末尾的 20 个项目, 小于 20 则在会话前填充零向量到 20; 模型参数均以均值为 0、标准差为 0.05 的正态分布进行初始化. 其他参数通过在两个数据集上进行多次实验得到.

3.4 对比实验结果分析 (问题 1)

首先分析本文所提模型 SR-GASPR 的整体推荐效果. 表 3 列出了在 Nowplaying 和 Diginetica 两个数据集上本文模型与基线模型的对比实验结果, 加粗数据表示最优结果, 加下划线数据表示次优结果.

综合来说, 基于传统方法的推荐模型推荐效果弱于基于深度学习方法的模型, 这证明了深度学习方法处理会话推荐任务的有效性. 传统方法在推荐过程中仅利用了会话序列中非常有限的信息, 导致其无法准确捕获用户会话偏好.

在传统方法中, POP 模型的结果最差, 仅依靠流行度推荐并不能满足用户的个性化需求. FPMC 利用矩阵分解和马尔可夫链进行推荐, 相比 POP 取得了更好的结果. Item-KNN 利用相似度进行推荐, 取得了最好的结果, 但它无法捕获项目之间的顺序转换信息.

GRU4Rec 的效果是基于 RNN 的模型中最差的, 因为用户兴趣是波动的, 会话序列不只包含了序列信息, RNN 作为序列模型无法充分捕获这种复杂信息.

NARM 的性能优于 GRU4Rec, 它在使用 RNN 的基础上额外引入了注意力机制来动态选择输入序列的不同部分, 这使得模型能够关注更重要的项目, 从而更准确地捕捉用户的会话偏好. STAMP 的性能同样优于 GRU4Rec, 其将会话序列的末项当作短期记忆, 并完全基于注意力机制. 这证明引入注意力机制来处理会话推荐任务是有效的, 因为在会话内用户的兴趣可能会随着时间的推移而发生改变, 模型应当关注用户更感兴趣的项目.

表3 对比实验结果

Method	Nowplaying				Diginetica			
	$P@10$	$P@20$	$MRR@10$	$MRR@20$	$P@10$	$P@20$	$MRR@10$	$MRR@20$
POP	1.86	2.28	0.83	0.86	0.76	1.18	0.26	0.28
Item-KNN	10.96	15.94	4.55	4.91	25.07	35.75	10.77	11.57
FPMC	5.28	7.36	2.68	2.82	15.43	22.14	6.2	6.66
GRU4Rec	6.74	7.92	4.4	4.48	17.93	30.79	7.73	8.22
NARM	13.6	18.59	6.62	6.93	35.44	48.32	15.13	16
STAMP	13.22	17.66	6.57	6.88	33.98	46.62	14.26	15.13
SR-GNN	14.02	18.67	7.24	7.54	38.53	51.6	16.99	17.89
Disen-GNN	15.95	21.18	<u>8.04</u>	<u>8.39</u>	<u>40.55</u>	53.57	<u>18.04</u>	<u>18.94</u>
TRHMCI-GNN	<u>15.99</u>	<u>21.89</u>	7.1	7.5	40.38	<u>53.86</u>	17.5	18.43
Ours	17.36	23.1	8.16	8.55	41.22	54.32	18.83	19.73

基于 GNN 的模型性能高于基于 RNN 的序列模型, 这证明了融合项目之间复杂转换信息对推荐的有效性. SR-GNN 将会话序列建模为图, 利用图结构信息提取会话序列内更为复杂的项目转换信息, 取得了不错结果. 同样基于 GNN 的模型 Disen-GNN 和 TRHMCI-GNN 的性能高于 SR-GNN. Disen-GNN 用多个因素表示项目, 从更加细粒度的层次提取会话序列中的特征信息, 这证明在会话推荐任务中考虑项目因素对捕获用户会话偏好是有效的. TRHMCI-GNN 构建主题关系异构跨项目全局图和自环星图, 分别用于捕获会话间和会话内项目之间的信息, 其证明了挖掘不同会话之间的高阶隐藏关系对提高推荐效果的有效性. 不同的模型可以从不同角度构建多种类型的图结构来解决对应问题, 这证明了图神经网络 (GNN) 的灵活性和会话的图结构信息对推荐的有效性.

本文提出的模型在利用自注意力机制学习项目表示的同时引入了会话的图结构信息, 在项目表示学习时削弱了融合弱相关项目特征信息的干扰. 另外, 通过对用户会话偏好进行更细粒度的分类识别, 提高了模型对用户真实偏好的捕获能力. 本文提出的模型对比所有基线模型达到了最优实验结果. 特别是相较于 Diginetica 数据集, 本文模型在 Nowplaying 数据集上的表现尤为突出, 相较于基线模型取得了更加显著的提升, 这是因为 Nowplaying 数据集中会话序列长度更长, 用户的兴趣更容易发生改变, 这证明了本文模型在处理长会话推荐任务中的有效性.

3.5 消融实验 (问题 2 和问题 3)

本节进行消融实验, 验证所提方法各模块的有效性.

3.5.1 验证图注意力网络的有效性 (问题 2)

现有的基于自注意力的会话推荐方法大多在学习项目嵌入时融合了会话内所有项目的特征信息, 没有

考虑弱相关项目的影响. 故为验证本文设计的图注意力网络能否摒除弱相关项目对项目嵌入学习的影响, 在利用图注意力网络学习项目嵌入时, 通过逐步增加融合图邻居项目的阶数 (如图 4), 来逐步增加融合的项目数量, 一般而言, 阶数越高的图邻居项目相关性越弱, 所以融合的弱相关项目的数量也会逐步增加.

如图 4, 邻居阶数为 1 则只融合会话图中的 1 阶邻居; 邻居阶数为 2 则融合会话图中的 1 阶和 2 阶邻居, 之后以此类推. 观察 $MRR@20$ 指标, 在 Nowplaying 和 Diginetica 数据集上只融合一阶邻居的效果是最好的, 观察 $P@20$ 指标, 在 Nowplaying 和 Diginetica 数据集上分别是只融合一阶邻居和一、二阶邻居效果是最好的.

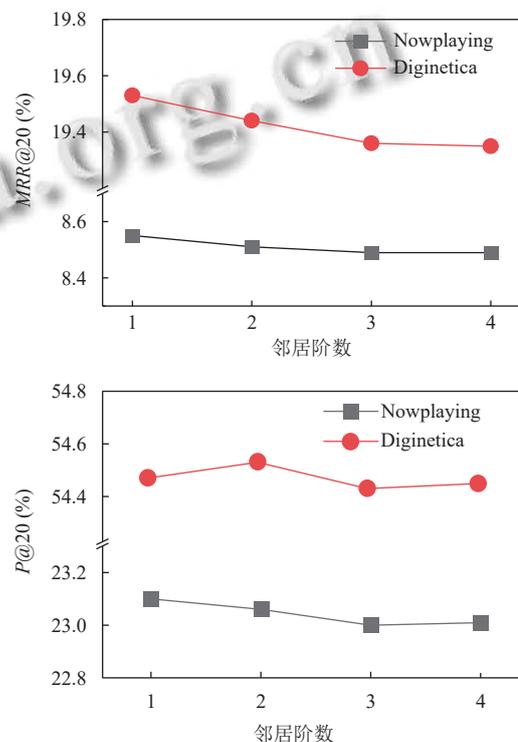


图4 融合不同数目邻居项目结果图

整体来看,在这两个数据集上通过图注意力网络学习项目嵌入时,均是在融合更近的、相关性更高的图邻居项目特征信息时达到最优推荐结果,而随着融合更多高阶弱相关图邻居项目特征信息时模型的推荐效果均会下降.这说明,已有的基于自注意力的推荐方法在学习项目嵌入时融合会话内所有的项目,这种做法确实会引入弱相关项目导致推荐效果受损.而本文提出的图注意力网络利用会话的图结构信息,能够通过动态调整融合图邻居项目的阶数来减少弱相关项目对项目嵌入学习的干扰,从而提高推荐效果,验证了本文所提图注意力网络的有效性.

3.5.2 验证会话偏好识别与嵌入学习方法的有效性(问题3)

为验证会话偏好识别与嵌入学习方法的有效性,去除 SR-GASPR 中的会话偏好识别与嵌入学习部分,并采用当下流行的反向位置编码作为替换,将此变体命名为 RP (reversed position).

反向位置编码的使用:去除本文模型中的用户会话偏好识别模块,在经节点项目嵌入学习模块得到会话中的项目表示后,将项目表示与反向位置编码融合,然后经过会话嵌入学习层后参与预测.流程如下:

$$u_i = \tanh(W^c[e_i \parallel p_{l-i+1}] + b_4) \quad (27)$$

$$\beta_i = q^T \text{Sigmoid}(W^d u_i + W^e \text{global} + b_5) \quad (28)$$

$$z = \sum_{i=1}^n \beta_i u_i \quad (29)$$

其中, u_i 表示融合了反向位置编码的项目表示, \parallel 表示拼接操作, β_i 表示在会话融合时会话内项目对应的权重, z 是会话表示, p_{l-i+1} 是可学习的位置编码, W^c 、 W^d 、 W^e 和 q^T 是可学习的参数, b_4 和 b_5 是偏置.

实验结果如图 5, Ours 表示本文模型 SR-GASPR, 在 $P@20$ 与 $MRR@20$ 两个指标上, SR-GASPR 的性能均优于 RP 变体,这是因为当用户会话偏好是发散的, SR-GASPR 能够动态识别会话偏好并关注整个会话的项目信息,而 RP 变体采用的反向位置编码固定地认为越靠近会话末尾的项目对推荐越重要,使得模型仍旧过多关注靠近会话末尾的项目,从而降低推荐效果.这证明会话偏好识别与嵌入学习方法是有效的.

3.6 探究超参数的影响(问题4)

3.6.1 相似度控制系数 ε 对推荐效果的影响

不同数据集的数据分布情况是不同的,因此本文

模型分别针对每个数据集探究相似度控制系数 ε 对模型推荐效果的影响.从 1.1 开始以 0.1 为增量,探究相似度控制系数 ε 在取值集合 $\{1.1, 1.2, 1.3, 1.4, 1.5\}$ 上时模型的推荐效果变化.

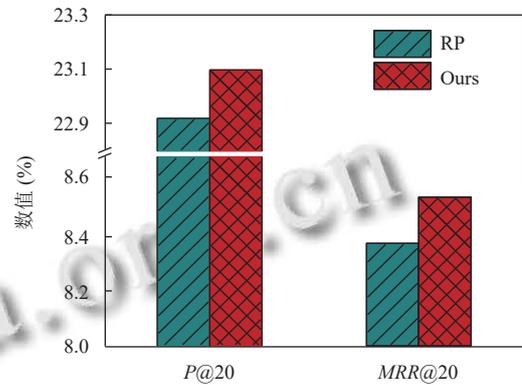


图5 Nowplaying 数据集上的消融实验结果

实验结果如图 6 所示,模型推荐的效果随着控制系数的增加,在 Diginetica 数据集上呈现下降趋势,而在 Nowplaying 数据集上呈现上升趋势.这表明,相对于 Diginetica 数据集,在 Nowplaying 数据集上需要更大的控制系数来判断当前会话中用户会话偏好是否集中,即说明 Nowplaying 数据集的会话反映的用户会话偏好更偏向于发散.这与事实相符,相较于 Diginetica 数据集,Nowplaying 数据集中的会话平均长度更长,用户的兴趣更易发生改变.

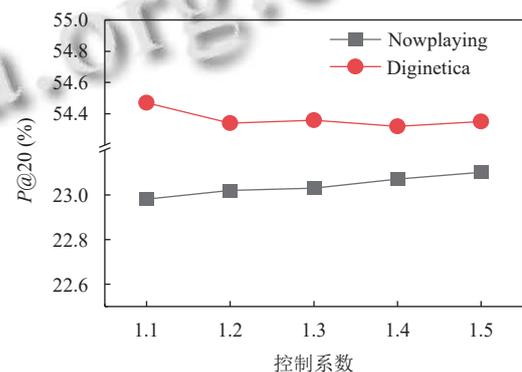


图6 相似度控制系数大小的影响

3.6.2 图注意力网络层数对推荐效果的影响

如图 7 所示,在 Nowplaying 数据集和 Diginetica 数据集上,图注意力网络层数分别为 1 和 4 时模型推荐效果最好.随着层数的增加,在两个数据集上模型的性能均出现下降,这表明堆叠过多的网络层数会增加模型过拟合的风险,从而损害模型的推荐效果.

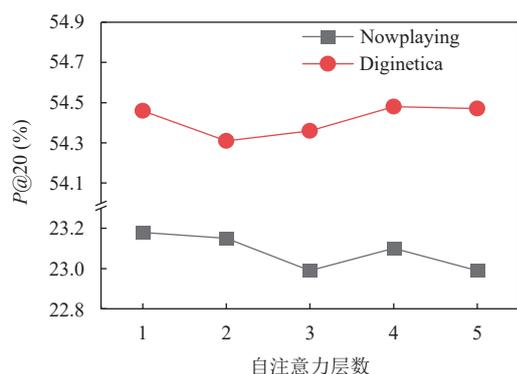


图7 图注意力网络层数的影响

4 结论与展望

为解决基于自注意力的会话推荐任务中引入弱相关项目对项目表示学习的干扰以及用户会话偏好识别不充分问题,本文提出了基于图注意力和会话偏好识别的会话推荐模型.模型将会话的图结构信息融入自注意力机制中,利用 μ -neighbor选择矩阵减少弱相关项目对项目表示学习造成的干扰;提取用户的当前兴趣和全局兴趣,通过阈值动态分类识别用户的会话偏好类型并加以增强;采用软注意力结合增强后的用户会话偏好来聚合会话中的项目表示得到会话表示,然后进行预测与推荐.本文在两个真实数据集上进行实验,相比于基线模型均达到了更优的性能,并通过消融实验验证了所提方法的有效性.项目类别信息往往能够反映用户的兴趣,在未来的研究中,通过融合类别信息,以进一步提高模型的性能.

参考文献

- 1 丁美荣,王雨航,曾碧卿.结合跳跃连接的多层图注意力网络会话推荐.计算机系统应用,2024,33(2):23–32.[doi:10.15888/j.cnki.csa.009390]
- 2 Ozbay B, Tugay R, Oguducu SG. A GNN model with adaptive weights for session-based recommendation systems. Proceedings of the 9th International Conference on Machine Learning Technologies. Oslo: ACM, 2024. 258–264.
- 3 林浩,陈平华.基于因子级特征与属性偏好联合学习的会话推荐.广东工业大学学报,2024,41(6):91–100.[doi:10.12052/gdutxb.230212]
- 4 Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms. Proceedings of the 10th International Conference on World Wide Web. Hong Kong: ACM, 2001. 285–295.

- 5 张雄涛,祝娜,郭玉慧.基于图神经网络的会话推荐方法综述.数据分析与知识发现,2024,8(2):1–16.
- 6 Hidasi B, Karatzoglou A, Baltrunas L, et al. Session-based recommendations with recurrent neural networks. arXiv: 1511.06939, 2015.
- 7 Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- 8 Liu WX, Zhang ZZ, Ding YH, et al. Global heterogeneous graph enhanced category-aware attention network for session-based recommendation. Expert Systems with Applications, 2024, 243: 122907. [doi: 10.1016/j.eswa.2023.122907]
- 9 Wan ZW, Liu X, Wang BY, et al. Spatio-temporal contrastive learning-enhanced GNNs for session-based recommendation. ACM Transactions on Information Systems, 2023, 42(2): 58.
- 10 Pang YT, Wu LF, Shen Q, et al. Heterogeneous global graph neural networks for personalized session-based recommendation. Proceedings of the 15th ACM International Conference on Web Search and Data Mining. ACM, 2022. 775–783.
- 11 Feng LX, Cai YQ, Wei EL, et al. Graph neural networks with global noise filtering for session-based recommendation. Neurocomputing, 2022, 472: 113–123. [doi: 10.1016/j.neucom.2021.11.068]
- 12 Chen JP, Li HY, Zhang XD, et al. SR-HetGNN: Session-based recommendation with heterogeneous graph neural network. Knowledge and Information Systems, 2024, 66(2): 1111–1134.
- 13 Wang SJ, Cao LB. Inferring implicit rules by learning explicit and hidden item dependency. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2020, 50(3): 935–946. [doi: 10.1109/TSMC.2017.2768547]
- 14 Ludewig M, Jannach D. Evaluation of session-based recommendation algorithms. User Modeling and User-adapted Interaction, 2018, 28(4-5): 331–390.
- 15 Rendle S, Freudenthaler C, Schmidt-Thieme L. Factorizing personalized markov chains for next-basket recommendation. Proceedings of the 19th International Conference on World Wide Web. Raleigh: ACM, 2010. 811–820.
- 16 Yang Y, Jang HJ, Kim B. A hybrid recommender system for sequential recommendation: Combining similarity models with Markov chains. IEEE Access, 2020, 8: 190136–190146.
- 17 Mlika F, Karoui W. Proposed model to intelligent

- recommendation system based on Markov chains and grouping of genres. *Procedia Computer Science*, 2020, 176: 868–877. [doi: [10.1016/j.procs.2020.09.082](https://doi.org/10.1016/j.procs.2020.09.082)]
- 18 Li J, Ren PJ, Chen ZM, *et al.* Neural attentive session-based recommendation. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. Singapore: ACM, 2017. 1419–1428.
- 19 Bach NX, Long DH, Phuong TM. Recurrent convolutional networks for session-based recommendations. *Neurocomputing*, 2020, 411: 247–258. [doi: [10.1016/j.neucom.2020.06.077](https://doi.org/10.1016/j.neucom.2020.06.077)]
- 20 Wang MR, Ren PJ, Mei L, *et al.* A collaborative session-based recommendation approach with parallel memory modules. *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. Paris: ACM, 2019. 345–354.
- 21 Wu S, Tang YY, Zhu YQ, *et al.* Session-based recommendation with graph neural networks. *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*. Honolulu: AAAI Press, 2019. 346–353.
- 22 Li AS, Cheng ZY, Liu F, *et al.* Disentangled graph neural networks for session-based recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(8): 7870–7882.
- 23 Wang ZY, Wei W, Cong G, *et al.* Global context enhanced graph neural networks for session-based recommendation. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. Xi'an: ACM, 2020. 169–178.
- 24 Yang F, Peng DL. A graph neural network with topic relation heterogeneous multi-level cross-item information for session-based recommendation. *Information Systems*, 2024, 123: 102380.
- 25 Qiao ST, Zhou W, Luo FJ, *et al.* Noise-reducing graph neural network with intent-target co-action for session-based recommendation. *Information Processing & Management*, 2023, 60(6): 103517.
- 26 Zhang ZH, Yu JX, Li X. Context-aware session-based recommendation with graph neural networks. *Proceedings of the 2023 IEEE International Conference on Knowledge Graph (ICKG)*. Shanghai: IEEE, 2023. 35–44.
- 27 Chen QZ, Jiang F, Guo XY, *et al.* Combine temporal information in session-based recommendation with graph neural networks. *Expert Systems with Applications*, 2024, 238: 121969. [doi: [10.1016/j.eswa.2023.121969](https://doi.org/10.1016/j.eswa.2023.121969)]
- 28 Wang ZY, Wei W, Zou D, *et al.* Exploring global information for session-based recommendation. *Pattern Recognition*, 2024, 145: 109911. [doi: [10.1016/j.patcog.2023.109911](https://doi.org/10.1016/j.patcog.2023.109911)]
- 29 Zhang Z, Yang B, Xu H, *et al.* Multi-level category-aware graph neural network for session-based recommendation. *Expert Systems with Applications*, 2024, 242: 122773. [doi: [10.1016/j.eswa.2023.122773](https://doi.org/10.1016/j.eswa.2023.122773)]
- 30 吴永庆, 朱月, 王钰涵. 基于注意力机制的自监督图卷积会话推荐. *计算机系统应用*, 2024, 33(5): 57–66. [doi: [10.15888/j.cnki.csa.009483](https://doi.org/10.15888/j.cnki.csa.009483)]
- 31 Wang Q, Cui H, Zhang JP, *et al.* Self-supervised global graph neural networks with enhance-attention for session-based recommendation. *Applied Soft Computing*, 2024, 150: 111026. [doi: [10.1016/j.asoc.2023.111026](https://doi.org/10.1016/j.asoc.2023.111026)]
- 32 Xu CF, Zhao PP, Liu YC, *et al.* Graph contextualized self-attention network for session-based recommendation. *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI-19)*. Macao: ijcai.org, 2019. 3940–3946.
- 33 Luo AJ, Zhao PP, Liu YC, *et al.* Collaborative self-attention network for session-based recommendation. *Proceedings of the 29th International Conference on International Joint Conferences on Artificial Intelligenc*. Yokohama: ijcai.org, 2021. 359.
- 34 Zhu XY, Zhang Y, Wang JY, *et al.* Graph-enhanced and collaborative attention networks for session-based recommendation. *Knowledge-based Systems*, 2024, 289: 111509. [doi: [10.1016/j.knsys.2024.111509](https://doi.org/10.1016/j.knsys.2024.111509)]
- 35 Wang JJ, Xie HR, Wang FL, *et al.* A Transformer-convolution model for enhanced session-based recommendation. *Neurocomputing*, 2023, 531: 21–33. [doi: [10.1016/j.neucom.2023.01.083](https://doi.org/10.1016/j.neucom.2023.01.083)]
- 36 Gan MX, Zhang XT. Integrating community interest and neighbor semantic for microblog recommendation. *International Journal of Web Services Research*, 2021, 18(2): 54–75. [doi: [10.4018/IJWSR.2021040104](https://doi.org/10.4018/IJWSR.2021040104)]
- 37 Liu Q, Zeng YF, Mokhosi R, *et al.* STAMP: Short-term attention/memory priority model for session-based recommendation. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. London: ACM, 2018. 1831–1839.

(校对责编: 王欣欣)