

文章编号: 2095-2163(2024)03-0212-06

中图分类号: TP391.1

文献标志码: A

基于极限学习机的兴趣点推荐模型

龚亚奇¹, 任建宇¹, 张 祯²

(1 沈阳化工大学 计算机科学与技术学院, 沈阳 110142; 2 营口理工学院 电气工程学院, 辽宁 营口 115014)

摘要: 随着移动社交平台的发展, 基于位置的社交网络服务(Location-Based Social Network, LBSN)已进入人们的视野。在LBSN中, 根据用户的签到数据进行兴趣点(Point-of-Interest, POI)推荐是近年来研究的热点问题。提出一种基于极限学习机(Extreme Learning Machine, ELM)的POI推荐算法, 提取用户的个人偏好、朋友偏好、类型偏好、流行度偏好等特征, 利用ELM提供的分类方法, 使用上述特征向量集合训练ELM分类器, 最终根据分类结果向用户推荐POI。本文使用Foursquare和Twitter数据集的实验结果表明, 该方法在精确率和效率方面均有所提高。

关键词: 兴趣点推荐; 特征提取; 极限学习机; 位置社交网络

Point-of-Interest recommendation model with Extreme Learning Machine

GONG Yaqi¹, REN Jianyu¹, ZHANG Zhen²

(1 School of Computer Science and Technology, Shenyang University of Chemical Technology, Shenyang 110142, China;

2 School of Electrical Engineering, Yingkou Institute of Technology, Yingkou 115014, Liaoning, China)

Abstract: With the development of mobile social platforms, Location-Based Social Network services have come into people's attention. Point-of-Interest recommendation based on users' check-in data in LBSN has been a hot topic in recent years. This paper proposes a POI recommendation algorithm based on Extreme Learning Machine, which extracts the user's personal preference, friend preference, type preference and popularity preference features. Using the classification method provided by ELM, the above feature vector set is used to train the ELM classifier, and the POIs are recommended to the user according to the classification results. In this paper, Foursquare and Twitter datasets are used to prove that the accuracy and efficiency of the proposed method are improved.

Key words: POI recommendation; feature extraction; Extreme Learning Machine; Location-Based Social Networks

0 引言

近些年, 移动通讯设备和个人生活逐渐紧密结合在一起, 人们在现实世界的位置上进行的签到活动变得流行。目前, 国内外出现了许多位置社交网络服务的终端应用, 如国内的美团、大众点评、微信, 国外的Foursquare、 Gowalla、Twitter等。这些位置服务平台不仅能让用户在某个位置和时间点进行签到活动, 还能通过社交平台与他人建立社交关系, 将用户的个人签到信息发布到平台上^[1-3]。用户使用这些社交平台产生签到数据, 这些数据也为分析用户的兴趣和行为习惯带来了新的机遇。但是, 用户往往很难从成千上万的数据中获得对自己有用的信息, 因此在LBSN中挖掘用户未来的兴趣偏好是近年来的重要研究课题^[4]。

用户在社交平台签到, 产生了用户签到数据集, 用户签到数据集中包含了丰富的用户属性与POI属性, 如何更精准地通过用户地理位置挖掘用户的偏好, 向用户提供更加适合的POI是亟待解决的关键问题。LBSN中的数据呈现出规模化、类型多样化等特点, 所需分类算法处理的数据量增多, 如何选择更高效、快速的分类算法是迫切需要解决的热点问题。

针对上述问题, 本文提取签到数据集中的个人偏好、朋友偏好、类型偏好、流行度偏好等特征, 组成POI偏好特征向量, 进一步对用户的POI偏好进行分析。针对复杂多样的大规模数据, 利用ELM分类器的快速学习速度来训练样本, 最终实现精准高效的POI推荐。

作者简介: 龚亚奇(1998-), 男, 硕士研究生, 主要研究方向: 数据挖掘; 任建宇(1996-), 男, 硕士研究生, 主要研究方向: 数据挖掘。

通讯作者: 张 祯(1991-), 女, 博士, 副教授, 主要研究方向: 基于位置的社交网络分析与挖掘。Email: zhangzhen@yku.edu.cn

收稿日期: 2023-03-06

1 相关工作

POI 推荐所采用的主要研究方法分为 5 个类别, 本文将对协同过滤、矩阵分解、链接分析、图嵌入、机器学习的相关研究方法进行分析。

(1) 协同过滤。协同过滤^[5-6]主要利用用户与 POI 之间的签到记录, 寻找用户未去过的 POI 与用户之前去过的 POI 的相似情况。该方法主要利用基于内存的推荐方法^[7-8], 一方面认为有类型喜好的朋友之间将会比较喜欢去相同的地方, 另一方面认为用户倾向于喜欢访问与之前去过的 POI 比较类似的新 POI。面对大规模的签到数据, 不可避免地会存在许多噪声数据, 但该方法不能消除噪声数据, 影响 POI 的推荐效果。

(2) 矩阵分解。矩阵分解经常用来解决数据稀疏性问题, 能够将用户的签到矩阵从高维空间映射到低维空间, 提取用户与 POI 存在的特征关系, 因此有研究人员将矩阵分解和其改进方法运用到推荐领域。Cao 等学者^[9]运用 Rank-FBPR 模型分析用户在服务器上的 POI 偏好, 并和存储在服务器中的用户签到地理位置信息进行交互, 以聚集 POIs, 再将用户签到产生的具体位置信息输入到模型中, 输出用户偏好的 POI, 该算法有效集成了用户在 LBSN 中签到的序列时间信息和实体位置信息。

(3) 链接分析。链接分析的推荐方法, 是通过在网络结构中节点之间进行链接形成节点关系对, 计算 2 个节点之间的关联程度。其中, 代表性的方法是 PageRank^[10]和 HITS^[11-12]。虽然该方法在推荐领域有很多应用, 但却不能将用户的个人偏好加入到推荐方法中, 因此无法针对不同的用户群体来制定 POI 推荐。

(4) 图嵌入。随着 LBSN 中用户在终端平台进行签到活动的数据规模进一步扩大, 传统的序列嵌入方法已不能满足海量的数据分析, 而图嵌入的方法运用到 POI 推荐中成为研究的热点, 该方法能够更加有效地得到用户与 POI 之间的潜在关系。如: Chen 等学者^[13]提出了一种多任务图嵌入的 POI 推荐模型, 对用户终端服务平台的历史记录信息进行建模。使用图嵌入的方式在用户对、用户和 POI 对以及用户和时间对关系图中, 得到影响用户偏好的因素。该方法通过加入多任务关系图嵌入的方法寻找各种影响因素, 有效地提高了推荐的性能。

(5) 机器学习。由于传统的推荐算法在构建较为复杂的推荐模型时训练效果不佳, 面对大规模数

据训练时间过长, 并且利用矩阵分解算法容易导致过度拟合。因此, 在近近年来已将机器学习^[14]应用于推荐问题, 得到了进一步的发展, 且提出了许多好的方法。如: Zhao 等学者^[15]提出了一种名为 GR-DELIM 的分布式极限学习机推荐系统, 系统同时考虑了大规模数据集中的朋友推荐和 POI 推荐, 该方法进一步提高了运用大规模数据进行 POI 推荐和朋友推荐的准确率和效率。Wu 等学者^[16]提出了用户长期和短期兴趣点偏好模型, 通过注意力机制和 LSTM 模型分别获得用户的长期和短期 POI 偏好。Wang 等学者^[17]利用自适应 POI 图, 构建 POI 之间的复杂地理依赖关系, 首次利用图结构为用户推荐下一个 POI。Yan 等学者^[18]利用时空超图卷积网络推荐 POI, 利用超图来获得当前用户和其他用户的历史轨迹信息, 优点是能够学习全局用户偏好并缓解冷启动问题。

综上所述, 经过对现阶段比较热门的推荐方法进行总结, 为进一步对 POI 推荐开展相关研究提供借鉴与参考。

2 ELM 模型

ELM 是由 Huang 等学者^[19]提出的一种单隐层前向神经网络 (Single-Hidden layer Feedforward Network, SLFN) 的训练算法, ELM 预测模型结构如图 1 所示。在 ELM 中, 隐藏节点参数是随机选择, 通常比传统的机器学习方法需要更少的训练时间, 并且倾向于达到最小的训练误差, 还有着比其他传统学习方法更快的学习速度^[20], 其分类性能也优于基于梯度的学习方法。

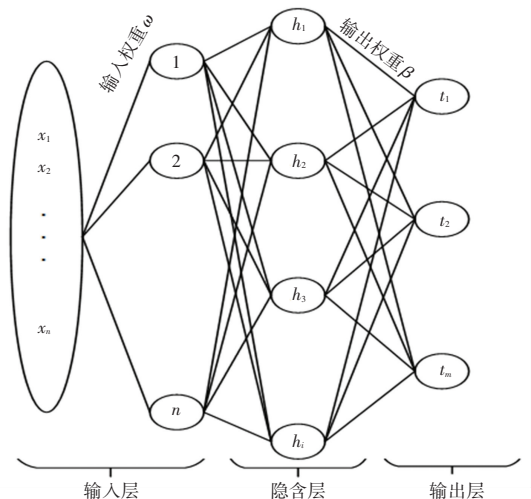


图 1 极限学习机网络结构

Fig. 1 Network structure of Extreme Learning Machine

假设:利用极限学习机对训练集进行训练,训练样本容量为 N , 每个样本表示为 (x_j, t_j) , 其中 $x_j = [x_{j1}, x_{j2}, \dots, x_{jn}] \in R^n, t_j = [t_{j1}, t_{j2}, \dots, t_{jm}] \in R^m$, 激活函数为 $g(x)$, SLFNs 输出函数如式 (1) 所示:

$$\sum_{i=1}^S \beta_i g_i(x_j) = \sum_{i=1}^S \beta_i g(w_i \cdot x_j + b_i) = o_j, \quad j = 1, \dots, N \quad (1)$$

其中, S 是隐含层节点数; $w_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$ 是第 i 个隐含节点的输入权重向量; $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ 是第 i 个隐含节点的输出权重向量; b_i 是第 i 个隐含节点的偏置; o_j 是输出的预测结构。

样本的最小输出误差表示为:

$$\sum_{j=1}^S \|t_j - o_j\| = 0 \quad (2)$$

存在 β_i, w_i, b_i 满足式 (3):

$$\sum_{i=1}^S \beta_i g(w_i \cdot x_j + b_i) = t_j, \quad j = 1, \dots, N \quad (3)$$

上述公式可以改写为 $H\beta = T$ 。

其中:

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} g(w_1 x_1 + b_1) \cdots g(w_s x_1 + b_s) \\ \vdots \\ g(w_1 x_N + b_1) \cdots g(w_s x_N + b_s) \end{bmatrix}_{N \times S} \quad (4)$$

其中, H 是隐含层输出矩阵, T 是输出结果矩阵, 输入权重 w_i 和隐含层偏置 b_i 随机产生, 输出权重 β 可通过 $\beta = H^+ T$ 计算, H^+ 是 H 的广义逆矩阵。

3 POI 推荐模型

本节将对本文提出的基于 ELM 的 POI 推荐模型展开论述, 首先分析基于 ELM 的 POI 推荐模型框架, 然后对用户产生的签到数据集, 运用特征提取方法, 提取用户的个人偏好、朋友偏好、类型偏好、流行度偏好等特征, 最后使用用户签到数据集中提取的上述特征, 为每一个 POI 形成一个特征向量, 将特征向量集合组成训练样本训练 ELM 分类器, 最终根据 ELM 分类器输出结果, 将 POI 推荐给用户。

3.1 模型框架概述

POI 推荐模型如图 2 所示。本文提出的基于

ELM 的 POI 推荐框架由 2 部分组成:

(1) 特征提取部分。为用户未访问的 POI 提取用户的个人偏好 (UF)、朋友偏好 (SF)、类型偏好 (CF)、流行度偏好 (PF) 等特征, 这些偏好特征都是通过用户签到数据集提取得来, 最终为每一个 POI 形成一个特征向量, 将 POI 特征向量集合作为 ELM 的输入层。

(2) ELM 分类器训练特征部分。特征向量组成的训练样本通过 ELM 分类器训练, 决定哪些用户未访问的 POI 应该推荐给当前用户。

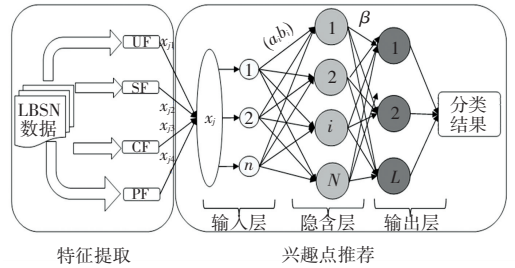


图 2 POI 推荐模型

Fig. 2 POI recommendation model

3.2 特征提取

个人偏好代表着用户个人对待推荐 POI 的喜好程度, 朋友偏好代表着用户的朋友对该 POI 的喜好程度, 类型偏好代表着该 POI 在其所属相同类型标签的 POI 集合中的受欢迎程度, 流行度偏好代表该 POI 在该区域人们的喜好程度。提取到的影响用户 POI 偏好的 4 个相关特征, 可以通过图 3 对这些偏好关系进行描述和理解。

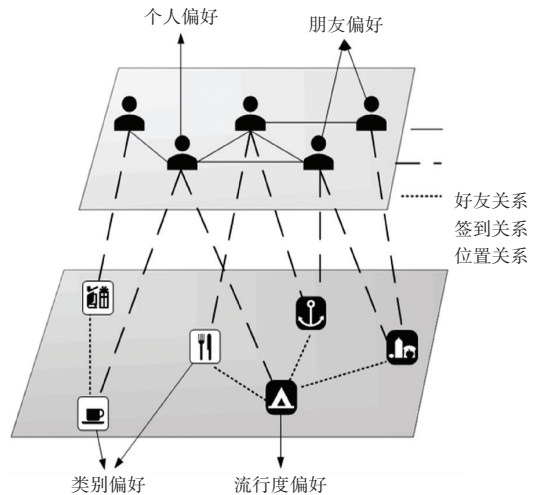


图 3 位置社交网络关系

Fig. 3 Location social network relationships

3.2.1 个人偏好特征

用户的个人偏好特征 UF , 对该 POI 的偏好起着重要的作用。向用户推荐的 POI 是用户没有进行过签到活动的 POI, 想要获取用户对该 POI 的偏好是比较困难的, 因此本文将从 2 方面来计算 UF 的值。一方面, 是寻找用户历史签到过并且与待推荐 POI 具有相同类型标签的 POI, 通过计算用户对上述 POI 的偏好程度, 间接获得用户对当前待推荐 POI 的偏好程度; 另一方面, 计算用户当前位置与待推荐 POI 之间的距离评分值, 在某一区域用户一般喜欢访问距离所处位置较近的 POI。定义 UF 公式如下:

$$UF = \frac{|C_u^p|}{|P_u| \cdot dis(l_u, l_p)} \quad (5)$$

其中, C_u^p 表示与待推荐 POI 所属同一类型标签下的 POI, 且用户 u 进行过签到活动; $|C_u^p|$ 表示签到的次数; $|P_u|$ 表示用户 u 签到的总次数; l_u, l_p 分别表示用户 u 与兴趣点 p 的经纬度坐标; $dis(l_u, l_p)$ 表示两者之间的距离。

3.2.2 朋友偏好特征

用户的行为习惯可能在一定程度上受到朋友的影响, 如用户可能经常与关系较好的朋友一起逛街, 则朋友比较喜欢去的地方用户也可能去。通过用户在社交平台的朋友集合 F 寻找用户的 POI 偏好得分, 由此定义 SF 公式如下:

$$SF = \frac{\sum_{f \in F} \frac{|p'_f|}{|P_f|} \cdot \frac{|F_u \cap F_f|}{|F_u \cup F_f|}}{|F|} \quad (6)$$

其中, f 表示用户朋友集合 F 中的一个朋友; $|p'_f|$ 表示朋友 f 在被推荐 POI 的签到次数; $|P_f|$ 表示朋友 f 的总签到次数; $|F_u \cap F_f|$ 表示用户 u 与朋友 f 共有的好友数量; $|F_u \cup F_f|$ 表示用户 u 与朋友 f 的朋友数量之和; $|F|$ 表示用户 u 的朋友数量。 $\frac{|F_u \cap F_f|}{|F_u \cup F_f|}$ 比值越高, 说明当前用户 u 与朋友 f 之间的亲密程度越强。

3.2.3 类型偏好特征

CF 用来判断该 POI 在同类型标签下的受欢迎程度, 如在一个区域会有很多餐饮类的 POI, 若要推荐给用户比较受欢迎的、并且认可度比较高的餐饮类 POI, 就需要通过计算来获得每个 POI 在同类型标签下的得分。定义 CF 公式如下:

$$CF = \frac{|p|}{\sum_{p' \in C^p} |p'|} \quad (7)$$

其中, $|p|$ 表示待推荐 POI 所有用户的总签到次数, C^p 表示与待推荐 POI 同类别的 POI。

3.2.4 流行度偏好特征

PF 用来判断在该区域待推荐 POI 的受欢迎程度, 如果在一个区域该 POI 签到次数比附近其他 POI 签到次数明显偏多, 则说明该 POI 比较受人们喜欢。定义 PF 公式如下:

$$PF = \frac{|p| - |p_{\min}^a|}{|p_{\max}^a| - |p_{\min}^a|} \quad (8)$$

其中, $|p|$ 表示待推荐 POI 所有用户的总签到次数; p_{\max}^a 表示当前待推荐 POI 附近签到次数最多的 POI; p_{\min}^a 表示当前待推荐 POI 附近签到次数最少的 POI。

4 训练 ELM 分类器

在特征提取之后, 得到每个用户与 POI 生成的训练样本, 用 (x_j, t_j) 表示。其中, x_j 由 UF 、 SF 、 CF 、 PF 组成; t_j 是一个结果向量, 其值为 0 或 1: 若 t_j 为 1 则将 POI 推荐给用户; 若 t_j 为 0 则不会将 POI 推荐给用户。

将样本集合输入 ELM 分类器的输入层进行样本训练, 最后通过测试阶段, 输出用户的 POI 预测结果。POI 推荐算法描述如下。

输入 训练样本数量 N , 测试样本数量 D

输出 POI 推荐结果

1. $x, t \leftarrow \emptyset$
2. for $j \leftarrow 1$ to N do
3. $x_j \leftarrow UF \cup SF \cup CF \cup PF$
4. end for
5. Training $M \leftarrow ELM(x_j, t_j)$
6. Test $R \leftarrow M(x_D)$
7. Return R

算法中, 输入训练样本数量为 N , 测试样本数量为 D 。

第一阶段进行特征提取, 为每个待推荐 POI 生成一个特征向量(第 3 行);

第二阶段训练 ELM 分类器, 算法第 5 行是将第一阶段提取的特征与结果向量 t_j 组合形成训练样本; 通过训练完成的 ELM 分类器, 将测试样本 x_D 输入分类器(第 6 行)获得分类结果; 最后返回分类结果。

5 实验结果分析

5.1 实验数据集

本文使用 Foursquare 与 Twitter 中美国纽约的数据集见表 1。其中, Foursquare 数据集中包括 40 543 名用户和 103 254 个签到记录, Twitter 数据集中包括 32 237 名用户和 78 642 个签到记录。将研究中, 数据集形成 $\langle user, POI \rangle$ 数据对, 每个签到记录代表一个 POI 位置。

表 1 用户和签到数据集

Table 1 User and check-in datasets

数据来源	用户	签到记录
Foursquare	40 543	103 254
Twitter	32 237	78 642

5.2 评价标准

本文使用精确率 (*precision*), 召回率 (*recall*) 作为评价指标, 计算公式如下:

$$Precision = \frac{|TP|}{|TP| + |FP|} \quad (9)$$

$$Recall = \frac{|TP|}{|TP| + |FN|} \quad (10)$$

其中, $|TP|$ 表示用户接受被推荐的数量; $|FP|$ 表示用户不接受但被推荐的数量; $|FN|$ 表示用户接受的 POI 没有被推荐的数量; $|TP| + |FN|$ 表示用户接受的总数量。

5.3 实验分析

为了验证本文提出的 POI 推荐模型的准确率和效率, 将其与 SVM 模型和误差反向传播算法 BP 模型进行对比。利用用户在 Foursquare 与 Twitter 中产生的签到数据集, 生成特征集合, 并将特征集合输入到 3 个模型中, 采用 75% 的数据分别训练 3 个模型, 25% 的数据作为测试集。其中, Foursquare 数据集实验结果如图 4 所示, Twitter 数据集实验结果如图 5 所示。

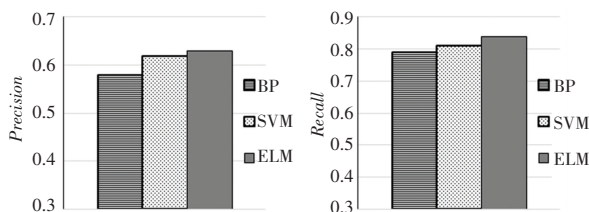


图 4 Foursquare 数据实验结果

Fig. 4 Experimental results of Foursquare data

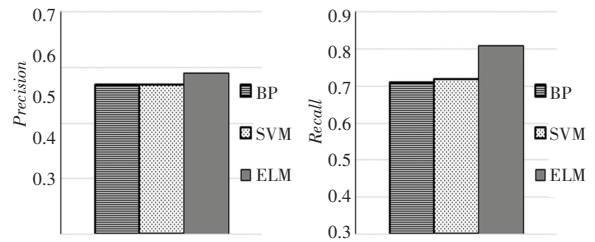


图 5 Twitter 数据实验结果

Fig. 5 Experimental results of Twitter data

从图 4、图 5 中可以看出, 本文提出的 POI 推荐模型运用 ELM 的分类技术, 在精确率和召回率都略优于传统的 SVM 和 BP 模型, SVM 模型略高于 BP 模型。模型运行效率如图 6 所示。

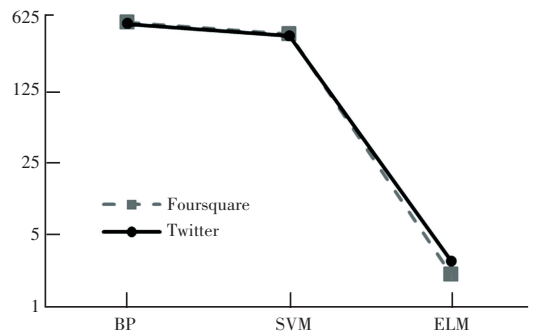


图 6 运行效率图

Fig. 6 Operation efficiency diagram

从图 6 中可以看出, 运用 ELM 模型的运行速度比 SVM 与 BP 模型高很多, 这是由于 ELM 输入权重和偏置是随机产生的, 通常比传统的机器学习方法需要更少的训练时间, 并且倾向于达到最小的训练误差。

通过实验对比, 证明了本文提出的特征提取方法与 ELM 相结合的 POI 推荐模型, 在推荐精度和效率上都有一定的提升, 相比 SVM 和 BP 模型更加适合 POI 推荐。

6 结束语

本文提出了一种基于 ELM 的 POI 推荐模型, 该模型综合考虑了个人偏好、朋友偏好、类型偏好、流行度偏好四个相关因素, 利用 ELM 提供的分类方法, 使用用户签到数据集中提取的 UF 、 SF 、 CF 、 PF 特征, 组成特征向量训练 ELM 分类器。实验表明, 与 BP 和 SVM 模型相比, 本文提出的 POI 推荐模型在准确率和效率上都有一定的提高。

参考文献

- [1] WEI Xuemei, QIAN Yang, SUN Chunhua, et al. A survey of location-based social networks: problems, methods, and future research directions[J]. *Geoinformatica*, 2022, 26(1): 159-199.
- [2] CRANSHAW J, TOCH E, HONG J, et al. Bridging the gap between physical location and - online social networks [C]// *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*. New York:ACM,2010: 119-128.
- [3] ZHENG Yu. Location-based social networks: Users [M]// *Computing with spatial trajectories*. New York, USA: Springer, 2011: 243-276.
- [4] IALAM M A, MOHAMMAD M M, DAS S S S, et al. A survey on deep learning based Point-of-Interest (POI) recommendations [J]. *Neurocomputing*, 2022, 472: 306-325.
- [5] LIU K, ZHENG W, XIAO Y, et al. POI Recommendation algorithm based on region transfer collaborative filtering [C]// *2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. Hangzhou:IEEE, 2022: 903-907.
- [6] ZENG Jun, TANG Haoran, HE Xin. RCFC: a region-based POI recommendation model with collaborative filtering and user context [C]// *International Conference on Collaborative Computing: Networking, Applications and Worksharing*. Cham: Springer, 2021: 656-670.
- [7] ZHANG Zhipeng, ZHANG Yao, REN Yonggong. Employing neighborhood reduction for alleviating sparsity and cold start problems in user-based collaborative filtering [J]. *Information Retrieval Journal*, 2020, 23(4): 449-472.
- [8] XUE Feng, HE Xiangnan, WANG Xiang, et al. Deep item-based collaborative filtering for top-n recommendation [J]. *ACM Transactions on Information Systems (TOIS)*, 2019, 37(3): 1-25.
- [9] CAO Keyan, GUO Jingjing, MENG Gongjie, et al. Points-of-interest recommendation algorithm based on LBSN in edge computing environment [J]. *IEEE Access*, 2020, 8: 47973-47983.
- [10] PAGE L, BRIN S, MOTWANI R, et al. The PageRank citation ranking: Bringing order to the web [C]// *Proceedings of the 7th International World Wide Web Conference*. Brisbane, Australia: dblp, 1998: 161-172.
- [11] KLEINBERG J M. Authoritative sources in a hyperlinked environment [J]. *Journal of the ACM (JACM)*, 1999, 46(5): 604-632.
- [12] CHAKRABARTI S, DOM B, RAGHAVAN P, et al. Automatic resource compilation by analyzing hyperlink structure and associated text [J]. *Computer Networks and ISDN Systems*, 1998, 30(1-7): 65-74.
- [13] CHEN Ling, YING Yuankai, LYU D, et al. A multi-task embedding based personalized POI recommendation method [J]. *CCF Transactions on Pervasive Computing and Interaction*, 2021, 3: 253-269.
- [14] 汤佳欣,陈阳,周孟莹,等. 深度学习方法在兴趣点推荐中的应用研究综述 [J]. *计算机工程*, 2022, 48(1): 12-23.
- [15] ZHAO Xiangguo, MA Zhongyu, ZHANG Zhen. A novel recommendation system in location-based social networks using distributed ELM [J]. *Memetic Computing*, 2018, 10(3): 321-331.
- [16] WU Yuxia, LI Ke, ZHAO Guoshuai, et al. Personalized long- and short-term preference learning for next POI recommendation [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2020, 34(4): 1944-1957.
- [17] WANG Zhaobo, ZHU Yanmin, WANG Chunyang, et al. Adaptive graph representation learning for next POI recommendation [C]// *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Taipei: ACM, 2023: 393-402.
- [18] YAN Xiaodong, SONG Tengwei, JIAO Yifeng, et al. Spatio-temporal hypergraph learning for next POI recommendation [C]// *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Taipei: ACM, 2023: 403-412.
- [19] HUANG Guangbin, ZHU Qinyu, SIEW C K. Extreme learning machine: A new learning scheme of feedforward neural networks [C]// *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541)*. Budapest, Hungary: IEEE, 2004, 2: 985-990.
- [20] 马萌萌. 基于深度学习的极限学习机算法研究 [D]. 青岛: 中国海洋大学, 2015.