

基于图卷积神经网络的家具推荐方法

黄雨洁¹, 李 昕¹, 胡基柔¹, 陶 卓²

(1.南京邮电大学计算机学院, 江苏 南京 210023;

2.南京邮电大学管理学院, 江苏 南京 210023)

✉ 847993517@qq.com; 1375758190@qq.com; b20031306@njupt.edu.cn; taozhuo@njupt.edu.cn



摘要:协同过滤是大数据推荐系统的重要算法之一,为提高家具推荐效果,文章首先介绍了协同过滤的经典算法,其次对基于图卷积的推荐算法进行梳理并改进,基于收集到的数据集构建了模型,最后选取其他协同过滤经典算法与之进行对比和分析。实验结果表明,相较于基于物品的协同过滤算法模型、基于用户的协同过滤算法模型,基于图卷积的协同过滤算法模型在评价指标上有更好的表现,更适合应用于家具推荐。

关键词:图卷积神经网络;推荐算法;家具推荐

中图分类号:TP391 **文献标志码:**A

A Furniture Recommendation Method Based on Graph Convolutional Neural Networks

HUANG Yujie¹, LI Xin¹, HU Jirou¹, TAO Zhuo²

(1.School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China;

2.School of Management, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

✉ 847993517@qq.com; 1375758190@qq.com; b20031306@njupt.edu.cn; taozhuo@njupt.edu.cn

Abstract: Collaborative filtering is one of the most important algorithms of big data recommendation system. In order to improve the effectiveness of furniture recommendations, this paper proposes to improve recommendation algorithm based on graph convolution. Firstly, classic algorithms of collaborative filtering are introduced. Then recommendation algorithms based on graph convolution are sorted out and improved, and a model based on the collected data set is built. Finally, other classic algorithms of collaborative filtering are selected for comparison and analysis. Experimental results show that compared with ItemCF and UserCF models, the collaborative filtering algorithm model based on graph convolution has better performance in evaluation indicators and is more suitable for furniture recommendation.

Key words: graph convolution neural network; recommendation algorithm; furniture recommendation

0 引言(Introduction)

随着人们消费水平的提升,对个性化家具的购买需求逐渐增大,而通常城市中家具城的地理位置相对偏远、退换货流程不便,加之人们的生活节奏加快,选购家具比较困难。针对以上问题,家具行业智能化推荐方法显得尤为重要。2019年由中华人民共和国工业和信息化部等十三个部门印发的《制造业设计能

力提升专项行动计划(2019—2022年)》也指出要实现传统优势产业设计升级,这对家具行业的转型升级提出了更高的要求^[1]。

推荐系统在实现家具行业智能化的过程中起到了重要的作用,并且近年来各类推荐算法也不断进步^[2]。有研究表明,卷积神经网络相比于传统的推荐算法,可以通过邻域聚合操作,实现对节点邻居信息的更精准统计^[3]。本文将基于图卷积神经网络的协同过滤算法应用于家具推荐,根据用户对部分家

具的评分以及其性别、年龄和薪资等特征属性和家具特征属性,进行二层图卷积神经网络的搭建,实现对家具评分的预测,帮助用户高效选购家具。

1 背景介绍(Background information)

1.1 家具行业发展现状及趋势

经济的持续快速发展与现代化建设的不断推进以及居民收入水平的不断提高,为中国家具行业提供了良好的发展环境。进入经济全球化、贸易全球化与互联网时代,人们对思想与审美的个性化追求与商品性价比的要求逐步提升,加之移动互联网的发展、各类3D模型技术的精进和推荐算法的不断优化,家具装饰移动应用应运而生,具有巨大的市场潜力。

家具行业属于创意性传统企业。此前,我国家具行业中企业数量较多,但行业集中度较低。随着居民生活节奏的加快,实体店的客流量逐渐减少。为适应消费者的消费需求,家具定制逐渐与互联网融合,电子商务快速发展,O2O(Online To Offline)模式的普及度逐渐提升,电商利用互联网聚集了大量的家具供应商,并运用大数据实现精细化管理,通过推送算法根据用户喜好为其推送关注的内容。此外,O2O模式得以与推送算法结合,给予消费者更多交互体验,能让消费者在虚拟的世界中获得更多的信息^[4]。家具装饰移动应用通过高性价比的功能与精细管理吸引了大量消费者。

1.2 协同过滤推荐系统发展

推荐系统随着互联网的发展而快速进步,而协同过滤作为推荐系统中的经典算法,在推荐领域中起着重要作用。1994年,GroupLens项目组在明尼苏达大学首次开发出一种全新的自动化协同过滤系统,它不仅首次将协同过滤的理论应用于实践,更进一步将推荐问题的系统模型纳入其中,这一创新性的技术将在未来几十年里极大地促进推荐系统的发展^[5]。该小组随后创建了一个试点研究网站MovieLens,其中包含迄今为止在推荐领域被广泛引用的数据集,影响深远。

协同过滤算法细分可以分为以用户为基础(User-based)的协同过滤算法、以项目为基础(Item-based)的协同过滤算法、以模型为基础(Model-based)的协同过滤算法。以模型为基础的协同过滤算法是前两种算法的进一步发展,也是目前应用最为广泛的协同过滤算法^[6]。

2016年,视频网站YouTube尝试通过结合深度学习的协同过滤算法满足视频推荐需求的增长,通过实验证明,推荐系统和深度学习的结合可以取得良好效果。自此,推荐系统与深度学习结合研究成为热点^[7]。近年来,将图神经网络(Graph Neural Network, GNN)应用至协同过滤算法中也是一个研究热点,其中包括图卷积神经网络(Graph Convolution Network, GCN)^[8]。研究表明:GCN相比传统的方法,能够更高效地利用推荐系统以及数据中存在的用户属性和商品属性信息构建图的结构属性和节点特征信息^[9-10]。因此,GCN在推荐系统中的应用有着良好的发展前景。

2 图卷积神经网络推荐算法介绍(Introduction to GCN recommendation algorithms)

在推荐系统中有两组集合分别是用户集 $U=\{u_1, u_2, u_3, \dots,$

$u_n\}$ 和家具集 $F=\{f_1, f_2, f_3, \dots, f_m\}$ 。用 $R=\{r_{uf}\}$ 表示用户 u 对家具 f 的评分。用以上数据集构建用户-家具的二部图 $G=(V, E)$ 。其中, $V=U\cup F$,即该二部图中的所有节点由用户节点和家具节点构成。 $E=R$ 表示用户对家具的评分值构成二部图中的边。由此可知,用户对家具的评分越高,则关联越强。构造的用户-家具二部图 G 如图1所示。 G 的邻接矩阵记为 A ,其中 A_{ij} 代表 i 节点和 j 节点是否连接。 G 的度矩阵记为 D , D_{ii} 表示节点 i 的度。GCN用来传播用户和家具的特征,其计算公式如下:

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

其中: $\mathbf{H}^{(l+1)}$ 代表节点在第 $l+1$ 层特征的表示; σ 为非线性激活函数ReLU; $\tilde{\mathbf{A}}=\mathbf{A}+\mathbf{I}$ (\mathbf{I} 为单位矩阵),单位矩阵是为了保留节点的固有特征,两边乘 $\tilde{\mathbf{D}}^{-\frac{1}{2}}$ 是为了实现对特征的归一化处理; $\mathbf{W}^{(l)}$ 为第 l 层特征的权重矩阵^[11]。

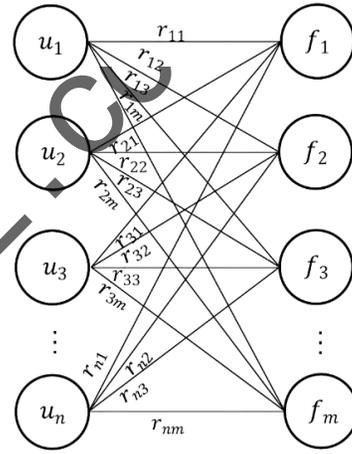


图1 用户-家具二部图

Fig. 1 User-furniture bipartite graph

3 图卷积神经网络推荐算法框架介绍(Introduction to GCN recommendation algorithm framework)

基于GCN的家具推荐方法的基本框架可分为输入层、卷积层及输出层。输入层主要负责生成相关图结构数据。卷积层传播用户和家具的特征,并不断优化用户和家具的特征表示。输出层将不同卷积层得到的用户和家具的特征表示聚合,预测用户对家具的评分。

输入层用于生成相关图结构数据以及初始化用户与家具的特征表示。图结构数据主要包括节点自身的特征向量以及图的邻接矩阵和度矩阵等。本层将每一条用户与家具信息转换为向量表示。在本文实验中,向量的维度设置为64,节点的特征向量可表示为[用户id/家具id,性别/价格, ...]。邻接矩阵体现了各个节点之间的关系,为卷积层中传播用户和家具特征奠定了基础。

卷积层主要完成传播特征及不断完善用户及家具的特征表示的任务。文献^[12]指出一旦使用多层卷积层,相关任务的传播效果会急剧下降,所以本实验采用两层卷积层,具体的计

算公式如公式(2)所示:

$$H^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l+1)} W^{(0)}) W^{(1)}) \quad (2)$$

其中: σ 为非线性激活函数 ReLU, $W^{(0)}$ 为第一层 GCN 的特征权重矩阵, $W^{(1)}$ 为第二层 GCN 的特征权重矩阵。通过堆叠多个图卷积网络层,模型可以输出各个节点更高维度的特征表示,利用这些特征表示可以实现后续节点的预测。

输出层的主要任务是利用卷积层得出各个节点特征预测用户对家具的偏好得分。在多层卷积后,可以得到多个用户 u 和家具 f 的特征表示,即 $H_u^{(1)}, H_u^{(2)}, H_u^{(3)}, \dots, H_u^{(l)}$ 和 $H_f^{(1)}, H_f^{(2)}, H_f^{(3)}, \dots, H_f^{(l)}$ 。不同卷积层得到的用户、家具特征都可以反映用户对各种家具偏好的不同,形成预测公式如式(3)所示:

$$R_{uf} = \sum_{l=0}^L W_l H_u^{(l)} H_f^{(l)} \quad (3)$$

其中: R_{uf} 表示用户 u 对家具 f 的偏好预测分数, W_l 为第 l 层时的参数, $H_u^{(l)}$ 为第 l 层时的用户特征表示, $H_f^{(l)}$ 为第 l 层时的家具特征表示。

基于 GCN 的家居推荐方法框架如图 2 所示。

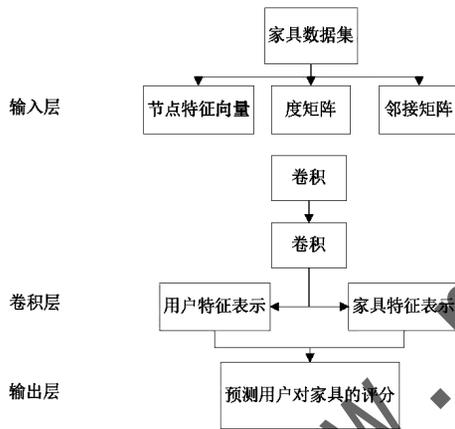


图 2 基于 GCN 的家具推荐方法框架

Fig. 2 A framework for furniture recommendation methods based on GCN

4 实验分析(Experimental analysis)

4.1 实验环境

本文构建的 GCN 模型基于深度学习框架 PyTorch 实现,CPU 处理器采用 Intel(R) Core(TM)i7-10710U,GPU 处理器采用 NVIDIA GeForce MX350。

4.2 数据集

为验证本文所提出的基于 GCN 的家具推荐方法的准确性,本实验共收集了 910 名用户对 50 种家具的 22 750 条评分记录,其中每个用户至少评价 25 种家具,评分的取值位于整数 1~5,评分越高,意味着用户越喜爱这种家具。本实验随机选择评分记录的 70% 作为训练集,剩余的 30% 作为测试集。

4.3 数据预处理

在实验中,为 910 名用户添加了性别、年龄和薪资等特征

属性,最终形成 910×4 的用户特征矩阵。用户特征属性如表 1 所示。

表 1 用户特征属性
Tab.1 User profile attributes

用户	性别	年龄	薪资
1	男性	25 岁以下	3 000 元以下
2	女性	26~35 岁	3 000~7 000 元
3	女性	36~45 岁	7 000~10 000 元
4	女性	45 岁以上	10 000 元以上

本文实验为 50 种家具添加了价格、材质、尺寸及颜色 4 种特征属性,最终形成 50×5 的家具特征矩阵。同时,利用收集到的用户对家具的评分可以构建用户-家具三元组 $[user_id, item_id, rating]$,可以用来描述用户节点与家具节点之间的关系,通过这个三元组可以构建邻接矩阵。

4.4 实验结果评估标准

本文实验算法的评估标准主要是三个方式:精确度、召回率和均方根误差值。

精确度(Precision)以预测的结果作为判断依据,表示为推荐结果中与真实数据相同的数据与推荐数据的比值,代表推荐数据中的正确率,如公式(4)所示:

$$Precision = \frac{\sum_u |R(u) \cap T(u)|}{\sum_u |R(u)|} \quad (4)$$

公式(4)中,对于用户 u 的推荐家具的集合为 $R(u)$,而用户 u 喜欢的家具集合为 $T(u)$ 。

召回率(Recall)表示预测结果与真实数据相同的正确数据与真实数据的比值,代表真实数据的推荐率,如公式(5)所示:

$$Recall = \frac{\sum_u |R(u) \cap T(u)|}{\sum_u |T(u)|} \quad (5)$$

均方根误差(RMSE)是预测值和真实值偏差的平方和与观测次数的比值的平方根,衡量的是预测数据与真实数据的偏差,对于异常数据较为敏感,如公式(6)所示:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - f(x_i))^2} \quad (6)$$

其中: N 表示观测次数, $Y_i - f(x_i)$ 表示观测值与真实值的偏差。

4.5 基线方法

为证明本文提出的基于 GCN 的家具推荐算法的准确性,选用 ItemCF 和 UserCF 与之进行比较^[13]。

ItemCF:通过用户行为计算物品相似度,根据物品之间的相似度与用户历史行为形成用户推荐表,适用于用户个性化推荐。

UserCF:根据用户之间相似度,给目标用户推荐相似用户产生过行为的物品。UserCF 的推荐更社会化,适用于新闻推荐。

4.6 实验结果分析

本文实验选取精确度、召回率与均方根误差作为推荐算法的性能衡量指标,精确度、召回率的数值越高,均方根误差的数值越低,代表推荐算法的性能越好。为研究不同模型对协同过滤算法性能的影响,本实验利用家具喜爱度问卷搜集到的数据集,分别对 GCN、ItemCF、UserCF 三种算法进行算法测试。在三组实验数据中,三种算法得到的精确度、召回率和均方根误差如表 2 所示,评估数据视图如图 3 所示。

表 2 协同过滤算法评估数据对比

Tab.2 Comparison of evaluation data for collaborative filtering algorithms

模型	精确度	召回率	均方根误差
GCN	0.817 2	0.815 9	0.259 9
ItemCF	0.862 5	0.888 4	1.807 4
UserCF	0.754 3	0.807 0	2.319 4

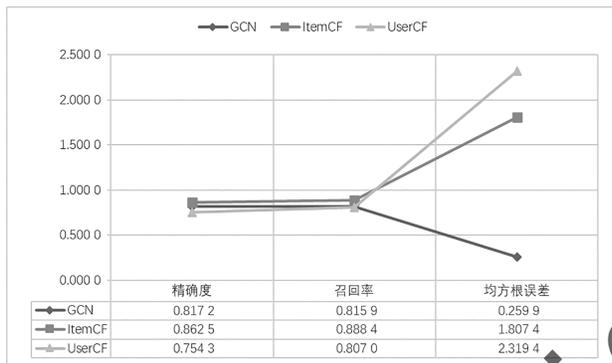


图 3 评估数据视图

Fig. 3 Assessment data view

根据图 3 得知,在精确度和召回率方面,ItemCF 的算法得分要高出其他两种算法,但在均方根误差指标方面,GCN 要明显低于其他两种算法的指标,可以达到约 0.26,ItemCF 和 UserCF 则都大于 1.8。对比结果可知,基于 GCN 的推荐算法相比 ItemCF 和 UserCF,在均方根误差上实现了更好的性能,而 ItemCF 在精确度与召回率上实现了更好的性能。本文提出的 GCN 算法优化是可取的,精确度与召回率相差较小的情况下,均方根误差越小,说明算法越好,拟合程度越高,即对于用户的推荐效果就越好。在用户或者物品较多的场合,ItemCF 与 UserCF 的物品相似度矩阵的计算代价较大,损失也较大,对比 ItemCF 与 UserCF,GCN 在 Loss 训练上的表现更好。综上所述,GCN 在数据量较多时,根据同时聚合的用户与家具的特征,捕捉全图信息,模型结果更加准确,预测值更加接近真实值,对于用户的推荐效果更好。

5 结论(Conclusion)

针对当下人们对个性化家具的购买需求的增大,本文提出了一种基于图卷积神经网络的家具推荐方法。该推荐方法通过建立图卷积神经网络获取不同层的用户-家具特征,将多层特征表示聚合后进行评分预测。开展实验与其他常见协同过

滤算法进行比较,结果表明该推荐方法在精确度、召回率和 Loss 训练上都有较好的效果。下一步将结合家具商品的实际市场,进一步优化模型,提升推荐效果。

参考文献(References)

- [1] 中华人民共和国工业和信息化部,中华人民共和国国家发展和改革委员会,中华人民共和国教育部,等. 制造业设计能力提升专项行动计划(2019—2022 年)[EB/OL]. (2019-11-13)[2023-03-15]. https://www.miit.gov.cn/jgsj/zfs/wjfb/art/2020/art_e0dfb53144d24784a416c8d03424d221.html.
- [2] 黄勃,严非凡,张昊,等. 推荐系统研究进展与应用[J]. 武汉大学学报(理学版),2021,67(6):503-516.
- [3] 刘雨昕,陈春雨. 基于图卷积神经网络的推荐算法[J]. 应用科技,2022,49(4):13-17.
- [4] 章彰,许柏鸣. 基于 O2O 电子商务模式的家具企业发展现状及趋势分析[J]. 家具与室内装饰,2012(11):64-65.
- [5] 高阳团. 推荐系统开发实战[M]. 北京:电子工业出版社,2019:2-3.
- [6] 赵俊逸,庄福振,教翔,等. 协同过滤推荐系统综述[J]. 信息安全学报,2021,6(5):17-34.
- [7] COVINGTON P, ADAMS J, SARGIN E. Deep neural networks for YouTube recommendations[C]//Association for Computing Machinery. Proceedings of the 10th ACM Conference on Recommender Systems. New York: ACM, 2016:191-198.
- [8] 吴国栋,查志康,涂立静,等. 图神经网络推荐研究进展[J]. 智能系统学报,2020,15(1):14-24.
- [9] 葛尧,陈松灿. 面向推荐系统的图卷积网络[J]. 软件学报,2020,31(4):1101-1112.
- [10] 魏文超,简广逢,廖开阳,等. 图网络层级信息挖掘分类算法综述[J]. 中国图象图形学报,2022,27(10):2916-2936.
- [11] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[DB/OL]. (2020-12-28)[2023-03-15]. <https://arxiv.org/abs/1609.02907>.
- [12] LI Q M, HAN Z C, WU X M. Deeper insights into graph convolutional networks for semi-supervised learning[J]. Proceedings of the AAAI Conference on Artificial Intelligence,2018,32(1):1-9.
- [13] 秦冲,赵铁柱,柳毅. 个性化推荐算法的研究及发展综述[J]. 东莞理工学院学报,2021,28(3):51-60.

作者简介:

黄雨洁(2002-),女,本科生。研究领域:信息安全。
 李 昕(2001-),女,本科生。研究领域:信息安全。
 胡基柔(2001-),女,本科生。研究领域:信息安全。
 陶 卓(1983-),女,博士,讲师。研究领域:软件产业政策研究。