

DOI:10.19651/j.cnki.emt.2108462

基于时序特征的图卷积协同过滤推荐算法*

陈一凡 朱民耀 朱晓强 宋海洋 陆小锋
(上海大学通信与信息工程学院 上海 200444)

摘要: 基于图卷积神经网络的协同过滤推荐算法框架是目前最先进的推荐算法框架,该框架在用户-项目交互嵌入向量的特征学习中并未关注交互发生的时序性,但实际情况中,用户-项目交互普遍具有明显的时序特征,且是影响推荐性能的重要因素。因此,本文提出一种基于时序特征的图卷积协同过滤推荐算法,重做多个数据集,保留数据集时序特征等原始信息,总结归纳数据集中用户-项目交互的历史时序信息,并对其进行参数化处理,作为重要特征输入到图卷积神经网络模型训练的高阶协同信号传递中。在3个公开的官方数据集 Gowalla, Yelp 和 Amazon-book 上设置多组消融性实验,采用公认的评价指标 ndcg 和 recall 对推荐算法性能进行评价,实验结果证明,在同样参数设置下,基于时序特征的图卷积协同过滤推荐算法性能超越了现有同类型图卷积协同过滤推荐算法,验证了时序特征对提升推荐效果的积极作用,提高了模型训练效率和预测命中率,更加高效地解决网络信息过载问题,满足了更高的应用需求。

关键词: 推荐算法;图卷积神经网络;协同过滤;时序特征

中图分类号: TP301 **文献标识码:** A **国家标准学科分类代码:** 520

Graph convolution collaborative filtering recommendation algorithm based on the time series features

Chen Yifan Zhu Minyao Zhu Xiaoqiang Song Haiyang Lu Xiaofeng

(School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China)

Abstract: The collaborative filtering recommendation algorithm framework based on graph convolutional neural network is the most advanced recommendation algorithm framework at present. The framework does not pay attention to the timing of interaction occurrence in the feature learning of user-item interaction embedding vector, but in actual situations, users-Item interaction generally has obvious timing characteristics and is an important factor affecting recommendation performance. Based on this, a graph convolution collaborative filtering recommendation algorithm based on time series features is proposed, which redo multiple data sets, retain the original information of the data sets, especially the time series features, and summarize the historical time series information of user-item interaction in the data set. It is parameterized and put as an important feature input to the high-order cooperative signal transmission of graph convolutional network model training. Set up multiple sets of ablative experiments on three publicly available official datasets-Gowalla, Yelp and Amazon-book, and use recognized evaluation indicators-ndcg and recall to evaluate the performance of the recommendation algorithm. The experimental results show that under the same parameter settings, the figure convolution collaborative filtering recommendation algorithm based on temporal characteristics performance beyond the existing same type figure convolution, collaborative filtering recommendation algorithm to verify the timing characteristics are recommended to improve effect of the positive role, improve the efficiency of model training and prediction, shooting more efficiently solve the problem of network information overload, to satisfy the higher application requirements.

Keywords: recommendation algorithm; graph convolutional neural network; collaborative filtering; the time series features

0 引言

推荐系统利用电子商务网站向客户提供商品信息和建

议,帮助用户决定应购买什么产品,进而达到模拟销售人员帮助客户完成购买过程的目的^[1]。个性化推荐则根据用户的兴趣特点和购买行为,向用户推荐其感兴趣的信息和商

收稿日期:2021-11-28

* 基金项目:上海市科委科技创新行动计划项目(21511102605)资助

品。随着电子商务规模的不断扩大,商品个数和种类快速增长,顾客需要花费大量时间才能找到目标商品。浏览大量无关信息和产品的过程,无疑会使淹没在信息过载问题中的消费者不断流失。为解决这些问题,个性化推荐系统应运而生^[2]。推荐系统的核心任务是对推荐算法性能提升的研究,主流的推荐系统主要分为传统的 3 类^[3]:基于内容过滤的推荐^[4]、协同过滤的推荐^[5]和混合推荐;分为基于深度学习的 4 类:基于深度神经网络的推荐、基于循环神经网络的推荐^[6-9]、基于卷积神经网络的推荐^[10-12]和基于图神经网络^[13]的推荐,其中图神经网络的推荐算法出现较晚,概念也相对较新,融合了循环神经网络和卷积神经网络的思想,重定义和设计了一种神经网络算法,用以处理非欧式空间的数据,该算法可以对节点之间链接不固定的数据图谱进行高效的建模并精确捕捉数据潜在联系,从而进行有效推荐^[14]。通过对图神经网络算法的研究发现,目前在公认指标评价体系下,性能表现最好的图神经网络算法是 Wang 等^[15]提出的 NGCF(neural graph collaborative filtering)模型以及 He 等^[16]改进后的 LightGCN(light graph convolution network)模型,基于图神经网络的协同过滤算法核心思想是利用 GCN(graph neural network)的消息传递机制堆叠多个图卷积层,实现二分图上高阶邻居的协同信号隐式传递,最后合并所有层的节点嵌入进行推荐,两者的不同点是后者取消了特征变换和非线性激活函数,并通过实验验证了方法的合理性和有效性。但该类算法不管是从图结构的建立还是协同信号的隐式传递上,都唯一考虑了不同用户和不同物品之间是否发生交互这一属性,也就是说通过高阶协同信号隐式传递的最终结果找到的相似用户,仅仅是通过用户与物品是否交互判定的,而用户与物品何时交互,交互次数并未作为信息参与信号传递,为此,本研究首次将用户与物品何时交互和交互次数总结为交互的时序属性^[17],并通过归纳将该属性参数化,作为重要信息加入到信号的传递中,从而在结果上表现为寻找到更加相似的用户,有效地提高推荐的命中率。

本研究的主要贡献:1)从 Gowalla, Yelp 和 Amazon-book 三个公认公开的原始数据集中,提取包含用户和项目交互时序信息的实验数据集,对数据集进行整合分割,作为算法模型训练的训练集以及模型验证的测试集;2)归纳总结用户和项目交互时序信息(包括交互发生的时间和频次)对推荐的贡献,并将其参数化;3)在高阶协同信号传递中,参数化形式引入用户和项目交互的时序信息,在通用硬件平台上,同一实验设置下,采用多种数据集,实验验证得出本研究提出的方法均对比现有方法提高了模型预测效果,提供了一种新型的推荐算法,更加高效地解决网络信息过载问题。

1 相关工作

1.1 图卷积协同过滤算法思想

图卷积协同过滤算法是借鉴融合了协同过滤和图卷积神经网络的算法思想。协同过滤推荐区别于传统的基于内

容的推荐,协同过滤简单来说是利用某兴趣相投、拥有共同经验之群体的喜好来推荐用户感兴趣的信息,个人通过合作的机制给予信息相当程度的回应(如评分)并记录下来以达到过滤的目的进而帮助别人筛选信息,回应不一定局限于特别感兴趣的,特别不感兴趣信息的记录也相当重要。基于协同过滤的推荐能够过滤机器难以自动内容分析的信息,如艺术品、音乐等;共用其他人的经验,避免了内容分析的不完全或不精确,并且能够基于一些复杂的,难以表述的概念(如信息质量、个人品味)进行过滤;有推荐新信息的能力,可以发现内容上完全不相似的信息,用户对推荐信息的内容事先是预料不到的;可以发现用户潜在的但自己尚未发现的兴趣偏好;推荐个性化、自动化程度高、能够有效的利用其他相似用户的回馈信息、加快个性化学习的速度。比较常见的基于协同过滤的算法思想是使用矩阵分解,通过对用户和项目的评分矩阵进行分解以获取用户和项目的嵌入表达,从而相乘得到任意用户和任意项目的交互预测评分,但问题在于,矩阵分解只能运算用户和项目之间的线性交互,对非线性交互却无能为力,针对这一问题,NGCF 以及优化后的 LightGCN 模型引入图卷积神经网络算法,学习协同过滤的非线性特征,并表现出良好的性能。

1.2 NGCF 简介

NGCF 模型框架包含 3 个模块^[15]。第 1 个模块是提供用户嵌入和项目嵌入的嵌入层和初始化;第 2 个模块是信息嵌入传播层,又分为信息构成和信息聚集;第 3 个模块是预测层,该层聚合来自不同传播层的精细化嵌入并输出用户和项目交互对的预测评分。整体上来看,NGCF 模型的性能目前来说均比当时其他各类模型的性能表型得更加优异。

1.3 LightGCN 简介

鉴于 NGCF 使用非线性激活函数和特征转换矩阵。然而,实验论证这两种操作对于协同过滤并不那么有用^[16]。在协同过滤中,交互图的每个节点只有一个 ID 作为输入,没有具体的语义。在这种情况下,执行多个非线性转换将无助于更好地学习特征;更糟糕的是,这可能会增加训练的难度。在一系列消融性实验中得出结论,添加特征变换会对模型产生负面影响,如果将其去除可以显著提高模型性能;加入非线性激活对包含特征变换模型的影响较小,但禁用特征变换模型的影响较大;从整体上看,特征变换和非线性激活对 NGCF 的影响是负面的,同时去除特征变换和非线性激活后,模型整体性能有所改进。

2 基于时序特征的图卷积协同过滤推荐算法

2.1 传统图卷积协同过滤算法问题分析

图卷积协同过滤的核心之一是利用用户项目的高阶连接来编码交互图结构的协作信号,用于解决交互规模巨大难以提取协作信号的挑战。如图 1 所示,解释说明了高阶连通性的概念,感兴趣的推荐用户为 u_i ,标记为与用户和

项目交互的左子图中的双圈图。右边的子图显示了展开的树结构,从 u_1 开始所到达的路径长度大于 1 的任何节点。不难发现 u_1 和 u_2 都与 i_1 发生了交互,在一阶邻居上认为二者是相似的,而 u_2 又与 i_4 和 i_5 发生了交互,故可以把 i_4 和 i_5 作为对 u_1 的推荐项目,同时发现 i_4 和 i_5 到 u_1 的连通路径均有两条,如果按照 NGCF 作者文中提出比较连通路径数量的方案,无法区分出 i_4 和 i_5 推荐给 u_1 的优先级,在数据量和交互量巨大的情况下,这种现象的出现是很常见的,无法有效区分出推荐项目的优先级,体现在嵌入聚合的评分结果上就会产生评分模糊和推荐偏差。

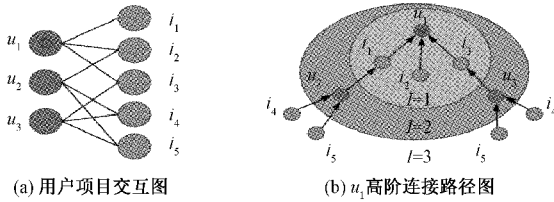


图1 用户项目交互和高阶连通性的示意图

针对这一问题,本研究提出解决方案,对路径进行加权处理,数据集中用户和项目的直接交互在路径图上体现为邻居关系,而任意项目到目标用户的路径体现为特定个相接邻居的路径和,在单个邻居的路径权值均等的情况下,会出现某个项目到目标用户有多条相同路径的情况,但如果对邻居路径进行加权处理,那么这种情况就会大大减少甚至消除,从而更加清晰地区分出推荐项目的优先级。本研究基于用户历史数据时序信息,归纳用户与直接交互项目的亲和度并对其参数化(详见本章 2.3 节),作为路径上任意相邻用户和项目的邻居路径权值。

2.2 数据集

本研究实验开展的数据集基于公开公认的 Gowalla, Yelp 和 Amazon-book 原始数据集,在 NGCF 和 LightGCN 模型中,提出者使用的数据集虽然也是基于这 3 种数据集,但他们对数据集的处理仅仅关注了用户和项目交互的这一属性,对于用户与项目交互的时序特征并未关注。本研究首先获取原始数据集,以用户和项目发生交互的时间为索引,重做了数据集,形式上保留了用户和项目的交互,内容上保留了交互发生的时序信息,包括交互的先后和频次,数据集基本属性如表 1 所示。

表 1 数据集基本属性

数据集	用户数量	项目数量	交互数量	交互密度
Amazon-book	9 087	100 000	670 311	0.000 74
Yelp	6 940	132 694	499 881	0.000 54
Gowalla	16 401	99 999	1 124 354	0.000 69

2.3 时序特征参数化

数据集中,任意用户交互过的项目集合,在经过处理后按时间排列,并且保留频次信息。每一个用户对应一个项

目集合,该集合中的每一个项目都是对应用户的一阶邻居,故只需求解该种情况下的一阶邻居路径权值,就可以在构建高阶连通图时得到目标用户和任意项目的路径和。本研究将用户和其一阶邻居项目的路径权值形象地表述为亲和度,根据一阶邻居与用户发生交互的时序特征求解亲和度,这种项目亲和度高低表现为用户再次选择该项目的倾向性大小,项目亲和度越高用户再次选择该项目的倾向性就越大,反之亦然。

一般来说,同时在后续章节的实验结果也证明了,用户对项目的亲和度,与用户对项目历史选择的次数,最后选择的时间和最后两次选择的间隔时间 3 个因素来决定。

$$\begin{cases} N \propto K \\ N \propto S \\ N \propto S/T \end{cases} \quad (1)$$

将选择的次数定义为频次 K ,最后选择的时间定义为 S ,最后两次选择的时间间隔定义为 T ,用户对项目的亲和度与 K 成正比,代表用户选择该项目的次数越多,亲和度越高;与 S 占据整个用户在线时段的比值成正比,代表用户越是近期选择该项目,亲和度越高;与 T 和最后选择时间 S 的比值成反比,代表用户最后对该项目的选择密度越大,亲和度越高。定义亲和度为 N ,其与各影响因素的关系如式(1)所示,定义用户完整在线时段为 L ,同时对频次和最后选择时间进行归一化处理,得到亲和度的最终表达式如式(2)所示。

$$N = \alpha \cdot \frac{K}{L} \cdot \frac{S}{L} \cdot \frac{S}{T} = \frac{\alpha K S^2}{T L^2} \quad (2)$$

2.4 引入时序特征的模型训练

本研究沿用了传统图卷积协同过滤网络模型对用户项目对的评分预测机制,通过改变卷积层数调节拟合高阶邻居相似性的阶数,通过多次损失函数的反向传播迭代更新用户和项目嵌入向量矩阵的权值,用户项目对的最终预测评分可以由最后一层输出的用户和项目嵌入向量的向量积表示,如式(3)所示。

$$y(u, i) = e_u^T e_i \quad (3)$$

其中, e_u 表示用户的最终嵌入表达, e_i 表示项目的最终嵌入表达,二者的内积表示用户对目标项目的预测评分 $y(u, i)$,通过排序特定用户对所有项目的预测评分,可以得到用户最感兴趣的前 n 个项目,计算预测得出的 n 个项目和实际测试集的 n 个项目的匹配程度可以评价模型的预测效果。因此提高模型预测效果的关键,就是提高用户和项目最终嵌入表达的准确度,在传统图卷积协同过滤模型中,参数通过多个卷积层的迭代更新,达到高阶协同信号的传播效果,参数的更新方式采用优化的成对 BPR 损失,该方式在推荐系统中得到了广泛的应用,具体地说,BPR 假设所观察到的交互更能反映用户的偏好,应赋予比未观测到的交互更高的预测值,目标函数为式(4):

$$Loss = \sum_{(u,i,j) \in O} -\Delta f(y_{ui} - y_{uj}) + \lambda_{(u,i,j)} \quad (4)$$

其中, (u, i, j) 代表数据集中的用户和项目, (u, i) 表示已知的用户和项目交互对, (u, j) 表示未知的用户和项目交互对, y_{ui} 和 y_{uj} 表示上述用户项目对的预测评分, 已知用户项目交互对预测值应该大于未知项目交互对预测值, 通过这种方式设计目标函数, 随机抽样一定批量的三元组 (u, i, j) 带入目标函数更新模型参数, $\lambda_{(u,i,j)}$ 控制 L_2 正则化强度以防止过拟合。但很明显, 使用该目标函数更新参数, 模型的优化指标仅仅与用户项目对是否存在已被观测到的交互有关, 结合 2.1 的讨论和 2.3 的准备工作, 本研究将目标函数进行了改进, 如式(5)所示。

$$Loss = \sum_{(u,i,j) \in O} -\Delta f(y_{ui} - y_{uj}) - \frac{\Delta f(y_{ui_{pos}} - y_{ui_{neg}})}{n} + \lambda_{(ui,uj,ui_{pos},ui_{neg})} \quad (5)$$

其中, 新增一项成对 BPR 损失 $\Delta f(y_{ui_{pos}} - y_{ui_{neg}})/n$, 用以更新模型参数。已知用户项目交互对和未知用户项目交互对是随机抽样的, 其中已知用户项目交互对, 就是用户和其一阶邻居项目交互对, 将已知用户项目交互对按照用户对项目的时序特征参数化的亲和度不同来非随机抽样, 令亲和度 N_{ui} 越高抽样概率 P_{ui} 越大, 亲和度 N_{ui} 越小抽样概率 P_{ui} 越小, 如式(6)所示。

$$P_{ui} = \frac{N_{ui}}{\sum_{(u,i)} N_{ui}} \quad (6)$$

定义 $y_{ui_{pos}}$ 表示大概率抽样到的已知用户项目交互对预测评分, $y_{ui_{neg}}$ 表示小概率抽样到的已知用户项目交互对预测评分, 前者预测评分应该大于后者预测评分, 据此规律从而实现模型训练过程中引入时序特征调节参数更新的目的。模型中, 则:

$$\Delta f(x) = -\ln(1 + e^{-x}) \quad (7)$$

由于抽样批量用户和项目, 但时序特征只体现在单个用户的一阶邻居项目子集中, 研究将反应时序特征的损失值乘上一个批量系数 $1/n$, 其中 n 代表抽样批量, 为了防止过拟合, 完善控制 L_2 正则化强度的函数为 $\lambda_{(ui,tj,ui_{pos},ui_{neg})}$, 将非等概抽样的用户项目对参数集合更新到防过拟合操作中。

3 实验结果与分析

3.1 基于时序特征的图卷积协同过滤模型性能对比

为了验证改进后的算法的可行性与有效性, 本研究进行了一系列消融性实验。主要与传统公认的效果很好的 NGCF 和 LightGCN 做了横向实验对比, 分别在 NGCF 和 LightGCN 模型上引入时序特征对算法做优化, 将 NGCF 模型的时序特征优化模型命名为 TNGCF (time series into neural graph collaborative filtering), 将 LightGCN 模型的时序特征优化模型命名为 TLGCN (time series into light graph convolution network)。实验硬件环境为配有显卡的 Windows 主机, 处理器参数为 Intel(R) Core(TM) i7-7700 CPU@3.60 GHz 3.60 GHz, 机带 RAM 为 16 GB, 显卡型号为 GeForce GTX 1080Ti, 实验软件环境为 Windows10 专业版和 Python3.6, Python 中安装的 tensorflow 版本为 1.8.0。整个实验进行过程中设置参数保持高度一致性, 各个模型均采用 3 层卷积层, 每层的嵌入大小均为 64, 抽样批量大小为 1 024, 学习率为 10^{-4} , 迭代次数为 50 次, 节点丢失率和信息丢失率均为 10^{-1} , 采用通用的评价指标^[18-19] ndcg@K 和 recall@K 对模型的预测效果进行评价。首先解释归一化折损累计增益 (normalized discounted cumulative gain, NDCG), 为得到 ndcg, 需计算累计增益 (cgcumulative gain, CG), 指测试集用户项目对的生成预测评分之和, 然后求解折损累计增益 (discounted cumulative gain, DCG), 即按照项目排序先后在累计时每一项除以一个折损系数, 旨在使得排序越靠前的项目对最后评分影响越高, 又由于测试集每个用户对应的项目数量不一致, 需要对 dcg 进行归一化处理, 即需要计算理想折损累计增益 (ideal discounted cumulative gain, IDCG), idcg 是指理想情况下的 dcg 值, 是取用户的模型预测评分结果最高的前 K 个用户项目对评分, 按照 dcg 的计算方式得到 idcg, 而 ndcg 就是 dcg 除以 idcg, 该比值最大为 1, 表示完全预测准确, 该值越接近 1, 表明模型的预测效果越好; recall 是指测试集通过预测模型评价后的召回率, 其详细的计算过程为, 首先得到测试集项目集合和模型预测评分排序前 K 个项目集合的交集数量, 用该数值除以测试集项目数量得到 recall 值, recall 值越大表明预测准确度越高, 模型预测效果越好, 默认 K 为 20, 实验结果如表 2 所示。

表 2 基于时序特征的图卷积协同过滤模型整体性能对比

模型	Gowalla		Yelp		Amazon-book	
	ndcg	recall	ndcg	recall	ndcg	recall
NGCF	0.173 75	0.052 79	0.068 78	0.018 40	0.055 96	0.016 70
TNGCF	0.177 09	0.054 29	0.070 70	0.018 80	0.061 49	0.018 32
提升/%	1.90	2.80	2.80	2.20	9.90	9.80
LightGCN	0.179 01	0.054 83	0.060 25	0.016 16	0.041 18	0.012 93
TLGCN	0.183 54	0.056 11	0.062 22	0.016 56	0.044 99	0.013 97
提升/%	2.50	2.30	3.30	2.50	9.25	8.04

在 Gowalla, Yelp 和 Amazon-book 三个数据集上优化算法,对比传统的 NGCF 和 LightGCN 算法,模型性能指标 recall 和 ndcg 均有不同程度的明显提升,可见引入时序特征的模型预测效果对比原始模型性能均有所提升,证明了本研究理论的可行性和合理性。

3.2 时序特征对模型性能的影响程度研究

目标函数中时序特征损失占总损失的比重反映了时序特征对模型参数更新的影响,决定了时序特征对模型最终性能的影响程度,本研究对代表时序特征的损失进行加权处理,将目标函数更新为式(8):

$$Loss = \sum_{(u,i,j) \in O} -\Delta f(y_{u_i} - y_{u_j}) - p \left(\frac{\Delta f(y_{u_i^{pos}} - y_{u_i^{neg}})}{n} \right) + \lambda (u_i, u_j, u_i^{pos}, u_i^{neg}) \quad (8)$$

通过改变比例系数 p 探究时序特征对模型性能的影响,以 Amazon-book 数据集为例设置对比实验。实验参数设置不变,仅仅改变比例系数 p ,实验结果如图 2 所示。

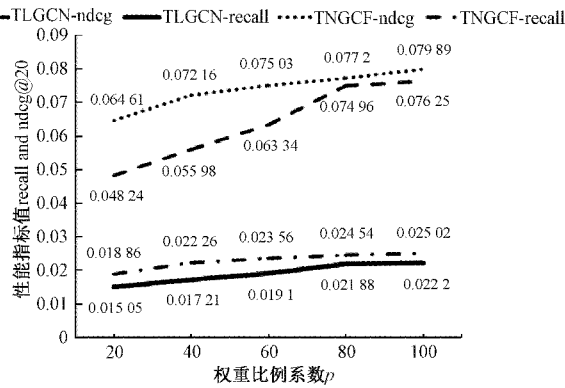


图2 不同比例系数 p 下基于时序特征的图卷积协同过滤算法模型的性能对比

由图 2 可知,随着时序特征的权重比例系数增加即 p 值的增加,模型的性能指标 recall 和 ndcg 逐渐提升,但 p 值到达 100 时曲线基本持平,一方面验证了加入时序特征对算法性能的提升有明显意义,另一方面探索研究出 p 值在 80~100 之间是较为合理的引入时序特征权重区间。

3.3 时序特征与不同信息传递阶数的关系研究

图卷积层数代表了协同过滤中高阶信息的传递阶数,默认使用 3 层图卷积层,为了探究时序特征在不同阶数下对协同过滤的性能优化差异性,实验在各项参数默认相同的情况下,设置不同的卷积层数比较基于时序特征的图卷积协同过滤算法模型的预测性能,实验结果如图 3 所示。

研究发现,TNGCF 模型在越多的卷积层数下表现的性能越好,在 3 层以后性能趋于稳定。而 TLGCN 模型在较少卷积层数下就表现出至少不差于较多卷积层数下的性能。结合前面章节的讨论,LightGCN 是在 NGCF 基础上做了简化,取消了特征变换和非线性激活函数,只保留了邻域聚合,即纯粹通过协同过滤寻找相似用户,而在

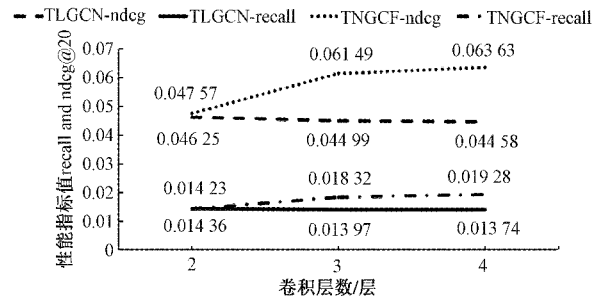


图3 不同卷积层数(即不同邻居阶数)下基于时序特征的图卷积协同过滤算法模型的性能对比

LightGCN 中加入时序特征后,TLGCN 通过较低阶邻居的协同过滤就能提前达到较高阶邻居的协同过滤预测效果,原因在于,在低阶邻居的连通路径图中(参考图 1),项目到目标用户的路径较短,如果未对路径做加权处理,就会出现大量路径相同大量用户相似的情况,推荐命中率会有所降低,但加入时序特征后对路径做加权处理,就可以基本避免这一问题,从而实现较短路径下较快寻找到相似用户集合的效果,提高了模型训练效率的同时,也基本保证了模型性能。

3.4 模型推荐效果应用与分析

为验证本研究算法的可行性与优越性,将训练后的模型应用于现有用户数据的推荐,并对比本研究模型和现有传统模型的推荐准确率 $Precision$,其计算方式如式(9)所示,其中 U 指代全体用户集合, $R(u)$ 指代经过算法模型预测后向用户进行推荐的项目集合, $T(u)$ 指代用户测试集中真实选择的项目集合,预测值越大反映模型推荐的项目与用户真实选择的重合度越高,模型推荐准确度越高,推荐性能越好。

$$Precision = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |R(u)|} \quad (9)$$

设计如图 4 所示的系统应用流程,以获取不同模型在实际应用过程中的性能表现。

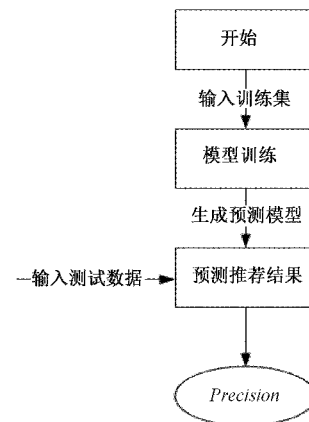


图4 推荐系统应用流程

在系统应用过程中,采用多种模型训练方法,但训练过程中的参数和输入的数据集保持一致,与 3.1 节中的设置相同,从而通过对比最终预测值 *Precision*,分析不同算法训练生成的模型性能差异,如表 3 所示,展示了在 Gowalla, Yelp 和 Amazon-book 三个数据集上 NGCF、LightGCN、TNGCF 和 TLIGHTGCN 模型的预测值情况。

表 3 基于时序特征的推荐算法推荐准确率提升对比

模型	<i>Precision</i>		
	Gowalla	Yelp	Amazon-book
NGCF	0.025 06	0.009 96	0.008 13
TNGCF	0.025 83	0.010 17	0.008 90
提升/%	3.07	2.11	9.47
LightGCN	0.026 07	0.008 90	0.006 12
TLGCN	0.026 79	0.009 17	0.006 60
提升/%	2.76	3.03	7.84

从表 3 中结果显示,基于时序特征的图卷积协同过滤推荐算法训练的推荐模型在不同数据集上相对传统图卷积协同过滤模型训练的模型在推荐的预测上都有不同程度的明显提升,表明本研究提出的算法在应用上达到预期的优化效果。

4 结 论

传统的图卷积协同过滤算法在模型训练过程中仅仅考虑了用户项目的交互信息,而忽略了交互以外的附加信息,本研究将目光聚集在用户-项目交互的时序特征上,从原始数据集中提取用户-项目交互的时序信息并将其参数化,作为重要组份输入到模型训练中,通过一系列消融性实验,验证了方法的有效性和合理性。整体上基于时序特征的图卷积协同过滤算法模型的预测效果对比传统模型性能均有明显提升,并进一步探究时序特征参数在目标损失函数中的比重以及邻居用户的阶数与模型性能之间的关系,大量实验验证了时序特征在推荐模型中的积极作用,同时满足了互联网迅速发展,用户项目交互信息更新频繁,对模型训练效率要求更高的应用需求。值得一提的是,本研究的算法已经在工程中得到了初步应用,在一套基于 VR 的个性化定制培训系统中,由于培训子项目和内容繁多,为实现对用户的个性化推荐和高效培训,引入本研究的算法,并取得了一定成效,后续将开展更多的实验,精细化研究时序特征在模型训练中的精确影响,优化算法寻求基于时序特征的图卷积协同过滤算法模型参数更新的最优解。

参考文献

[1] 聂林海. 我国电子商务发展的特点和趋势[J]. 中国流通经济, 2014, 28(6): 97-101.
 [2] 陈干, 肖博. 基于 Spark 平台的推荐算法研究[J]. 国外

电子测量技术, 2020, 39(3): 71-74.

- [3] 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.
 [4] 宣鹏程, 唐彦, 王汗送. 基于项目类型的群组推荐方法[J]. 电子测量技术, 2019, 42(7): 54-58.
 [5] 陈琦, 吕杰, 张世超. 一个解决协同过滤推荐系统相关问题的新算法[J]. 电子测量技术, 2016, 39(5): 66-69.
 [6] HUANG Z, YU C, NI J, et al. An efficient hybrid recommendation model with deep neural networks[J]. IEEE Access, 2019, 7: 137900-137912.
 [7] OKURA S, TAGAMI Y, ONO S, et al. Embedding-based news recommendation for millions of users[C]. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017: 1933-1942.
 [8] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks[C]. Advances in Neural Information Processing Systems, 2014: 3104-3112.
 [9] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
 [10] HUANG G, LIU Z, VAN D MAATEN L, et al. Densely connected convolutional networks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700-4708.
 [11] VAN DEN O A, DIELEMAN S, SCHRAUWEN B. Deep content-based music recommendation [J]. Advances in Neural Information Processing Systems, 2013, 26: 2643-2651.
 [12] GENG X, ZHANG H, BIAN J, et al. Learning image and user features for recommendation in social networks[C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 4274-4282.
 [13] 余晓霞, 汤宝平, 魏静, 等. 强背景噪声条件下自适应图卷积神经网络的航空发动机附件机匣故障诊断方法[J]. 仪器仪表学报, 2021, 41(8): 78-86.
 [14] 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.
 [15] WANG X, HE X, WANG M, et al. Neural graph collaborative filtering [C]. Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2019: 165-174.

- [16] HE X, DENG K, WANG X, et al. Lightcn: Simplifying and powering graph convolution network for recommendation [C]. Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, 2020: 639-648.
- [17] 龚小龙,孔玲爽,袁川来,等.非线性时间序列粒子群优化B样条网络预测模型[J].电子测量与仪器学报,2017,31(12):1890-1895.
- [18] HE X, LIAO L, ZHANG H, et al. Neural collaborative filtering [C]. Proceedings of the 26th International Conference on World Wide Web, 2017: 173-182.
- [19] YANG J H, CHEN C M, WANG C J, et al. HOP-rec: High-order proximity for implicit recommendation [C]. Proceedings of the 12th ACM

Conference on Recommender Systems, 2018: 140-144.

作者简介

陈一凡,工学硕士,主要研究方向为推荐算法、图像处理、虚拟现实等。

E-mail:1303583067@qq.com

朱民耀,工学硕士,高级工程师,主要研究方向为图像处理等。

E-mail:zhmy68@staff.shu.edu.cn

朱晓强,工学博士,副教授,主要研究方向为虚拟现实等。

E-mail:xqzhu@shu.edu.cn

宋海洋,工学硕士,主要研究方向为图像处理等。

E-mail:shshy@shu.edu.cn

陆小锋(通信作者),工学硕士,高级实验师,主要研究方向为智能信息处理等。

E-mail:luxiaofeng@shu.edu.cn