

自适应的邻接关系：去相关图方法在下一个兴趣点推荐的应用

王世杰, 李艳红*, 徐昊翔, 张法

(中南民族大学 计算机学院, 武汉 430074)

摘要 下一个兴趣点(POI)推荐在社交网络服务中扮演了相当重要的角色,它的任务是基于用户的历史访问足迹,预测用户在下一个时刻最可能访问的POI.现有方法普遍使用预定义图的方式为POI构建关系图,仅有少量的工作对自适应图方法进行研究,而自适应图表示学习能学习到更具潜在意义的图结构,使得图神经网络(GNN)后续的传播过程能够学习到更有意义的邻接关系,从而得到更有价值的POI嵌入,使得下游的序列建模任务能够更好地捕获到POI之间的潜在关系.然而现有工作对自适应图的研究仍处于较为初步的阶段.为下一个POI推荐任务提出了去相关图表示增强的注意力网络(DGRAN).此外,还探讨了自注意力机制与自适应图学习之间的关系,为该领域的自注意力方法加入额外残差连接以加大梯度,保证自适应图结构高质量更新.在两个真实数据集上的结果证明所提出的方法超越了现有最先进基线的性能.

关键词 下一个兴趣点推荐;自适应图;额外残差连接

中图分类号 TP399.41 **文献标志码** A **文章编号** 1672-4321(2026)02-0191-11

doi:10.20056/j.cnki.ZNMDZK.20250831

Adaptive adjacency relationships: Application of decorrelated graph methods in next POI recommendation

WANG Shijie, LI Yanhong*, XU Haoxiang, ZHANG Fa

(College of Computer Science, South-Central Minzu University, Wuhan 430074, China)

Abstract The task of next Point-of-Interest (POI) recommendation plays a significant role in social networking services. Its goal is to predict the next POI that a user is most likely to visit, based on their historical check-in records. Existing methods generally construct relational graphs for POIs using predefined graphs, while only a few studies have explored adaptive graph approaches. Adaptive graph representation learning can capture more meaningful graph structures, enabling the subsequent propagation process in Graph Neural Networks (GNNs) to learn more significant adjacency relationships, and thus obtain more valuable POI embedding. This enables downstream sequence modeling tasks to capture potential relationships between POIs better. However, research on adaptive graphs remains in its early stages. In this paper, we propose a Decorrelated Graph Representation-enhanced Attention Network (DGRAN) for the next POI recommendation task. Additionally, this paper explores the relationship between the self-attention mechanism and adaptive graph learning, introducing extra residual connections to self-attention methods in this field to increase the gradient and ensure the high quality updates of the adaptive graph structure. Results on two real-world datasets demonstrate that the method of this paper outperforms state-of-the-art baselines.

Keywords next POI recommendation; adaptive graph; extra residual connections

收稿日期 2025-01-13

* **通信作者** 李艳红(1973-),女,教授,博士,研究方向:时空数据处理、网络大数据、DB与AI融合,E-mail: liyanhong@mail.scuec.edu.cn

基金项目 湖北省自然科学基金资助项目(2017CFB135);中央高校基本科研业务费专项资金资助项目(CZY23019);网络创新及应用型人才课程实践教学研究资助项目(2019年第一批)

随着移动设备的普及,用户在 Foursquare、 Gowalla 等基于位置的社交网络(Location-Based Social Networks, LBSNs)上产生了大量的兴趣点(Point-of-Interest, POI)访问信息,如商场、酒吧等.这些平台为了提升用户体验^[1],致力于提供更精准的推荐服务,因此 POI 推荐成为了研究的重点.下个 POI 推荐,作为 POI 推荐的一种特殊形式,主要通过分析用户的历史签到数据和当前位置信息,捕捉用户兴趣的动态变化,从而为用户提供更加个性化的推荐^[2-4].该任务的特点在于,它受到时间和地理因素的双重影响^[5-6].

早期的研究针对个性化推荐方法进行了广泛的探索,最初的模型主要集中于捕捉序列中元素之间的关系变化,通常采用传统方法,如马尔可夫链和矩阵分解等.随着深度学习的兴起,基于循环神经网络(Recurrent Neural Network, RNN)的方法也开始被引入下个 POI 推荐领域.然而,经典的 RNN 结构(包括传统 RNN 和长短期记忆网络 LSTM)尽管在自然语言处理领域表现出色,但是对于 POI 推荐任务而言,它们并不能很好地把握时间和空间依赖关系.为此,研究者们尝试引入时间和空间等辅助信息,或者通过增加辅助任务和调整损失函数来提升推荐效果^[7-9].后来伴随着 Transformer 的横空出世,人们意识到自注意力机制在建模序列关系时的强大能力,使得许多学者转向研究自注意力网络在下个 POI 推荐任务中的应用^[10-12].

最近,基于图的方法受到了广泛关注^[13-14],特别是图神经网络(Graph Neural Network, GNN)在 POI 表示学习中的潜力备受瞩目.近期的研究^[15-16]主要通过用户的签到序列构建图结构,从而进行 POI 表示学习.尽管这种方法能够生成质量较高的 POI 嵌入,但预定义的图结构往往受制于人为因素.这一问题促使了自适应图方法的出现^[17].自适应图方法继承了 GNN 中的邻居聚合和非线性变换思想,将邻接矩阵的构建过程交由 POI 嵌入来完成.这种方法在模型的学习过程中动态地生成图结构,从而提升学习到的 POI 嵌入质量.

尽管自适应图方法在一定程度上取得了良好的效果,但在缺乏明确指导的情况下,生成的 POI 嵌入往往不够理想.此外,作为图神经网络的一种,自适应图方法难以避免 GNN 自身的两个核心问题——过平滑^[18]和过相关^[19],如何有效地缓解这两个问题,仍是当前面临的关键挑战.

另一方面,作为一种序列任务,无论将 POI 嵌入输入 RNN 类模型还是自注意力类模型,都不能忽视

POI 图结构学习与下游模型之间的联系.WANG 等^[17]的研究中,将 POI 嵌入直接用于扩展的自注意力机制.然而,这种简单的组合方式忽略了反向传播中从注意力模块传递至自适应图模块的梯度关系.由于这一缺陷,在多次前向传播和反向传播的过程中,自适应图模块难以学习到更优的图结构,从而无法生成更高质量的 POI 嵌入,最终导致注意力模块的预测效果未达到最佳水平.

为了解决上述问题,本文为下个 POI 推荐任务提出了去相关图表示增强的注意力网络(Decorrelated Graph Representation-enhanced Attention Network, DGRAN).为了在图学习的过程中减少人为因素的影响,本文通过探索图结构学习^[18]提出去相关的自适应图方法,以在下个 POI 推荐任务中建模 POI 之间的复杂的地理依赖关系,从而学习富有表达能力的 POI 嵌入.具体来说,本文为简单 GCN 图卷积网络(Simple Graph Convolutional Network, SGCN)^[19]重新构造了自适应邻接矩阵,然后将其每一层看作单独的图降噪过程,修改其前向传播的方法;接着,本文在做序列建模时,为注意力模块加入了额外的残差连接,使得反向传播时能为自适应图模块提供更大的梯度,进一步提升 POI 嵌入的质量.对两个真实世界 LBSN 数据集的研究表明, DGRAN 优于最先进的基线方法.

本文的主要贡献如下:

(1) 本文提出使用去相关的自适应简单图神经网络来学习 POI;

(2) 本文提出了 DGRAN 模型,在扩展的注意力网络中加入了额外的残差连接,阐释了注意力网络和自适应图网络之间的梯度关系,使得自适应图网络能够更有效地训练 POI 嵌入;

(3) 在两个真实世界 LBSN 数据集上执行了实验,结果表明本文提出的 DGRAN 模型在下个 POI 推荐任务中超出了 SOTA 基线模型的性能.

1 相关工作

本节将简要介绍下个 POI 推荐以及图结构学习这两方面的工作.

1.1 下个 POI 推荐

下个 POI 推荐基于用户的历史签到数据,为用户推荐下一个可能访问的 POI.早期的工作主要是基于矩阵分解和马尔科夫链.CHENG 等^[4]使用了矩

阵分解技术为每个用户学习一个个性化的过渡矩阵,这种方法并不能够有效处理高阶的序列模式,并且在处理未观测到的数据时难以为其分配转变概率.FENG等^[2]使用度量嵌入方法将POI映射到低维欧氏空间,通过欧氏距离来度量POI之间的相似度,这种映射方式并不能很有效地建模POI之间复杂的关系,使用过程也不能充分捕获用户的访问模式以及兴趣的变化.随着技术的更迭,包括RNN以及自注意力网络在内的深度学习模型被认为是更有效的序列建模方法.因此后续的学者们尝试对这两大类基础方法进行修改,以达到更高质量的推荐效果.LIU等^[8]通过计算两个连续访问之间的时间和空间间隔,将其加入RNN以更好地建模连续访问的关系.HUANG等^[11]通过将时间作为相似性判别方法进行显式的注意力计算,以此为注意力权重聚合LSTM的结果.

LIAN等^[12]将地理坐标编码送入自注意力网络以构建位置关系,将该模块整合到Transformer上,以获得签到序列点对点的交互信息.LUO等^[10]显式计算了所有签到之间的时空距离,通过一个双层注意力结构来考虑个性化项的频次.QIN等^[20]使用频域的自相关方法捕获子序列之间的关系.但是这些基于序列的方法都是直接将POI转化为嵌入,忽视了对POI嵌入质量的追求,这就会导致下游任务在这种质量不高的嵌入上不能很好地把握用户偏好的变化.

近期的工作尝试使用图表示学习方法构建更具表达性的POI表示.LI等^[13]通过图注意力网络将POI的嵌入以及类别信息的表示一同加入学习,从而生成具有类别信息的嵌入表示.YANG等^[21]构建了一个全局的轨迹流图,然后以此为基础通过GCN来学习拓扑信息.RAO等^[16]对图的权重做了进一步研究,除节点之间的连接性外,构建了富含多种关系的图结构,然后使用GCN学习其中的转变关系.尽管有了这些使用图神经网络的方法,但是如何构建一个高质量的图仍未被充分研究.YIN等^[22]使用知识图谱的方法在复数域提取POI之间的关系,通过聚类方法来动态构建图结构.WANG等^[17]设计了一种图结构随模型训练而更新的自适应图,通过这种结构来补充未被观测到的访问关系并去除噪声信息,进而学习到更具表达能力的POI嵌入.总的来说,基于预定义图的方法需要很多的领域知识才能达到不错的效果,图结构学习的方法却需要设计复杂的关系映射手段和结构选择方法,导致图的设计

过于冗余.

1.2 图神经网络中的图结构学习

图神经网络(GNN)在许多监督、半监督以及零样本学习任务下大放异彩,GNN的成功主要得益于它的消息传播机制^[23],归根结底是邻接矩阵的设计,因此优秀的图结构的设计有助于图表示的学习.对于推荐任务而言,节点的异质性对于图结构设计而言是一个较大的挑战.目前预定义图的设计方案仍旧占据主导地位^[16,24-26],仅有少数方法^[17,27]采用了图结构学习的思路.本文使用简化的图学习方案,并将图降噪方法融入其中,从而减少对先验知识的依赖.

2 准备工作

本节给出问题定义以及要进行改进的功能模块.

2.1 问题定义

定义 1(签到序列) 令 $U = \{u_1, u_2, \dots, u_m\}$ 表示所有的用户, $L = \{l_1, l_2, \dots, l_e\}$ 表示所有POI,其中每个POI l_i 都与一个独一无二的地理坐标 $(\text{long}_i, \text{lat}_i)$ 相关联.每个用户的签到记录 $r_{u_k - t_q}$ 可以表示为三元组 $(u_k, l_{i_q}^u, t_q)$,表示用户 u_k 在时刻 t_q 访问了POI $l_{i_q}^u$.用户 u_k 的签到序列可以表示为 $R(u_k) = \{r_{u_k - t_1}, r_{u_k - t_2}, \dots, r_{u_k - t_{|C(u_k)|}}\}$.

定义 2(下个POI推荐) 给定一个用户 u_k 以及他从时刻 t_1 到 t_q 的签到序列 $R(u_k)$,下个POI推荐任务的目标是基于他的历史轨迹产生下个时刻 t_{q+1} 时该用户最有可能访问的top- k POI列表.

2.2 简单图卷积网络

本文使用了SGCN^[25].对于一个无向图 $G = (V, E, A)$,其中 $V, E, A \in \mathbb{R}^{N \times N}$ 分别表示顶点集、边集、邻接矩阵,以此为基础的GCN第 l 层向后的传播规则为:

$$H^{(l+1)} = \sigma \left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right), \quad (1)$$

其中 σ 表示激活函数, $\tilde{A} = A + I \in \mathbb{R}^{N \times N}$ 表示加入自连接的邻接矩阵, $\tilde{D}_{i,i} = \sum_j \tilde{A}_{i,j}$, W^l 是可学习的矩阵.

令 $\tilde{A} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$,SGCN将GCN传播规则线性化:

$$Y = \tilde{A} \tilde{A} \dots \tilde{A} H^{(0)} W^{(0)} W^{(1)} \dots W^{(l)} = \tilde{A} H^{(0)} W, \quad (2)$$

与传统GNN不同的是,它直接构造了输出 Y ,而非定义每一层的传播过程,在将SGCN与GCN相对应时,可以将SGCN的传播过程视作:

$$H^{(l+1)} = \tilde{A} H^{(l)}, \quad (3)$$

值得注意的是, $H^{(l+1)}$ 不是最后的输出,当需要用到

SGCN 的最后结果时,需要对 $H^{(l+1)}$ 做一次无偏置项的线性变换.

2.3 时空间隔感知的自注意力机制

WANG 在 LI 等提出的时间感知自注意力^[28]的基础上拓展了空间信息^[17].对于某个用户 u_k ,假定他的签到序列为 $R(u_k) = \{r_{u_k-t_1}, r_{u_k-t_2}, \dots, r_{u_k-t_n}\}$,计算该签到序列中任意两个访问之间的时间、空间间隔矩阵 $M^{t,u_k} \in \mathbb{R}^{n \times n}$ 、 $M^{s,u_k} \in \mathbb{R}^{n \times n}$,分别将这两组矩阵投影为两个嵌入矩阵 $E^{t,u_k}, E^{s,u_k} \in \mathbb{R}^{n \times n \times d}$ 和 $E^{s_k,u_k}, E^{s_r,u_k} \in \mathbb{R}^{n \times n \times d}$.将用户序列中的每个 POI 转化为嵌入表示 $e_i^{u_k} \in \mathbb{R}^d$.用户 u_k 的序列经过注意力模块的聚合结果 $h_i^{u_k}$ 为:

$$h_i^{u_k} = \sum_{j=1}^i \alpha_{ij} (e_j^{u_k} W^V + e_{i,j}^{t,u_k} + e_{i,j}^{s,u_k} + p_j), \quad (4)$$

其中 $W^V \in \mathbb{R}^{d \times d}$ 是可学习的矩阵, $e_{i,j}^{t,u_k} \in \mathbb{R}^d$ 是 E^{t,u_k} 的第 i 行第 j 列元素, $e_{i,j}^{s,u_k} \in \mathbb{R}^d$ 是 E^{s,u_k} 的第 i 行第 j 列元素, p_j 是第 j 个位置的位置嵌入. α_{ij} 表示注意力系数,它的计算方式如下:

$$\alpha_{ij} = \frac{\exp a_{ij}}{\sum_{k=1}^i \exp a_{ik}}, \quad (5)$$

$$a_{ij} = \frac{\text{LN}(e_i^{u_k}) W^Q (e_j^{u_k} W^K + e_{i,j}^{t,u_k} + e_{i,j}^{s,u_k} + p_j)^\top}{\sqrt{d}}, \quad (6)$$

其中 $\text{LN}(\cdot)$ 代表层归一化(Layer Normalization)操作, $W^Q, W^K \in \mathbb{R}^{d \times d}$ 是可学习矩阵, $e_i^{t,u_k} \in \mathbb{R}^d$ 是 E^{t,u_k} 的第 i 行第 j 列元素, $e_i^{s,u_k} \in \mathbb{R}^d$ 是 E^{s,u_k} 的第 i 行第 j 列元素.

3 提出的方法

DGRAN 模型的结构图如图 1 所示,可分为两部分:(1)去相关图学习模块.首先是将简单图设计成自适应的结构,然后再加入去相关方法修改图的传播过程;(2)具有额外残差连接的自注意力聚合模块,这部分在 2.3 节的注意力机制上添加了额外的残差连接.

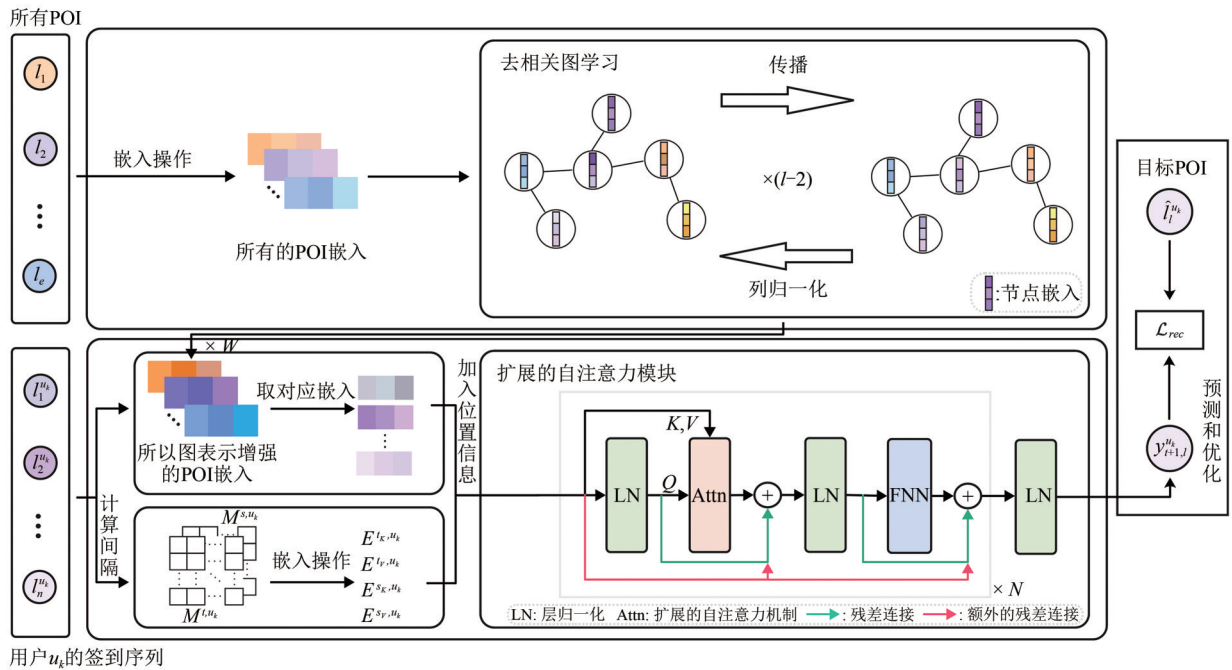


图 1 DGRAN 的整体架构

Fig. 1 The overall architecture of DGRAN

3.1 去相关图学习

作为 POI 推荐领域提出并使用自适应图的先驱, WANG 等^[17]在构造图的时候沿袭了前人的度量方法以及设计模式(包括生成邻接矩阵时的权重矩阵、图的稀疏化以及用平均池化的方法缓解过平滑问题)^[27],并为图表示学习设计了额外的正则化方法.但他们的方法仍旧有些复杂,因此本文使用了一种更简单直接的思路来解决这些问题.

3.1.1 自适应简单图构建

在构造图结构之前,需要先为所有 POI 初始化嵌入 $E \in \mathbb{R}^{n \times d}$ 以便于后续 POI 之间关系的学习.紧接着便是图结构的构建.在构造自适应图时,选择哪种距离度量方法构造邻接矩阵是首要问题.跟随现有的工作^[17,27],本文选择余弦相似度作为度量方法,但他们不同的是,本文舍弃了权重矩阵 W ,直接计算 POI 嵌入之间的距离.邻接矩阵的每一项 $A_{i,j}$

计算方式如下:

$$A_{i,j} = \cos(e_i, e_j), \quad (7)$$

其中 $\cos(\cdot, \cdot)$ 表示计算两个向量之间的余弦相似度. 注意, $A_{i,j} \in [-1, 1]$, 负数作为权重是不合理的. 因此现有工作会选择某个值作为阈值来筛选不合理的数值, 比如 WANG 等^[17]选择了 0, 相当于在每个数值上使用了 Relu 函数, 这会导致一部分邻接矩阵的元素无法反向传播, 限制了 POI 嵌入的梯度大小, 因此本文使用 Sigmoid 函数将其映射到 $[0, 1]$, 严谨地来说是 $[\frac{1}{1+e^{-1}}, \frac{1}{1+e^1}]$, 由于 $[-1, 1]$ 输入的范围避开了 Sigmoid 的饱和区间, 那么当反向传播梯度流经邻接矩阵时, 也不容易出现梯度爆炸或梯度消失问题, 这时候邻接矩阵可以表示为:

$$A = \text{Sigmoid}(\cos(E, E)), \quad (8)$$

其中 $\text{Sigmoid}(\cdot)$ 表示逐元素的 Sigmoid 计算, 本文用 $\cos(A, B)$ 表示先对矩阵 A 和 B 进行行归一化得到 A_m 和 B_m , 然后进行矩阵乘法 $A_m B_m^T$. 在后续的图传播过程中, 本文仍保持 $\tilde{A} = A + I$ 的做法, 其中 I 是单位阵, 以加强顶点自身的信息.

3.1.2 去相关方法

仅通过上述环节构造的图, 效果显然是不够的, 因为它失去了非线性激活, 而且也没有一个可靠的方法指引它的学习过程, 还会面临过平滑问题与过相关问题, 这就会导致它学到低质量的 POI 嵌入. 受到文[19]的启发, 本文将 SGCN 的每层看作图降噪过程, 通过将正交的正则化图信号降噪方法的梯度整合到图的传播过程中来提升学到的 POI 嵌入的质量. 正交的正则化图信号降噪问题如下:

$$\min_E J(E) = \frac{1}{2} \|E - E_{in}\|_F^2 + \frac{\lambda_1}{2} \text{tr}(E^T \tilde{L} E) + \frac{\lambda_2}{4} \|E^T E - I\|_F^2, \quad (9)$$

其中 λ_1, λ_2 是超参数. $\tilde{L} = I - \tilde{A} = I - \tilde{D}^{-\frac{1}{2}}(\text{Sigmoid}(\cos(E, E)) + I)\tilde{D}^{-\frac{1}{2}}$. 使用公式 (9) 对 SGCN 逐层降噪时需要计算它对 E 的梯度, 在对除了第 0 层以外的层求梯度时, 公式如下:

$$\nabla J(E) \Big|_{E=E^{(l)}} = E^{(l)} - E_{in}^{(l)} + \lambda_1 \tilde{L} E^{(l)} + \lambda_2 E^{(l)}(E^{(l)T} E^{(l)} - I), \quad (10)$$

但值得注意的是, 公式 (9) 在对第 0 层的 E 求导时, \tilde{L} 与 $E^{(0)}$ 是有关的, 公式 (9) 中第 2 项对 $E^{(0)}$ 的微分计算如下:

$$\begin{aligned} & d\left(\text{tr}(E^T \tilde{L} E)\right) \Big|_{E=E^{(0)}} = \\ & d\left(\text{tr}(E^T (I - \tilde{A}) E)\right) \Big|_{E=E^{(0)}} = \\ & (\text{dtr}(E^T E) - \\ & \text{dtr}\left(E^T \tilde{D}^{-\frac{1}{2}}(\text{Sigmoid}(\cos(E, E)))\tilde{D}^{-\frac{1}{2}} E\right) - \\ & \text{dtr}(E^T \tilde{D}^{-1} E)) \Big|_{E=E^{(0)}} = \\ & (\text{tr}(2E^T dE) - \\ & 2\text{tr}\left(E^T \tilde{D}^{-\frac{1}{2}} \text{Sigmoid}(\cos(E, E))\tilde{D}^{-\frac{1}{2}} dE\right) - \\ & \text{tr}\left(E^T \tilde{D}^{-\frac{1}{2}} d(\text{Sigmoid}(\cos(E, E)))\tilde{D}^{-\frac{1}{2}} E\right) - \\ & \text{tr}(2E^T \tilde{D}^{-1} dE)) \Big|_{E=E^{(0)}}, \end{aligned} \quad (11)$$

虽然上式并没有完全求解, 但是可以看到其中复杂的矩阵乘法过程, 值得注意的是矩阵 E 和 D 的大小与 POI 的数目相关, 所以在此直接求梯度就会成为一个相当耗费时间和空间的过程, 本文建议直接忽略掉这一步, 直接使用公式 (10) 来代替这一过程.

所以 POI 嵌入 E 在第 0 层的传播规则为:

$$E^{(l)} = E^{(l-1)} - \gamma \nabla J(E) \Big|_{E=E^{(l-1)}} = (1 - \gamma - \gamma \lambda_1 - \gamma \lambda_2) E^{(l-1)} + \gamma E_{in} + \gamma \lambda_1 \tilde{A} E^{(l-1)} - \gamma \lambda_2 E^{(l-1)}(E^{(l-1)T} E^{(l-1)}), \quad (12)$$

将 SGCN 看作逐层的单步图降噪, 令 $E_{in} = E^{(l-1)}$, 修改后的传播规则为:

$$E^{(l)} = (1 - \gamma \lambda_1 + \gamma \lambda_2) E^{(l-1)} + \gamma \lambda_1 \tilde{A} E^{(l-1)} - \gamma \lambda_2 E^{(l-1)}(E^{(l-1)T} E^{(l-1)}), \quad (13)$$

为了避免梯度爆炸, 本文在传播后对嵌入进行列归一化:

$$E^{(l)} = \text{Norm}(E^{(l)}), \quad (14)$$

因此, 当图具有 l_{num} 层时, 去相关自适应简单图的做法是, 先对输入 $E^{(0)}$ 做 K 次带有列归一化的传播, 然后做一次无偏置项的线性变换. 经处理后的 POI 嵌入矩阵记为 $\tilde{E} \in \mathbb{R}^{N \times d}$.

3.2 序列建模

在获得富有表达性的 POI 表示后, 接下来介绍如何使用扩展的自注意力机制对用户的签到序列建模.

3.2.1 签到序列编码

下一个 POI 推荐任务以用户的历史轨迹为依据来推测用户下一个时刻最有可能访问的位置. 首先将用户的访问足迹处理为固定长度 n , 如果长度大于 n 就

截断,只保留最近的 n 条数据;如果长度小于 n ,那就向左填补直到长度为 n . 随后为处理后的访问序列 $R(u_k)$ 得到通过图方法获得的嵌入,将该序列的表示向量记为 $E_c(u_k) = \{\tilde{e}_1^{u_k}, \tilde{e}_2^{u_k}, \dots, \tilde{e}_n^{u_k}\}$,其中 $\tilde{e}_i^{u_k} \in \tilde{E}$.

因为自注意力机制对位置先后顺序不敏感,所以此处使用了 FOUMANI 等^[29]提出的绝对位置编码方式 tAPE 来感知位置信息:

$$p_i(2k) = \sin i\omega_k, p_i(2k+1) = \cos i\omega_k, \omega_k = \frac{10000^{-\frac{2k}{d}} \times d}{n}, \quad (15)$$

其中 k 在区间 $[0, \frac{d}{2}]$ 内, $2k$ 和 $2k+1$ 分别用于表示嵌入维度中偶数和奇数位置, n 是序列长度, d 是嵌入维度. 该方法考虑了输入嵌入维度和长度,使得较低维度的嵌入在计算时也能正确表达距离关系. 加入位置的用户访问序列记为 $E_c(u_k) = \{\hat{e}_1^{u_k}, \hat{e}_2^{u_k}, \dots, \hat{e}_n^{u_k}\}$.

3.2.2 时空依赖编码

沿用 WANG 和 LI 等^[17,28]的方法,每个用户分别构建一个时间矩阵 M^{t, u_k} 和空间矩阵 M^{s, u_k} ,用于记录任意两个签到之间的时间间隔和空间间隔. 对于两个矩阵中的元素,计算方式如下:

$$t_{i,j}^{u_k} = \min \left(t_{\text{thr}}, \left\lfloor \frac{|t_i^{u_k} - t_j^{u_k}|}{t_{\text{min}}^{u_k}} \right\rfloor \right), \quad (16)$$

$$s_{i,j}^{u_k} = \min \left(s_{\text{thr}}, \left\lfloor \text{haversine}(l_i^{u_k}, l_j^{u_k}) \right\rfloor \right), \quad (17)$$

其中 $t_{\text{min}}^{u_k} = \min(|t_i^{u_k} - t_j^{u_k}|)$,表示用户 u_k 时间间隔的最小值. $\text{haversine}(\cdot, \cdot)$ 表示 Haversine 公式,用于计算两者之间坐标的距离. $t_{\text{thr}}, s_{\text{thr}}$ 分别代表时间间隔和空间距离的阈值,是超参数. 在得到这些时间和空间关系之后,通过嵌入操作来分别获得两组嵌入矩阵 $E^{t, u_k}, E^{s, u_k} \in \mathbb{R}^{n \times n \times d}$ 和 $E^{s, u_k}, E^{t, u_k} \in \mathbb{R}^{n \times n \times d}$.

3.2.3 扩展的自注意力模块

接下来展示如何使用处理好的序列信息以及时空上下文信息,即图 1 中的扩展的注意力模块.

WANG 等^[16]虽然在时空信息方面扩展了注意力机制,但是没有充分考虑自适应图与注意力模块的梯度关系. 他们沿用了 LI 等^[28]使用的自注意力模型,这是一种类似于 Pre-LN 型^[30]Transformer 的 Encoder 的实现形式,其中 Attention 表示时空间隔感知自注意力机制,图中简化了它原有的结构(具体的信息聚合过程没有画出,dropout 层也被省去),FNN 表示前馈神经网络. 接下来的表述中,用“Attn”代替“Attention”,“LN”代替“Layer Norm”. 假设输入

为 $x \in \mathbb{R}^d$,该模块的输出 Attn_{out} 可以表示为:

$$\begin{aligned} \text{Attn}_{\text{out}} = & \\ & \text{FNN} \left(\text{LN} \left(\text{Attn} \left(\text{LN}(x), x, x \right) + \text{LN}(x) \right) \right) + \quad (18) \\ & \text{LN} \left(\text{Attn} \left(\text{LN}(x), x, x \right) + \text{LN}(x) \right), \end{aligned}$$

上式对 x 的梯度为:

$$\begin{aligned} \frac{\partial \text{Attn}_{\text{out}}}{\partial x} = & \\ & \frac{\partial \text{FNN} \left(\text{LN} \left(\text{Attn} \left(\text{LN}(x), x, x \right) + \text{LN}(x) \right) \right)}{\partial x} + \quad (19) \\ & \frac{\partial \text{LN} \left(\text{Attn} \left(\text{LN}(x), x, x \right) + \text{LN}(x) \right)}{\partial x}, \end{aligned}$$

值得注意的是,该结构内部较短的残差机制并不能给输入较大的梯度,受 TAKASE 等^[30]的启发,本文在该模块中加入额外的残差机制,修改后自注意力模块的输出为:

$$\begin{aligned} \text{Attn}_{\text{out}} = & \\ & \text{FNN} \left(\text{LN} \left(\text{Attn} \left(\text{LN}(x), x, x \right) + \text{LN}(x+x) \right) \right) + \quad (20) \\ & \text{LN} \left(\text{Attn} \left(\text{LN}(x), x, x \right) + \text{LN}(x+x) \right) + x, \end{aligned}$$

此时对 x 的梯度为:

$$\begin{aligned} \frac{\partial \text{Attn}_{\text{out}}}{\partial x} = & \\ & \frac{\partial \text{FNN} \left(\text{LN} \left(\text{Attn} \left(\text{LN}(x), x, x \right) + \text{LN}(x+x) \right) \right)}{\partial x} + \quad (21) \\ & \frac{\partial \text{LN} \left(\text{Attn} \left(\text{LN}(x), x, x \right) + \text{LN}(x+x) \right)}{\partial x} + I, \end{aligned}$$

在反向传播时,更大的梯度值经过自注意力层传递到自适应图部分,能为自适应图习得更有效的 POI 嵌入,使得在下一轮前向传播时构建更优质的图结构.

此时注意力模块的处理过程为,将输入的用户签到序列 $E_c(u_k)$ 送入堆叠了 N 块的自注意力块:

$$\begin{aligned} x_i^{u_k} = & \\ & \text{FNN} \left(\text{LN} \left(\text{Attn} \left(\text{LN}(\hat{e}_i^{u_k}), \hat{e}_i^{u_k}, \hat{e}_i^{u_k} \right) + \text{LN}(\hat{e}_i^{u_k}) + \hat{e}_i^{u_k} \right) \right) + \quad (22) \\ & \text{LN} \left(\text{Attn} \left(\text{LN}(\hat{e}_i^{u_k}), \hat{e}_i^{u_k}, \hat{e}_i^{u_k} \right) + \text{LN}(\hat{e}_i^{u_k}) + \hat{e}_i^{u_k} \right) + \hat{e}_i^{u_k}, \\ & \hat{e}_i^{u_k} = x_i^{u_k}, \end{aligned}$$

其中 $i \in [1, n]$. 将第 N 块的输出结果进行层归一化,得到扩展的自注意力模块的结果:

$$h_i^{u_k} = \text{LN}(x_i^{u_k}).$$

3.3 预测和优化

3.3.1 模型预测

经过扩展的自注意力模块处理后得到的输出

为 $H(u_k) = \{h_1^{u_k}, h_2^{u_k}, \dots, h_n^{u_k}\}$, 该序列充分捕获了用户的长期依赖和短期兴趣转变的倾向, 因此当前用户在时刻 $t + 1$ 时对每个 POI l_i 的兴趣分数可以通过如下方式计算:

$$\hat{y}_{t+1,l}^{u_k} = \frac{\exp(h_t^{u_k} \tilde{e}_l^\top)}{\sum_{j=1}^{|I|} \exp(h_t^{u_k} \tilde{e}_j^\top)}, \quad (23)$$

其中 \tilde{e}_i 表示 \tilde{E} 的第 i 行, 是通过去相关图学习方法得到的 POI 嵌入, $h_t^{u_k}$ 表示用户 u_k 在时刻 t 的兴趣状态. 在推荐评估阶段, 根据兴趣分数为当前用户排列出他可能访问的 POI, 得到一个 top- k 列表.

3.3.2 模型优化

本文采用交叉熵损失函数来度量预测 POI 与目标 POI 之间的距离:

$$\mathcal{L}_{\text{rec}} = - \sum_{u_k \in U} \sum_{i \in C(u_k)} \sum_{j=1}^n y_{i,j}^{u_k} \log(\hat{y}_{i,j}^{u_k}), \quad (24)$$

其中 $y_{i,j}^{u_k}$ 是真正例标签, 当用户在时刻 i 访问位置 l_j 时, 其值为 1, 否则为 0. 因此, 在加入多头注意力的正则化项后, 最终的损失函数为:

$$L = \mathcal{L}_{\text{rec}} + \tau \|\theta\|_2, \quad (25)$$

其中 τ 是超参数, $\|\theta\|_2$ 是模型所有参数的 L2 正则项.

4 实验

4.1 实验设置

4.1.1 数据集

本文在两个分别从 Foursquare (<https://sites.google.com/site/yangdingqi/home/foursquare-dataset>) 和 Gowalla (<https://snap.stanford.edu/data/loc-Gowalla.html>) 上收集到的真实数据集上执行实验. 其中 Foursquare 数据集包括了从 2012 年 4 月到 2013 年 9 月在新加坡的签到数据, Gowalla 数据集包括了从 2009 年 3 月到 2010 年 10 月在美国的加利福尼亚州和内华达州的签到数据. 沿用现有工作的设定, 实验中将被访问次数小于 5 的 POI 以及签到数目小于 5 的用户视作不流行的 POI 和不活跃的用户, 在数据预处理时将其剔除. 处理后的两个数据集的统计信息如表 1 所示. 在数据集划分时, 本文采用了留一法策略.

表 1 数据集统计信息

Tab. 1 Statistic of datasets

数据集	用户数量	POI 数量	平均签到序列长度
Foursquare	4638	9731	64.44
Gowalla	7293	18079	44.34

4.1.2 基线算法

本文将 DGRAN 与传统方法 (FPMC)、基于 RNN 的方法 (Flashback)、基于注意力的方法 (TiSASRec 和 GeoSAN)、基于 GNN 的方法 (Graph-Flashback 和 AGRAN) 进行比较.

FPMC^[31]: 使用矩阵分解和马尔可夫链来建模 POI 之间转变关系.

Flashback^[9]: 使用时间和空间间距来增强上下文感知能力的基于 RNN 的模型.

TiSASRec^[28]: 基于自注意力机制的模型, 通过使用签到序列中 POI 间的时间间隔信息, 使得注意力机制能更好地感知时间关系.

GeoSAN^[12]: 基于自注意力机制的模型, 通过网格映射来编码地理位置关系, 在序列处理时额外建模位置关系.

Graph-Flashback^[16]: 为学习 POI 表示构建了一个时空知识图谱. 为了公平性, 本文去除了它的社会关系.

AGRAN^[17]: 使用自适应图学习 POI 表示, 并为自注意力机制扩展了时间和空间信息.

4.1.3 评估指标

本文使用了推荐系统里的两个标准评估方法, 截断到前 k 的召回率 Recall (Recall@ k) 以及截断到前 k 的归一化折损累计增益 NDCG (NDCG@ k):

$$\text{Recall}@k = \frac{1}{|U|} \sum_{u_k \in U} \frac{|L_{u_k}^T \cap L_{u_k}^R|}{|L_{u_k}^T|}, \quad (26)$$

其中 $L_{u_k}^T$ 是真正例列表, $L_{u_k}^R$ 是用户 u_k 的 top- k 推荐列表.

$$\text{NDCG}@k = \begin{cases} \frac{1}{\log_2(\text{rank}_l + 1)}, & \text{rank}_l \leq k \\ 0, & \text{rank}_l > k \end{cases}, \quad (27)$$

其中 rank_l 表示目标 POI l_i 在排名列表里的位置. k 的取值范围为 $\{2, 5, 10\}$.

4.1.4 实现细节

所有实验在一个 NVIDIA RTX 4060ti 16 GB GPU 上执行. 为了公平起见, 本文将所有模型的 POI 嵌入维度设置为 64, 其他参数使用模型表现最好的参数. 对于 DGRAN, Foursquare 的 L2 正则项系数 τ 设置为 0.0005, Gowalla 的 L2 正则项系数 τ 设置为 0.0001; Foursquare 的图卷积的层数设置为 5, Gowalla 的图卷积的层数设置为 4; Foursquare 的 γ 设置为 0.01, Gowalla 的 γ 设置为 0.005; Foursquare 的 λ_1 设置为 140, Gowalla 的 λ_1 设置为 220, λ_2 设置为 0.001, 学习率设置为 0.001, 丢失率设置为 0.4, 自注意力模块

设置为 2 层,多头机制设置为 2 头,空间距离和时间距离的上界 s_{thr}, t_{thr} 都设置为 256.

4.2 对比实验

表 2 和表 3 展示了所有模型在 Foursquare 和 Gowalla 上的推荐性能,可以得出如下结果:

(1) DGRAN 超越了所有类型的基线. 本文将 DGRAN 的性能提升归因于两方面: 1) 去相关的自适应图能够学到更优质的图结构,进而为 POI 学到更丰富的表示,差异化 POI 表示的各个维度能让该向量表示更具表达能力,从而使得后续的自注意力网络能更好把握住 POI 之间的潜在关系以及用户的兴趣变化趋势; 2) 额外的残差机制能够给予图方法更大的梯度,在帮助 POI 表示学习的同时,产生更优质的图结构;

(2) 传统方法 FPMC 由于马尔科夫链自身的弊病,无法构建更高阶的关系模式,从而无法准确建模用户的潜在偏好,导致其预测出效果不佳;

(3) 基于 RNN 类的方法 Flashback 虽然在建模过程中额外建模了时空信息对当前时刻的周期性影响,但是由于 RNN 类方法更适合建立短期序列关

系并具有一定的遗忘性且无法像自注意力机制一样直接建模点对点的关系,导致其无法充分把握当前时刻之前的所有时刻对当前的重要性,导致其最终预测效果不足;

(4) 基于注意力的模型源自于其点对点的注意力计算方式,使得其最终结果优于传统方法和基于 RNN 的模型. 虽然 GeoSAN 和 TiSASRec 建模时间的方式略有不同,但 GeoSAN 还从 POI 的位置信息中获益,从而达到了更好的预测效果. 不过这些方法忽略了输入嵌入自身的质量,这导致仅靠自注意力机制也无法精确获取用户兴趣的潜在偏好和转移关系;

(5) 使用图神经网络的模型通过建模 POI 关系图,为 POI 学习到富含表达能力的向量表示,去除了 POI 之间的噪声信息,这使得它们能胜过其他模型. 其中 AGRAN 更是因为使用了自适应图方法,使其在建模 POI 图时更胜一筹,因此该模型的 POI 表示质量更高,不过由于它忽视了自适应图与自注意力网络的梯度关系,导致反向传播时不能为图网络提供足够的梯度,影响了图结构的更新,从而导致模型的性能不足.

表 2 不同推荐算法在 Foursquare 数据集上的性能比较

Tab. 2 Performance comparison of different recommendation algorithms on the Foursquare dataset

模型	Recall@2	Recall@5	Recall@10	NDCG@2	NDCG@5	NDCG@10
FPMC	0.1233	0.1509	0.1826	0.1178	0.1301	0.1402
Flashback	0.1302	0.1802	0.2535	0.1102	0.1305	0.1501
TiSASRec	0.1257	0.1937	0.2631	0.1064	0.1458	0.1682
GeoSAN	0.1361	0.2025	0.2687	0.1188	0.1485	0.1698
Graph-Flashback	0.1455	0.2239	0.2884	0.1283	0.1709	0.1863
AGRAN	0.1630	0.2473	0.3235	0.1428	0.1815	0.2048
DGRAN	0.1751	0.2639	0.3335	0.1498	0.1896	0.2121
Improvement	7.42%	6.71%	3.09%	4.90%	4.30%	3.56%

表 3 不同推荐算法在 Gowalla 数据集上的性能比较

Tab. 3 Performance comparison of different recommendation algorithms on Gowalla dataset

模型	Recall@2	Recall@5	Recall@10	NDCG@2	NDCG@5	NDCG@10
FPMC	0.1230	0.1481	0.1696	0.1213	0.1295	0.1363
Flashback	0.1251	0.1821	0.2286	0.1121	0.1378	0.1528
TiSASRec	0.1352	0.1936	0.2588	0.1260	0.1459	0.1640
GeoSAN	0.1391	0.2095	0.2642	0.1199	0.1514	0.1691
Graph-Flashback	0.1503	0.2209	0.2915	0.1307	0.1725	0.1880
AGRAN	0.1634	0.2487	0.3114	0.1439	0.1820	0.2023
DGRAN	0.1846	0.2679	0.3392	0.1638	0.2013	0.2244
Improvement	12.97%	7.72%	8.93%	13.83%	10.60%	10.92%

4.3 消融实验

为了检验 DGRAN 中主要设计的影响,本文执行了消融实验,移除了不同模块的 3 个变种如下:

● w/o decor: 这个变种移除了去相关机制,即

使用简单图的传播方法.

● w/o extra res: 这个变种移除了额外的残差连接.

● w/o decor & extra res: 这个变种移除了局部

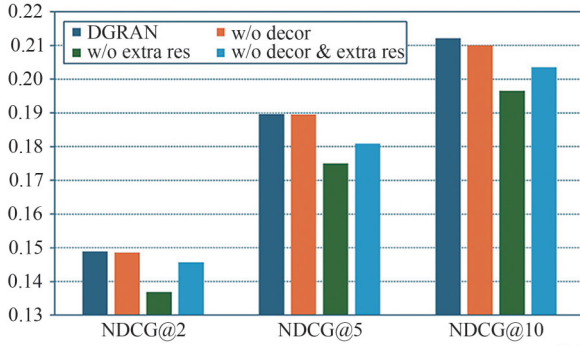
注意力和额外的残差连接.

图 2 展示了两个数据集上的消融实验结果, 对其分析可得:

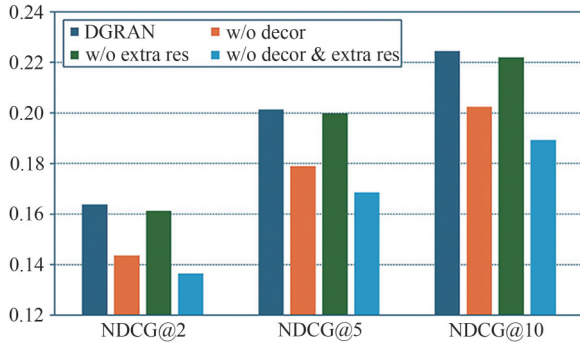
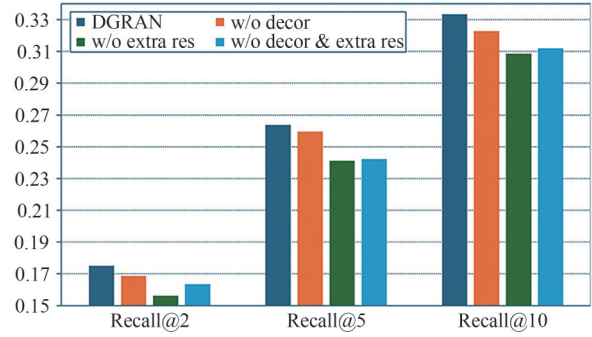
(1) 去相关方法通过减小表示不同维度间的相似度, 有助于生成更有价值的自适应图结构, 使得图方法能够学习到更具表达力的 POI 表示, 进而影响模型

的预测能力, 这是一种需要更少先验知识的有效方法;

(2) 额外的残差机制的有无影响了模型的表现, 即从注意力模块流向自适应图的梯度显著影响着自适应图的学习能力, 即使有着其他各种方法的加持, 在没有足够梯度的情况下, POI 的质量也难以得到保证.



(a) Foursquare 数据集上的消融实验结果



(b) Gowalla 数据集上的消融实验结果

图 2 两个数据集上的消融实验结果

Fig. 2 Ablation results on two datasets

4.4 超参数研究

为探究去 DGRAN 的图层数 l_{num} 对模型效果的影响, 本文在区间 $\{2, 3, 4, 5\}$ 测试模型的性能. 从图 3 可以看出在 Gowalla 数据集上层数为 4 时效果达到最

优, 而在 Foursquare 数据集上层数为 4 和 5 时效果相似, 这说明在更稀疏的 Gowalla 数据集上, 过度采样更远邻居的信息不利于自身信息的表达.

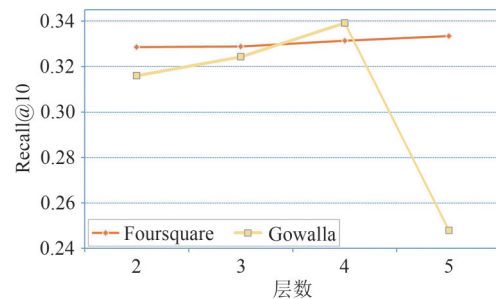
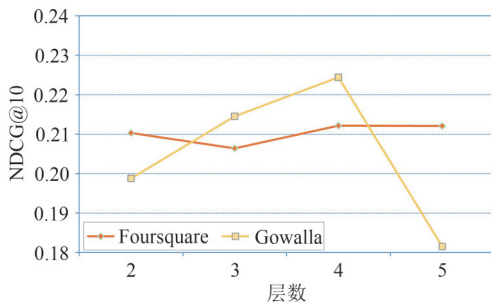


图 3 两个数据集上不同图层数的实验结果

Fig. 3 Experimental results of different graph layers on two datasets

5 结论

本文提出了为下个 POI 推荐而建的 DGRAN 模

型, 进行了对自适应图学习方法的探索, 以及对自注意力机制在应用中的改进; 同时, 本文还在理论与实践上说明了额外残差连接对自适应图与自注

意力机制相结合时的重要性.两个真实数据集上的结果证明了DGRAN性能的优越性.接下来的工作将继续探索自适应图方法的优化.

参 考 文 献

- [1] YU Z, LIAN J, MAHMOODY A, et al. Adaptive user modeling with long and short-term preferences for personalized recommendation[C]//IJCAI. International Joint Conferences on Artificial Intelligence. Macao: Morgan Kaufmann, 2019: 4213-4219.
- [2] FENG S, LI X, ZENG Y, et al. Personalized ranking metric embedding for next new POI recommendation[C]//IJCAI. International Joint Conferences on Artificial Intelligence. Buenos Aires: Morgan Kaufmann, 2015: 2069-2075.
- [3] HAN P, LI Z, LIU Y, et al. Contextualized point-of-interest recommendation[C]//IJCAI. International Joint Conferences on Artificial Intelligence. Yokohama: Morgan Kaufmann, 2020: 2484-2490.
- [4] CHENG C, YANG H, LYU M R, et al. Where you like to go next: Successive point-of-interest recommendation[C]//IJCAI. International Joint Conference on Artificial Intelligence. Beijing: Morgan Kaufmann, 2013: 2605-2611.
- [5] QIAN T, LIU B, NGUYEN Q V H, et al. Spatiotemporal representation learning for translation-based POI recommendation[J]. ACM Transactions on Information Systems(TOIS), 2019, 37(2): 1-24.
- [6] ZHAO S, ZHAO T, YANG H, et al. STELLAR: Spatial-temporal latent ranking for successive point-of-interest recommendation[C]//AAAI. Proceedings of the AAAI Conference on Artificial Intelligence. Phoenix: AAAI, 2016: 315-322.
- [7] KONG D, WU F. HST-LSTM: A hierarchical spatial-temporal long-short term memory network for location prediction[C]//IJCAI. International Joint Conferences on Artificial Intelligence. Stockholm: Morgan Kaufmann, 2018: 2341-2347.
- [8] LIU Q, WU S, WANG L, et al. Predicting the next location: a recurrent model with spatial and temporal contexts[C]//AAAI. Proceedings of the AAAI conference on Artificial Intelligence. Phoenix: AAAI, 2016: 194-200.
- [9] YANG D, FANKHAUSER B, ROSSO P, et al. Location prediction over sparse user mobility traces using RNNs: Flashback in hidden states! [C]//IJCAI. International Joint Conference on Artificial Intelligence. Yokohama: Morgan Kaufmann, 2020: 2184-2190.
- [10] LUO Y, LIU Q, LIU Z. Stan: Spatio-temporal attention network for next location recommendation [C]//IC3C2. The Web Conference. Ljubljana: ACM, 2021: 2177-2185.
- [11] HUANG L, MA Y, WANG S, et al. An attention-based spatiotemporal LSTM network for next POI recommendation [J]. IEEE Transactions on Services Computing, 2019, 14(6): 1585-1597.
- [12] LIAN D, WU Y, GE Y, et al. Geography-aware sequential location recommendation [C]//ACM. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Diego: ACM, 2020: 2009-2019.
- [13] LI Y, CHEN T, LUO Y, et al. Discovering collaborative signals for next POI recommendation with iterative Seq2Graph augmentation [C]//IJCAI. International Joint Conference on Artificial Intelligence. Montreal: Morgan Kaufmann, 2021: 1491-1497.
- [14] 姚静怡,李艳红,黄银峰,等.灵活的属性社区搜索方法[J].中南民族大学学报(自然科学版),2024,43(3): 358-369.
- [15] WANG Z, ZHU Y, ZHANG Q, et al. Graph-enhanced spatial-temporal network for next POI recommendation[J]. ACM Transactions on Knowledge Discovery from Data, 2022, 16(6): 1-21.
- [16] RAO X, CHEN L, LIU Y, et al. Graph-flashback network for next location recommendation [C]//ACM. ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Washington D C: ACM, 2022: 1463-1471.
- [17] WANG Z, ZHU Y, WANG C, et al. Adaptive graph representation learning for next POI recommendation [C]//ACM. International ACM SIGIR Conference on Research and Development in Information Retrieval. Taipei: ACM, 2023: 393-402.
- [18] WU Z, PAN S, CHEN F, et al. A comprehensive survey on graph neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 32(1): 4-24.
- [19] LIU H, HAN H, JIN W, et al. Enhancing graph representations learning with decorrelated propagation[C]//ACM. ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Long Beach: ACM, 2023: 1466-1476.
- [20] QIN Y, FANG Y, LUO H, et al. Next point-of-interest recommendation with auto-correlation enhanced multi-modal transformer network [C]//ACM. International ACM SIGIR Conference on Research and Development in Information Retrieval. Madrid: ACM, 2022: 2612-2616.
- [21] YANG S, LIU J, ZHAO K. GETNext: Trajectory flow map enhanced transformer for next POI recommendation[C]//ACM. ACM SIGIR Conference on research and development in information retrieval. Madrid: ACM, 2022: 1144-1153.
- [22] YIN F, LIU Y, SHEN Z, et al. Next POI recommendation with dynamic graph and explicit dependency[C]//AAAI.

- Proceedings of the AAAI Conference on Artificial Intelligence. Washington D C: AAAI, 2023: 4827-4834.
- [23] WU F, SOUZA A, ZHANG T, et al. Simplifying graph convolutional networks [C]//IMLS. International Conference on Machine Learning. Long Beach: PMLR, 2019: 6861-6871.
- [24] WANG Z, ZHU Y, LIU H, et al. Learning graph-based disentangled representations for next POI recommendation [C]//ACM. International ACM SIGIR Conference on Research and Development in Information Retrieval. Madrid: ACM, 2022: 1154-1163.
- [25] WANG J, YANG B, LIU H, et al. Global spatio-temporal aware graph neural network for next point-of-interest recommendation[J]. Applied Intelligence, 2023, 53(13): 16762-16775.
- [26] TAO H, ZENG J, WANG Z, et al. Next POI recommendation based on spatial and temporal disentanglement representation [C]//IEEE. IEEE International Conference on Web Services. Chicago: IEEE, 2023: 84-90.
- [27] CHANG J, GAO C, ZHENG Y, et al. Sequential recommendation with graph neural networks [C]//ACM. International ACM SIGIR Conference on Research and Development in Information Retrieval. Montreal: ACM, 2021: 378-387.
- [28] LI J, WANG Y, MCAULEY J. Time interval aware self-attention for sequential recommendation [C]//ACM. ACM International Conference on Web Search and Data Mining. Houston: ACM, 2020: 322-330.
- [29] FOUMANI N M, TAN C W, WEBB G I, et al. Improving position encoding of transformers for multivariate time series classification [J]. Data Mining and Knowledge Discovery, 2024, 38(1): 22-48.
- [30] TAKASE S, KIYONO S, KOBAYASHI S, et al. B2t connection: Serving stability and performance in deep transformers [C]//ACL. Annual Meeting of the Association for Computational Linguistics. Toronto: ACL, 2023: 3078-3095.
- [31] RENDLE S, FREUDENTHALER C, SCHMIDT-THIEME L. Factorizing personalized Markov chains for next-basket recommendation [C]//IW3C2. International Conference on World Wide Web. Raleigh: ACM, 2010: 811-820.

(责编 曹东,校对 雷建云)