引文格式:牛雪磊,杨军,闫浩文.顾及微地图推荐因子的深度信念网络[J].武汉大学学报(信息科学版),2025,50(7):1415-1424.DOI:10.13203/j.whugis20230053

Citation: NIU Xuelei, YANG Jun, YAN Haowen. Deep Belief Networks Considering WeMaps' Recommendation Factors[J]. Geomatics and Information Science of Wuhan University, 2025, 50(7):1415-1424.DOI:10.13203/j.whugis20230053

顾及微地图推荐因子的深度信念网络

牛雪磊[1] 杨 军[1,2] 闫浩文[1]

1 兰州交通大学测绘与地理信息学院,甘肃 兰州,730070

2 兰州交通大学电子与信息工程学院,甘肃 兰州,730070

摘 要:在微地图推荐系统的冷启动阶段,针对传统深度信念网络无法高效挖掘数据中的细粒度信息而导致推荐精度下降的问题,提出了顾及微地图推荐因子的深度信念网络(deep belief network, DBN)。利用聚类的结构化信息和DBN中丰富的语义信息解决推荐系统在冷启动阶段推荐精度低的问题,并且挖掘出用户和微地图的隐语义交互信息,达到"千人千面"的推荐效果。首先,采用针对噪声数据的基于密度的空间聚类算法对用户与微地图历史样本进行聚类;然后,选取出核心点(即推荐因子)、边界点及离群点,将聚类后的推荐因子构建成不同的可视层与隐含层;最后,将受限玻尔兹曼机的可视神经元替换为高斯单元,利用DBN对输入数据从低层到高层进行逐层微调训练,以提升推荐的准确性。所提方法在美食、自驾、旅游、校园4类微地图数据集上推荐的精度分别为0.775 32、0.768 18、0.775 18、0.774 64,推荐的均方根误差分别为0.190 78、0.194 76、0.190 33、0.190 92,可以为用户准确地推荐感兴趣的微地图信息。

关键词:微地图;推荐系统;DBN;推荐因子;特征值

中图分类号:P208 **文献标识码**:A **收稿日期**:2023-04-24

DOI:10.13203/j.whugis20230053 **文章编号**:1671-8860(2025)07-1415-10

Deep Belief Networks Considering WeMaps' Recommendation Factors

NIU Xuelei[1] *YANG Jun*[1,2] *YAN Haowen*[1]

1 Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou 730070, China

2 School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China

Abstract: **Objectives:** In the cold start phase of the WeMaps' recommendation system, traditional deep belief networks have low efficiency in mining fine-grained information, resulting in a decrease in recommendation accuracy. Therefore, we propose a deep belief network (DBN) considering WeMaps' recommendation factors (DBRFact). The structured information of clustering and rich semantic information in the deep belief networks are used to adress the problem of low recommendation accuracy during the cold start. In addition, the implicit semantic interaction information between the users and the WeMaps is mined to implement customized recommendations. **Methods:** First, a density-based spatial clustering of application with noise clustering algorithm is employed to cluster the historical samples of the users and the WeMaps. Second, the core points (i.e., recommendation factors), boundary points, and outliers are selected, and the recommendation factors are constructed into different visual and hidden layers. Finally, the visual neurons of the restricted Boltzmann machine are replaced with Gaussian units. A deep belief network is used to fine-tune the input data from low level to high level to improve the recommendation accuracy. **Results:** The accuracies on the food maps, driving maps, tour maps, and school maps datasets are 0.775 32, 0.768 18, 0.775 18 and 0.774 64, respectively, and the root mean square errors are 0.190 78, 0.194 76, 0.190 33 and 0.190 92, respectively. **Conclusions:** The experimental results have showed that the proposed algorithm can accu-

基金项目:国家自然科学基金(42261067);2021年度中央引导地方科技发展资金(2021-51);兰州市人才创新创业项目(2020-RC-22);兰州交通大学天佑创新团队(TY202002)。

第一作者:牛雪磊,博士,主要从事深度学习、地理信息系统等研究。NXLnxl186186@163.com

通信作者:杨军,博士,教授。yangj@mail.lzjtu.cn

rately recommend the interested WeMaps' information for users.

Key words: WeMaps; recommendation system; DBN; recommendation factors; eigenvalue

现代社会正在经历从工业化到信息化的转变,自媒体是在互联网技术的发展普及和后现代文化的双重影响下出现的[1-2],以平民化、个性化、低门槛、易操作、微内容、交互强、传播快[3]为特点的自媒体信息传播模式已经对各行各业产生深远的影响。地图是地理空间信息的记录载体和传播工具,对人类社会至关重要[4-7]。传统地图由于制作门槛高、更新难、传播慢等特点,无法满足自媒体时代大众地理信息传播的需求。在自媒体时代制作地图需有大众用户的实时参与,在分发和传播上需要及时、方便、快捷,在信息传播时不但具有由点到面的广播功能,而且要有点到点[8]和面到面的互播功能,因此微地图的概念应运而生。微地图[8]是面向平民大众的"草根"地图,对精度等数学基础要求不高,制作者无需进行严格的专业培训,地图用户能够参与地图制作,地图可以如微信一样方便、快捷地交互传播和应用,在投入使用后势必产生海量的数据,可能因信息量远超微地图用户信息需求、信息处理及信息利用能力,导致用户无法准确挑选与运用有效信息。所以,为满足千人千面的个性化[9]推荐,有必要从历史数据中尽可能挖掘潜在的经济和社会价值,从而预测用户的需求并推荐给用户感兴趣的内容。

推荐算法[10]为用户推荐相关领域感兴趣的内容,也为商业决策提供帮助。传统的商品、电影推荐算法一般采用显式评分的方式获取用户偏好和历史行为,结合用户的历史行为采用协同过滤的推荐方法[11]、采用基于内容的推荐方法[12]或采用混合推荐方法[13]就能够实现较准确的推荐结果。其中,文献[14]使用幂函数的方式进行协同过滤,利用用户与项目之间的交互信息为用户推荐,然而该算法遭遇严重的冷启动问题(新用户和新项目往往没有历史评分数据)。采用内容的推荐方法,文献[15]使用用户感兴趣的项目寻找其他类似属性的候选项目,该算法严重依赖于人工设计特征,其有效性及可扩展性非常有限,降低了基于内容推荐方法的性能。微地图推荐相对于传统电影、商品推荐主要存在以下两个方面的困难与挑战:(1)用户的偏好往往以隐式的方式存在于历史签到数据,以签到次数或签到频率表示偏好,致使无法准确理解用户具体的偏好程度;(2)现实世界中位置众多,而每个用户访问的地点只是其中极少的一部分,用户没有在访问的地点进行签到行为。由此造成了微地图历史签到信息存在数据量大、间歇性、稀疏性等问题,用户对推荐的内容体验效果较差,亟需收到与自己兴趣相关的高质量信息。

随着互联网中越来越多的数据能够被获取,包括图像、文本、标签在内的多源异构数据,其蕴含着丰富的用户行为信息以及个性化需求信息,能够缓解推荐系统中冷启动问题,因此,融合多源异构辅助信息的混合推荐方法[16-17]越来越受到重视。由于辅助信息往往具有多模态、数据异构、大规模、数据稀疏和分布不均匀等复杂特征,融合多源异构数据的混合推荐方法依然面临着严峻的挑战[18-19]。

微地图个性化推荐[20-21]取得了突破性进展,为微地图推荐系统的研究带来新的机遇。为解决冷启动阶段的推荐精度低的问题,挖掘用户与微地图的隐语义交互信息,聚类分析是推荐系统中分析数据与挖掘应用的有效方法,通过衡量研究对象之间的相似度,将相似度高的对象划分为同类。针对噪声数据的基于密度的空间聚类算法[22](density-based spatial clustering of application with noise, DBSCAN)的主要目标是寻找被低密度区域分离的高密度区域,无需指定簇的个数,能够对任意形状的数据集进行聚类,常用于离群点监测。文献[23]提出了嵌入异构情感意识的兴趣推荐,整合基于社区情感提取的潜在概率分布得到情感与感知网络嵌入模型,学习用户和文本感情等重要因素在低维嵌入网络中表达,并利用联合训练对异构网络中所有的边进行交替采样。然而,该算法偏向于学习相邻特征的组合,难以适用于无时序关系的数据。受限玻尔兹曼机[24](restricted Boltzmann machine, RBM)在玻尔兹曼机(Boltzmann machine, BM)的基础上,采用对比散度算法[25],只需迭代一次就可获得对模型的估计,训练效率在一定程度上得以提升。深度信念网络[26](deep belief network, DBN)是一种由多层非线性变量连接组成的生成模型,通过逐层抽取输入信息的无监督学习和用固定标签微调整个网络的监督学习,解决了学习多个隐含层参数难度大的问题。

综上所述,随着微地图推荐系统应用场景的不断扩大及复杂性的增强,现有研究成果在单独

使用DBN进行模型训练时,未深入研究用户的显式特征,忽略了用户的差异性,致使用户自身的多样性特征没有被充分挖掘和利用,导致推荐算法的灵活性较差;而若单独使用DBSCAN聚类算法分别对用户、微地图数据进行聚类,不能有效挖掘用户与微地图交互的隐式特征,由于隐式特征具有灵活多变、不易捕捉的特点,缺乏隐式特征时往往存在难以跨越从低阶特征到高阶特征的语义鸿沟,导致数据丰富但语义信息匮乏的问题,从而降低了微地图在冷启动阶段推荐的精度。因此,本文结合DBSCAN和DBN,提出顾及微地图推荐因子的DBN(DBN considering We-Maps recommendation factors, DBRFact),可以融合描述微地图推荐系统中微地图、用户之间的关系模型,使得微地图推荐算法不再局限于分析用户或地图之间关系,从而使用户获得个性化的精准推荐。

1　原理方法

用户与微地图数据不仅数量大,而且存在噪声,导致用户与微地图数据对DBN的预测精度产生较大影响。如果将所有的用户与微地图数据都加入到预测模型中,会带来大量的数据噪声,降低推荐的精度。本文提出DBRFact,在DBN训练前,先使用DBSCAN对微地图用户进行预处理,选取出核心点(微地图的推荐因子),即从影响用户选择的优先程度、用户因素界定的难易程度等不同方面考虑,细粒度挖掘推荐因子,在编码多尺度特征向量时为DBN提供有效的推荐因子进行模型训练,以提高微地图推荐系统在冷启动阶段的推荐精度。

1.1　模型架构

DBRFact的整体结构如图1所示,主要由4部分组成:输入数据、DBSCAN、DBN和输出微地图推荐结果。

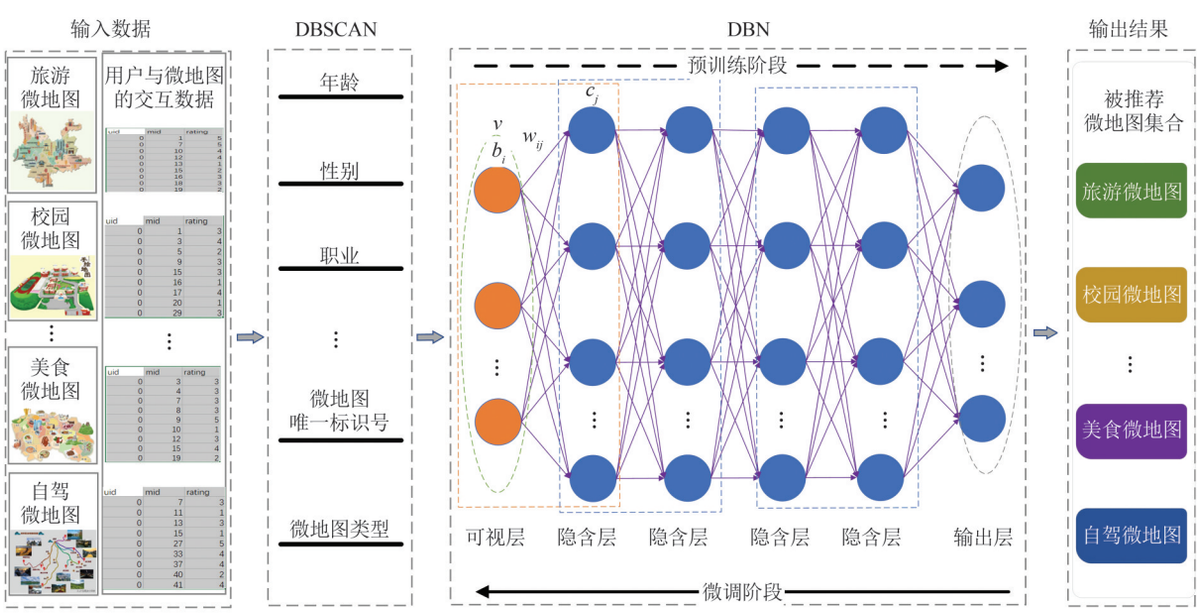


图1　DBRFact的整体结构图

Fig. 1　Overall Structure of DBRFact

1)输入数据。通用的各类专题微地图的历史数据,以及用户与各类专题微地图的历史交互数据。

2)DBSCAN。利用DBSCAN算法对微地图数据集、微地图与用户交互的历史样本数据进行聚类,并选择出核心点即影响微地图推荐的推荐因子。

3)DBN。DBN输入的可视层必须是二元值,因此将其第一层RBM替换为高斯二元受限玻尔兹曼机(Gaussian-binary RBM, GRBM),GRBM的输出作为第二层RBM的输入,以此类推,直到第五层RBM。如图1所示,DBN结构中的v为输入,$O(v)$为模型的输出,b_i表示可视层节点i的偏置,c_j表示隐含层节点j的偏置。在开始训练网络之前,随机初始化权重值w_{ij}、可视层v的阈值和隐含层h的阈值,然后逐层预训练和微调更新w_{ij}、v和h的值,直到达到收敛。

4)输出微地图推荐结果。模型最终输出用户感兴趣的微地图推荐结果,如旅游微地图、校园微地图、美食微地图和自驾微地图等。

1.2 DBSCAN

微地图数据集中的密度分布是对微地图空间现象的一种场表达,能够反映不同用户群体对不同专题微地图的偏好程度。在实际的地理空间中,人类的社会经济活动在不同的空间尺度下往往表现为若干兴趣点的聚集,并在不同的空间统计单元中形成强度等级各异的聚集中心。不同微地图区域下聚集不同的中心区,通过密度噪声应用空间聚类能够较全面地反映微地图的空间分布,表达不同层次微地图中心的显著程度。基于密度的噪声应用聚类方法的主要目标是寻找被低密度区域分离的高密度区域,通过过滤低密度区域发现稠密度样本点,能够形成不同形状的簇,并且可以消除孤立点和噪声对聚类质量的影响。本文采用DBSCAN算法,在未知训练样本标记信息的情况下识别出噪声点并生成任意形状的簇,有效区分各类别之间的差异。该算法通过参数(ε, M_i)来描述样本分布紧密程度。其中,ε是邻域的半径,M_i是聚类为一簇所需要的最少采样点的数量。

1.3 DBN

RBM是一种满足玻尔兹曼分布与热力学能量分布的概率图模型,如图2所示。通过训练不断更新网络参数的权重值,使模型中的数据分布尽可能拟合训练数据的分布。由n个可视层节点与m个隐含层节点组成RBM的能量函数$E(v,h)$为:

$$E(v,h)=-\left(\sum_{i=1}^{n}\sum_{j=1}^{m}v_i w_{ij}h_j+\sum_{i=1}^{n}b_i v_i+\sum_{j=1}^{m}c_j h_j\right) \tag{1}$$

式中,v_i表示第i个可视层节点;h_j表示第j个隐含层节点;w_{ij}表示可视节点i与隐含节点j之间的权重值。

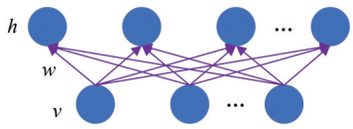


图2 RBM的结构图

Fig. 2 Structure of RBM

给定能量函数后,所有可视层与隐含层的联合概率分布$p(v,h)$的计算公式为:

$$p(v,h)=\frac{e^{-E(v,h)}}{Z} \tag{2}$$

式中,Z为配分函数,也被称为归一化因子,其计算公式为:

$$Z=\sum_{v}\sum_{h}e^{-E(v,h)} \tag{3}$$

在采样过程中,当给定可视层节点状态v时,第j个隐含层节点h_j的激活概率为:

$$p(h_j=1|v)=f\left(c_j+\sum_{i=1}^{n}w_{ij}v_i\right) \tag{4}$$

其中,$f(x)=\dfrac{1}{1+e^{-x}}$。

同理,当给定隐含层节点状态h时,第i个隐含层节点v_i的激活概率为:

$$p(v_i=1|h)=f\left(b_i+\sum_{j=1}^{m}w_{ij}h_j\right) \tag{5}$$

DBN是由多层RBM叠加而成的神经网络,其训练过程使用无监督贪婪算法。具有两个隐含层的DBN的训练过程为:(1)充分训练第一层RBM;(2)先确定第一层RBM的权重和阈值,然后使用其隐含神经元的状态作为第二层RBM输入向量;(3)在完成训练第二层RBM之后,将第二层RBM堆叠在第一层RBM之上。在所有RBM训练完成后,给定原始输入和目标输出作为监督信号,使用梯度下降法对网络进行监督训练,是一个微调过程。DBN的训练过程见图3。

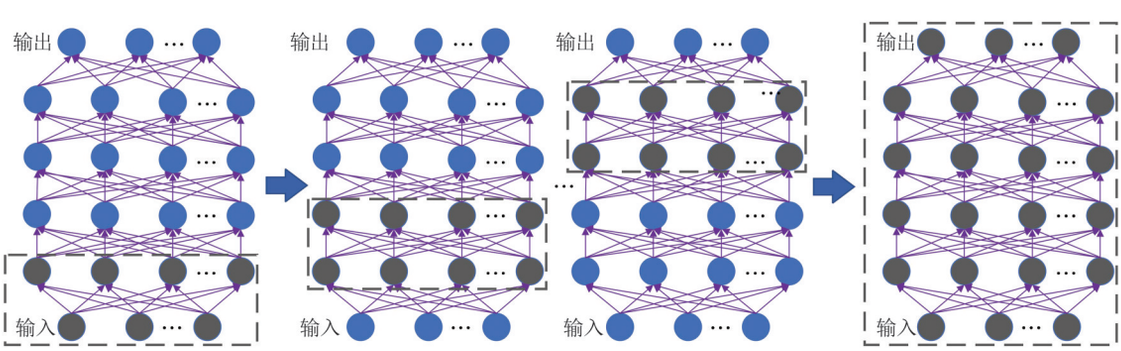


图3 DBN训练过程

Fig. 3 Training Process of DBN

1.4 GRBM

RBM可视层节点的神经元必须是二值的,然而,微地图推荐模型的输入数据是连续分布的。因此,本文将DBN可视层的第一层替换为

GRBM,其能量函数$E(v,h)$的计算公式为:

$$E(v,h)=\sum_{i=1}^{n}\frac{(v_i-b_i)^2}{2\sigma_i^2}-\sum_{j=1}^{m}c_jh_j-\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{v_i}{\sigma_i^2}h_jw_{ij} \tag{6}$$

式中,σ_i是样本数据的方差。

可视层节点的激活概率$p(v_i=1\mid h)$的计算公式为:

$$p(v_i=1|h)=f(b_i+\sigma_i\sum_{j=1}^{m}w_{ij}h_j,\sigma_i^2) \tag{7}$$

为有效挖掘用户对不同专题微地图的长期兴趣及短期兴趣,有必要从用户与微地图的交互数据中提取用户特征向量和微地图特征向量,其中用户特征向量包括用户的显式特征和隐式特征,微地图特征向量包括微地图的标号唯一标识符和微地图的名称。因此,本文采用激活函数来学习用户特征及用户与微地图特征之间潜在的非线性交互特征,从而提供更丰富的语义表达,为微地图推荐精度的提升奠定一定的基础。此外,本文采用的激活函数在训练过程中可以有效避免梯度消失的问题,从而加快反向传播算法的收敛速度。

2　DBRFact实验及结果分析

2.1　实验环境及数据集

实验的硬件环境为 Intel Core i9-9900k+RTX 3090,软件环境为 Windows 10+CUDA 11.1+cuDNN8.5.0+TensorFlow+Python3.6。实验迭代次数为100次,学习率为0.005,DBN的RBM层数为5层。

使用爬虫技术对百度图片进行爬取,并增加手绘地图,经过数据清洗、语义对齐等操作构建微地图数据集,其记录了2 915张微地图数据相关的地图信息,包括978张旅游微地图数据、600张校园微地图数据、682张美食微地图数据、655张自驾微地图数据等信息。微地图数据集的字段信息如表1所示,包括微地图的唯一标识号、微地图名称。

表1　微地图数据集中的字段信息

Table 1　Field Information of WeMaps DataSets

旅游微地图		校园微地图		美食微地图		自驾微地图	
唯一标识号	名称	唯一标识号	名称	唯一标识号	名称	唯一标识号	名称
1	天桥山	1	上海视觉艺术学院	1	撒汤	1	武威
2	洞口县	2	武汉东湖学院	2	幺铺毛肚	2	兰州
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

目前,由于DBRFact推荐系统在冷启动阶段缺少相应的用户数据,通过调查问卷的方法收集用户信息和相应的用户行为,共收到有效数据604份,字段信息如表2所示。从影响用户选择的优先程度、用户因素界定的难易程度等不同方面考虑,将用户属性情景分为显式和隐式两种。显式特征是描述用户的基本情况,属性相对清晰,短时间内稳定不变,不会发生突变情况。

表2　微地图用户显式属性信息

Table 2　Explicit Attribute Information of WeMaps Users

因素	阈值或选项
性别	男/女
年龄	18岁以下、18~24岁、25~34岁、35~44岁、45~49岁、50~55岁、56岁以上
职业	未定义、教育从业者、艺术家、行政人员、研究生、客服、医生、管理者、农民、家庭主妇、律师、程序员、退休人员、销售、科学家、创业者、工程师、工人、失业者、作家
⋮	⋮

本文除了分析用户年龄、性别等基础用户信息外,结合微地图大众性的特点,还将重点分析用户职业,更能表现出用户对微地图的掌握及使用情况。用户隐式特征主要研究用户在不同情景下的心理变化情况,有着灵活多变、不易捕捉的特点。在微地图用户研究中,具体研究用户需求心态、行为习惯等因素,把握用户在制图和用图过程中的操作习惯,如表3所示。在用户使用微地图的过程中,隐式属性对用户需求具有隐性持续的影响,但由于用户不能准确表达,相关算法也不能准确捕捉,无法进行定质、定量的描述,所以在用户表现出需求和实现需求的操作过程中,对用户隐式特征进行捕捉分类。先将隐性属性因素进行说明,参考相关心理学文献[27],对相

关因素阈值或选项进行分类。

表3 微地图用户隐式属性信息

Table 3 Implicit Attribute Information of WeMaps Users

因素	描述	阈值或选项
需求心态	影响微地图的使用人群范围	工作、生活、社会需求
识图能力	影响用户获取地图信息的难易程度	初级、中级、高级
分享意识	影响微地图的传播范围	薄弱、一般、良好
⋮	⋮	⋮

2.2 评价指标

微地图推荐算法的性能采用准确率与均方根误差(root mean square error, RMSE)来评估。

1)准确率$f(A)$。准确率是指微地图推荐算法在训练过程中采样得到的样本量符合用户偏好样本量的占比,准确率越高,说明微地图地名推荐越准确,其计算公式为:

$$f(A)=\frac{N_{TP}+N_{TN}}{N_{TP}+N_{FP}+N_{TN}+N_{FN}} \tag{8}$$

式中,N_{TP}表示预测值为1、真实值为1的真正例的数量;N_{FP}表示预测值为1、真实值为0的假正例的数量;N_{TN}表示预测值为0、真实值为0的真负例的数量;N_{FN}表示预测值为0、真实值为1的假负例的数量。

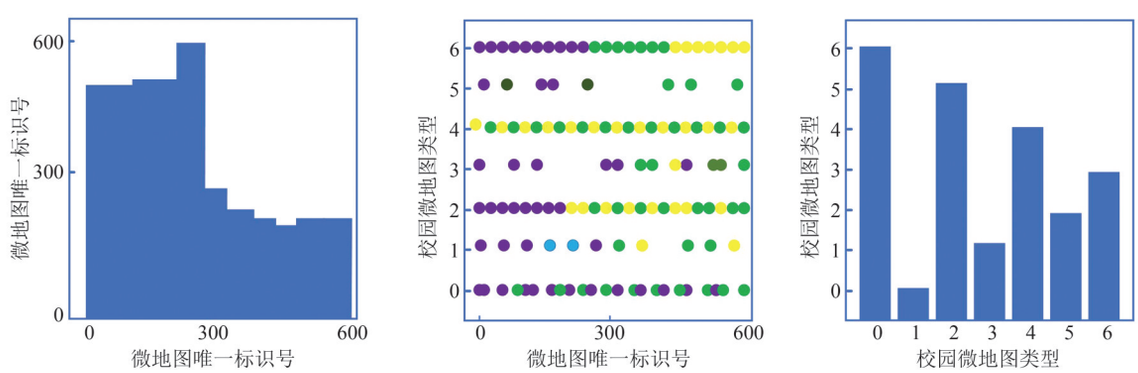
2)RMSE $f(R)$。RMSE是指推荐算法训练过程中采样后的样本$\hat{y}(i)$与原始样本$y(i)$的偏差,其计算公式为:

$$f(R)=\sqrt{\frac{1}{t}\sum_{i=1}^{t}[y(i)-\hat{y}(i)]^2} \tag{9}$$

式中,t表示推荐微地图的数量;$y(i)$表示推荐列表中原始微地图标签真实值;$\hat{y}(i)$表示推荐列表中采用吉布斯采样后微地图标签值。

2.3 校园微地图的DBSCAN

以校园微地图为例,使用DBSCAN算法进行聚类,探究校园微地图的推荐因子。该算法通过参数(ε,M_i)来描述样本分布紧密程度,本文设置$\varepsilon=2$,$M_i=12$,聚类结果如图4所示。将所有的数据点绘制在图上,可以直观呈现校园微地图唯一标识号特征与校园微地图类型特征的分布情况,用来探究微地图唯一标识号与校园微地图类型是否是影响校园微地图推荐的推荐因子。可以看出,唯一标识号共有600个簇(即数据集中包含600张校园微地图),校园微地图类型共有7个簇(用数字0~6分别表示综合类院校、艺术类院校、理工类院校、医科类院校、财经农业师范类院校、技术学院类、其他类)。从校园微地图类型角度分析,占比最多的是综合类院校,占比第二多的是理工类院校,占比最少的是艺术类院校。校园微地图存在长尾分布,导致校园微地图样本分布不均衡,对于热门推荐的校园微地图样本,模型对此类样本学习效果较好,而对长尾效应的冷门推荐校园微地图,模型对此类的样本理解不够充分,长此以往导致总体推荐精度下降。因此,为了能够有效挖掘具有长尾效应的校园微地图,可为用户个性化推荐不同类型的校园微地图,验证了本文算法在对校园微地图数据集历史样本的聚类过程中,能够细粒度挖掘校园微地图的推荐因子(即校园微地图的类型),为DBN在编码多尺度特征向量时有效结合推荐因子信息进行训练奠定了一定的基础。



(a) 微地图唯一标识号的核密度估计图 (b) 微地图唯一标识号与校园微地图类型之间的核密度散点图 (c) 微地图类型的核密度估计图

图4 校园微地图的DBSCAN结果图

Fig. 4 DBSCAN Results for SchoolWeMaps

2.4 基于DBSCAN的校园微地图用户数据集聚类

对用户数据使用DBSCAN算法进行聚类,探究校园微地图的用户数据集推荐因子与微地图推荐结果的相关性。仍然通过参数(ε,M_i)来描述样本分布紧密程度,本文设置$\varepsilon=2$,$M_i=12$,

聚类结果如图5所示。

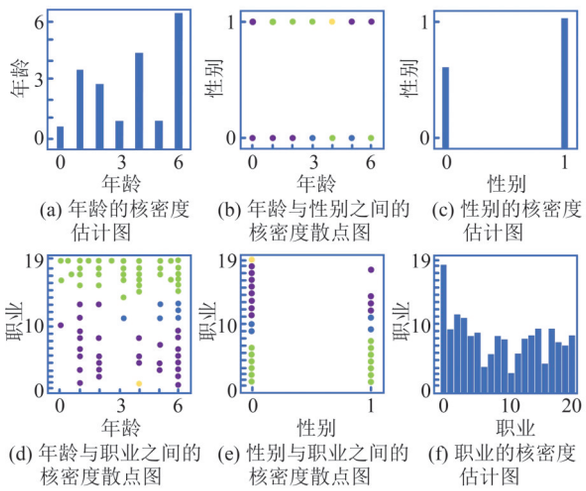


图5　校园微地图用户的DBSCAN聚类结果图

Fig. 5　DBSCAN Clustering Result of SchoolWeMap Users

将所有的数据点绘制在图5中,可以直观呈现校园微地图的用户数据集中年龄特征与性别特征、年龄特征与职业特征的分布情况,用来探究年龄、性别和职业等推荐因子对校园微地图的推荐结果的影响。图5中,年龄分为7个簇(用数字0～6分别表示18岁以下、18～24岁、25～34岁、35～44岁、45～49岁、50～55岁及56岁以上)。性别分为2个簇(用数字0和1分别表示男性和女性)。职业分为20个簇(用数字0～19分别表示未定义、教育从业者、艺术家、行政人员、研究生、客服、医生、管理者、农民、家庭主妇、律师、程序员、退休人员、销售、科学家、创业者、工程师、工人、失业者、作家)。从年龄角度分析可知,占比最多用户群体的年龄在56岁以上,占比第二多的用户群体为45～49岁年龄段的用户,占比第三多的用户群体为18～ 24岁,占比最少的用户群体是18岁以下的用户。从职业角度分析可知,占比最多的为未定义用户,占比第二多的用户群体是艺术家,占比第三多的用户群体是行政人员,占比最少的用户群体是律师。由此,验证了在对校园微地图的用户数据集的历史样本聚类过程中,本文算法能够细粒度挖掘校园微地图用户的推荐因子,即用户的年龄、性别和职业,为DBN在编码多尺度特征向量时有效结合上下文信息进行训练奠定一定的基础,使用DBRFact产生推荐结果后,可以研究不同年龄、不同职业、不同性别的用户个性化推荐因子与推荐结果的相关性。

2.5　校园微地图预测

以214号用户为例,其对不同类型校园微地图的关注程度汇总如表4所示。214号用户分别对综合类院校评分数据为46条,艺术类院校为3条,理工院校为41条等,其中综合类院校占比最多,其次是理工类院校,占比最少的是艺术类院校。按照本文算法为214号用户预测感兴趣的10个校园微地图,预测结果如表5所示,推荐得分排名前3位的校园微地图分别为78号上海海洋大学、322号广西大学西校区、407号江西理工大学。由此验证了本文算法在编码多尺度特征向量时,能够有效结合校园微地图推荐因子的语义信息。此外,本文推荐算法还为214号用户预测了技术学院类院校,说明算法能够有效挖掘具有长尾效应的校园微地图,可为用户个性化推荐不同类型的校园微地图。

表4　214号用户对不同类型校园微地图的关注程度汇总

Table 4　No. 214 User Prefer Different Types of SchoolWeMaps Summary

编号	校园微地图类型	数量/条
214	综合类院校	46
214	艺术类院校	3
214	理工类院校	41
214	医科类院校	4
214	财经农业师范类院校	21
214	技术学院类院校	9
214	其他类院校	6

2.6　不同专题微地图推荐对比

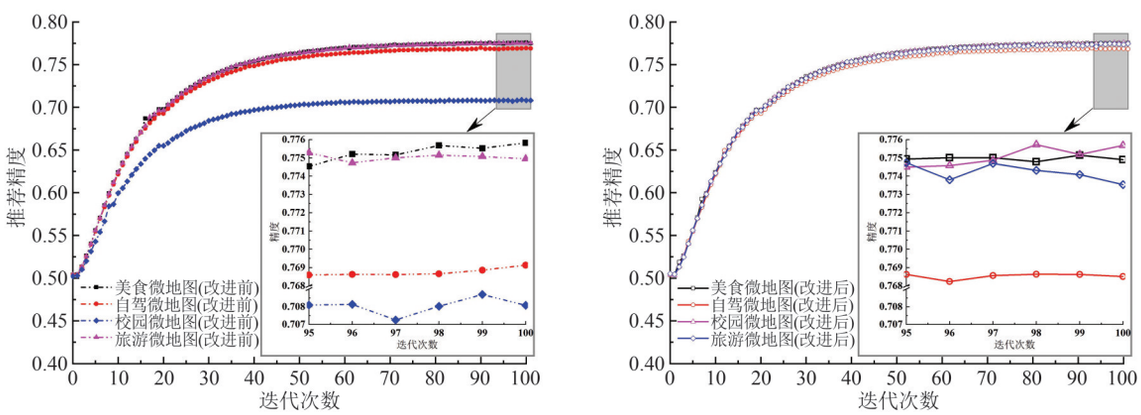
图6和图7分别给出了本文算法在旅游、校园、美食和自驾微地图数据集上推荐的精度、RMSE与迭代次数的统计结果,方框为实验结果(灰色阴影区域)的局部放大图。可以看出,在训练初期,由于网络训练过程中不断优化参数,由DBN网络学习到的特征对微地图数据集的描述准确度不断提高,随着迭代次数的增加,推荐的精度逐渐提高,推荐的RMSE呈下降趋势。当迭代次数达到100次时算法收敛,推荐的精度和RMSE均趋于稳定,模型趋于收敛,说明算法中的参数已达到最优,此时再增加迭代次数对推荐精度和RMSE几乎没有影响。实验结果充分验证了本文算法在训练过程中能够不断提取用户与微地图交互的细粒度语义信息,有效地为用户推荐感兴趣的微地图。

为了探究本文算法的有效性,分别在旅游、校园、美食及自驾微地图数据集上进行训练并测试,实验结果如表6所示。

表5 为214号用户预测10个感兴趣的校园微地图

Table 5 Predict 10 Preferred SchoolWeMaps for No. 214 User

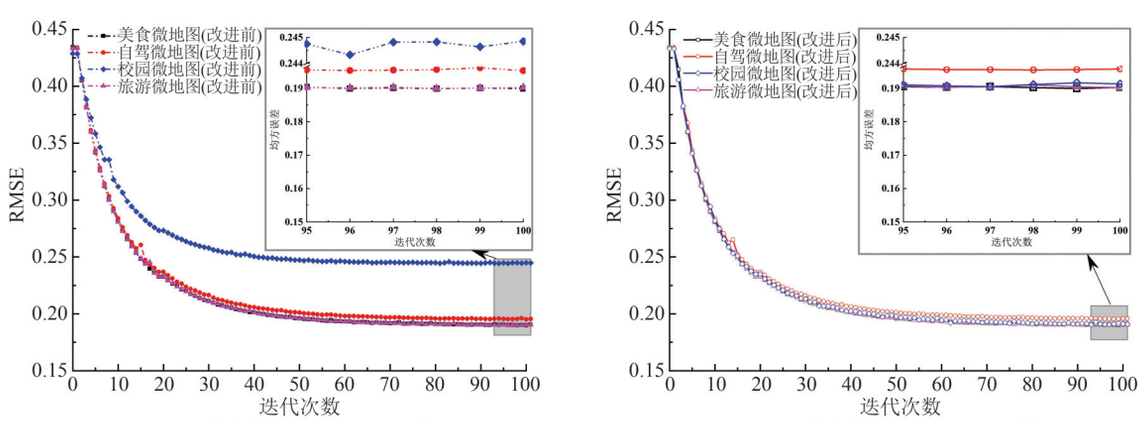
微地图唯一标识号	题目	用户唯一标识号	校园微地图类型	推荐得分
78	上海海洋大学	214	综合类院校	1.000 000
322	广西大学西校区	214	综合类院校	1.000 000
407	江西理工大学	214	理工类院校	0.999 900
311	暨南大学	214	综合类院校	0.999 981
167	西安翻译学院	214	技术类院校	0.999 951
82	香港大学	214	综合类院校	0.999 928
353	广西大学	214	综合类院校	0.999 763
358	华北水利水电大学	214	理工类院校	0.999 725
197	厦门大学	214	综合类院校	0.999 686
113	华南师范大学	214	财经农业师范类院校	0.998 262



(a) 推荐算法改进前推荐精度对比 (b) 推荐算法改进后推荐精度对比

图6 推荐算法改进前后推荐精度对比

Fig. 6 Comparison of Recommendation Accuracy Before and After Algorithm Improvement



(a) 推进算法改进前RMSE对比 (b) 推荐算法改进后RMSE对比

图7 推荐算法改进前后推荐RMSE对比

Fig. 7 Comparison of Recommendation RMSE Before and After Algorithm Improvement

表6 DBRFact与DBN推荐精度与RMSE对比

Table 6 Accuracy and RMSE Comparison of Recommendation Between DBRFact and DBN

微地图类型	推荐精度		RMSE	
	DBN	DBRFact	DBN	DBRFact
美食微地图	0.774 87	0.775 32	0.190 34	0.190 78
自驾微地图	0.768 92	0.768 18	0.195 47	0.194 76
旅游微地图	0.775 31	0.775 18	0.190 21	0.190 33
校园微地图	0.708 04	0.774 64	0.244 74	0.190 92

在这4类专题微地图上只采用DBN的推荐算法与DBRFact算法进行推荐精度和RMSE对比。DBRFact在美食、自驾、旅游和校园微地图数据集上的推荐精度分别为0.775 32、0.768 18、0.775 18、0.774 64,微地图推荐的RMSE分别为0.190 78、0.194 76、0.190 33、0.190 92。微地图推荐精度提升最为明显的是校园微地图,比DBN的推荐精度提升了9.4%;微地图推荐的RMSE

下降最为明显的是校园微地图,比DBN的RMSE降低了21.9%,主要原因是DBN采用微调学习策略能够捕捉细粒度用户与微地图交互的语义信息,同时DBN在编码多尺度特征向量时可以有效结合上下文信息。由于DBRFact更关注于校园与美食微地图的多样性与丰富性,而在一定程度上忽略了旅游与自驾微地图的线路推荐,所以在美食微地图与校园微地图推荐精度提升,而在旅游微地图与自驾微地图推荐精度有所下降。此外,由于旅游微地图与自驾微地图推荐强调线路合理性,不仅局限于推荐的多样性,且用户与这两类微地图交互数据存在数据稀疏性问题,说明这两类专题地图的隐式属性更难挖掘,用户对旅游微地图、自驾微地图的兴趣转移较快且不宜捕捉,但DBRFact也能为用户推荐旅游微地图与自驾微地图。

3　结　语

在微地图推荐系统中的冷启动阶段,针对传统DBN无法高效挖掘数据中的细粒度信息,本文提出了DBRFact算法。该算法首先通过DBSCAN对用户和微地图历史样本进行聚类,选取出核心点(即推荐因子)、边界点以及离群点;其次,把用户与微地图历史样本聚类后的推荐因子构建成不同的可视层与隐含层,利用DBN对输入数据微调训练,从而实现数据的降维并提升推荐的准确性。该算法不仅考虑了用户使用微地图的偏好习惯,而且考虑了微地图地名之间的内部联系,所获得的个性化推荐结果合理且符合个人生活偏好。在美食、旅游、校园及自驾四类微地图数据集上验证了DBRFact算法的推荐精度、RMSE都获得了令人满意的结果。

微地图推荐致力于个性化服务,但是由于用户使用微地图的偏好不同,想达到多维传播的效果,需要更具有个性化的地图服务。由于目前数据只构建了旅游、美食、自驾和校园微地图这4类专题微地图,正在研发的DBRFact推荐系统手机应用端还需构建更多种类的专题微地图,利用深度学习等智能化处理方法分析用户行为,以满足各种不同场景及不同使用人群的需求。未来的研究工作将会针对旅游与自驾微地图中的线路推荐的合理化进行动态研究,对不同的使用人群进行调查问卷和建模,综合考虑推荐系统的可解释性等因素,以满足不同需求的用户,进一步提高推荐效率,完善现有的微地图推荐系统。

参　考　文　献

[1] 赵前卫,马缘园. 自媒体时代微信谣言传播特点初探[J]. 新闻研究导刊,2014,5(16):11.
ZHAO Qianwei, MA Yuanyuan. A Preliminary Study on the Spread Characteristics of WeChat Rumors in the Media Era[J]. *Journal of News Research*, 2014, 5(16): 11.

[2] 王海鹰,闫浩文,田江鹏,等. 后现代哲学视野下的微地图[J]. 武汉大学学报(信息科学版),2022,47(12):2026-2037.
WANG Haiying, YAN Haowen, TIAN Jiangpeng, et al. WeMaps in the Perspective of the Post-Modernist Philosophy[J]. *Geomatics and Information Science of Wuhan University*,2022, 47(12): 2026-2037.

[3] 申金霞. 自媒体的信息传播特点探析[J]. 今传媒,2012,20(9):94-96.
SHEN Jinxia. Analysis on the Characteristics of Information Dissemination from the Media[J]. *Today's Massmedia*, 2012, 20(9): 94-96.

[4] 郭仁忠,陈业滨,赵志刚,等. ICT时代地图的科学概念及表达框架[J]. 武汉大学学报(信息科学版),2022,47(12):1978-1987.
GUO Renzhong, CHEN Yebin, ZHAO Zhigang, et al. Scientific Concept and Representation Framework of Maps in the ICT Era[J]. *Geomatics and Information Science of Wuhan University*, 2022, 47(12): 1978-1987.

[5] 王家耀,武芳,闫浩文. 大变化时代的地图学[J]. 测绘学报,2022,51(6):829-842.
WANG Jiayao, WU Fang, YAN Haowen. Cartography: Its Past, Present and Future[J]. *Acta Geodaetica et Cartographica Sinica*, 2022, 51(6): 829-842.

[6] 刘瑜,郭浩,李海峰,等. 从地理规律到地理空间人工智能[J]. 测绘学报,2022,51(6):1062-1069.
LIU Yu, GUO Hao, LI Haifeng, et al. A Note on GeoAI from the Perspective of Geographical Laws[J]. *Acta Geodaetica et Cartographica Sinica*, 2022, 51(6): 1062-1069.

[7] 应申,候思远,苏俊如,等. 论游戏地图的特点[J]. 武汉大学学报(信息科学版),2020,45(9):1334-1343.
YING Shen, HOU Siyuan, SU Junru, et al. Characteristics of the Game Map[J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(9): 1334-1343.

[8] 闫浩文,张黎明,杜萍,等. 自媒体时代的地图:微地图[J]. 测绘科学技术学报,2016,33(5):520-

523.
YAN Haowen, ZHANG Liming, DU Ping, et al. WeMap: A New Type of Map in the Era of We Media[J]. *Journal of Geomatics Science and Technology*, 2016, 33(5): 520-523.

[9] 李志林, 蓝天, 遆鹏, 等. 从马斯洛人生需求层次理论看地图学的进展[J]. 测绘学报, 2022, 51(7): 1536-1543.
LI Zhilin, LAN Tian, TI Peng, et al. Advances in Cartography from the Perspective of Maslow's Hierarchy of Needs[J]. *Acta Geodaetica et Cartographica Sinica*, 2022, 51(7): 1536-1543.

[10] ADOMAVICIUS G, TUZHILIN A. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2005, 17(6): 734-749.

[11] SU X Y, KHOSHGOFTAAR T M. A Survey of Collaborative Filtering Techniques[J]. *Advances in Artificial Intelligence*, 2009, 2009(1): 421425.

[12] VERBERT K, MANOUSELIS N, OCHOA X, et al. Context-Aware Recommender Systems for Learning: A Survey and Future Challenges [J]. *IEEE Transactions on Learning Technologies*, 2012, 5(4): 318-335.

[13] MOONEY R J, ROY L. Content-Based Book Recommending Using Learning for Text Categorization [C]//The 5th ACM Conference on Digital Libraries, San Antonio, USA, 2000.

[14] YE M, YIN P F, LEE W C. Location Recommendation for Location-Based Social Networks [C]// The 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, San Jose, USA, 2010.

[15] MA C, ZHANG Y X, WANG Q L, et al. Point-of-Interest Recommendation: Exploiting Self-attentive Autoencoders with Neighbor-Aware Influence [C]//The 27th ACM International Conference on Information and Knowledge Management, Torino, Italy, 2018.

[16] AAVIDSON J, LIEBALD B, LIU J N, et al. The YouTube Video Recommendation System [C]// The 4th ACM Conference on Recommender Systems, Barcelona, Spain, 2010.

[17] BORDES A, USUNIER N, GARCIA-DURÁN A, et al. Translating Embeddings for Modeling Multi-relational Data [C]//The 26th International Conference on Neural Information Processing Systems, New York, USA, 2013.

[18] WANG H, WANG N Y, YEUNG D Y. Collaborative Deep Learning for Recommender Systems[C]// The 21th ACM International Conference on Knowledge Discovery and Data Mining, Sydney, Australia, 2015.

[19] ZHANG F Z, YUAN N J, LIAN D F, et al. Collaborative Knowledge Base Embedding for Recommender Systems[C]//The 22nd ACM International Conference on Knowledge Discovery and Data Mining, San Francisco, USA, 2016.

[20] 何阳, 闫浩文, 王卓, 等. 面向微地图的地标提取方法及个性化寻路应用[J]. 地球信息科学学报, 2022, 24(5): 827-836.
HE Yang, YAN Haowen, WANG Zhuo, et al. Landmark Extraction Method and Personalized Wayfinding Application for WeMap [J]. *Journal of Geo-Information Science*, 2022, 24(5): 827-836.

[21] 牛雪磊, 杨军, 闫浩文. 融合知识图谱与协同过滤的微地图推荐[J]. 地球信息科学学报. 2024, 26(4): 967-977.
NIU Xuelei, YANG Jun, YAN Haowen. WeMap Recommendation by Fusion of Knowledge Graph and Collaborative Filtering[J]. *Journal of Geo-Information Science*, 2024, 26(4): 967-977.

[22] 代少升, 刘小兵, 赖智颖, 等. 网格化局部自适应DBSCAN聚类算法[J]. 重庆邮电大学学报(自然科学版), 2022, 34(2): 250-257.
DAI Shaosheng, LIU Xiaobing, LAI Zhiying, et al. Gridded Local Adaptive DBSCAN Clustering Algorithm [J]. *Journal of Chongqing University of Posts and Telecommunications* (*Natural Science Edition*), 2022, 34(2): 250-257.

[23] XIONG X, QIAO S J, HAN N, et al. Affective Impression: Sentiment-Awareness POI Suggestion via Embedding in Heterogeneous LBSNA [J]. *IEEE Transactions on Affective Computing*, 2022, 13(1): 272-284.

[24] FISCHER A, IGEL C. Training Restricted Boltzmann Machines: An Introduction[J]. *Pattern Recognition*, 2014, 47(1): 25-39.

[25] HIOTON G E. A Practical Guide to Training Restricted Boltzmann Machines[J]. *Lecture Notes in Computer Science*, 2012, 7700: 599-619.

[26] HIOTON G E, OSINDERO S, THE Y W. A Fast Learning Algorithm for Deep Belief Nets[J]. *Neural Computation*, 2006, 18(7): 1527-1554.

[27] 张淑华, 李海莹, 刘芳. 身份认同研究综述[J]. 心理研究, 2012, 5(1): 21-27.
ZHANG Shuhua, LI Haiying, LIU Fang. Review on Research of Identity [J]. *Psychological Research*, 2012, 5(1): 21-27.