

基于隐式反馈和加权用户偏好的推荐算法

夏翔, 刘姜*, 倪枫, 陆劲宇

(上海理工大学管理学院, 上海 200093)

摘要:针对现有隐式反馈算法中正负样本划分不合理、忽略用户操作频次、无法准确建模用户偏好等问题,提出一种基于隐式反馈和加权用户偏好的推荐算法(IFW-LFM)。该算法考虑了用户操作频次与正负样本划分间的关系,学习并改进wALS算法,根据用户操作频次从缺失值中重新挖掘潜在正负样本,将用户操作频次大于1时的样本设置为正样本,用户操作频次为1或0时的样本为正样本或负样本,不再需要人为引入负样本;根据用户操作频次对用户偏好程度的影响,定义了置信度,明确用户偏好,并将其应用在隐因子模型的框架中;利用用户收听歌曲起止时间、收听时长等隐式反馈数据,提高隐式反馈样本利用度。在两个音乐数据集上的对比实验结果说明,该算法在准确率、召回率与NDCG值上与5个经典隐式反馈算法(UserCF、ItemCF、LFM、BPR、SVD)相比最大平均提升了45.81%、83.83%和60.33%,具有更优的推荐效果。

关键词:推荐算法;隐式反馈;操作频次;用户偏好;音乐推荐

中图分类号:TP181

文献标识码:A

文章编号:1673-629X(2024)03-0140-07

doi:10.3969/j.issn.1673-629X.2024.03.021

Recommendation Algorithms Based on Implicit Feedback and Weighted User Preferences

XIA Xiang, LIU Jiang*, NI Feng, LU Jin-yu

(School of Management, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract:In view of the unreasonable division of positive and negative samples, ignoring the frequency of user operations and failing to accurately model user preferences, we propose a recommendation algorithm based on implicit feedback and weighted user preferences (IFW-LFM). The algorithm considers the relationship between user operation frequency and positive and negative sample division, learns and improves wALS algorithm, and re-mines potential positive and negative samples from missing values according to user operation frequency. It sets the samples with user operation frequency greater than 1 as positive samples and those with user operation frequency of 1 or 0 as positive or negative samples, eliminating the need to artificially introduce negative samples; defines the confidence level according to the influence of user operation frequency on the degree of user preference, specifies the user preference and applies it to the framework of the hidden factor model; uses the user listening to song start and end time, listening duration and other implicit feedback data to improve the utilisation of implicit feedback samples. The results of the comparison experiments on two music datasets illustrate that the accuracy, recall and NDCG values of the proposed method have a maximum average improvement of 45.81%, 83.83% and 60.33% respectively compared with the five classical implicit feedback algorithms of UserCF, ItemCF, LFM, BPR and SVD, which has better recommendation results.

Key words:recommendation algorithms; implicit feedback; frequency of operation; user preferences; music recommendation

0 引言

随着互联网技术和短视频平台的飞速发展,用户获取信息的能力不断增强,方式不断增多,各类信息的种类和数量也急剧增加。在海量信息资源条件下,用户由于其认知能力和特点的限制,很难在有限的时间内选择到符合需求的信息,这就呈现出“信息过载”和

“信息缺失”的矛盾^[1]。与以往搜索引擎需要用户输入关键字才能从数据库中检索所需内容不同,推荐系统根据每个客户的用户资料自动生成推荐列表,从而显示出更大的便利性^[2]。

推荐算法是推荐系统的核心,主要分为3类:(1)基于内容的推荐^[3-4];(2)协同过滤推荐^[5];(3)混合推

收稿日期:2023-04-28

修回日期:2023-08-28

基金项目:国家自然科学基金资助项目(11701370)

作者简介:夏翔(1998-),女,硕士,研究方向为推荐算法、机器学习;通信作者:刘姜(1983-),女,硕导,博士,研究方向为符号计算、非线性系统、深度学习。

荐^[6]。其中协同过滤(CF)算法广泛应用于推荐领域,通常依赖于用户的历史评分或行为(如购买或观看、收听等)。核心思想是具有相同或相似的价值观、思想观、知识水平和兴趣偏好的用户,其对信息的需求也是相似的^[7]。

用户行为数据通常在项目推荐中起着至关重要的作用,基本分为两类:显式反馈^[8]和隐式反馈^[9]。显式反馈是指用户主动或者在一定的提示和指引下,根据自身偏好对某个推荐对象或者某种服务的主观使用或者体验感受进行打分、评级或是进行语言评价等行为,以此来很明显地表达用户体验感受和偏好程度^[10]。现实生活中,显式反馈的数据收集依赖于用户的主动配合,数据十分稀疏。而隐式反馈是在用户并不知情的情况下,系统记录的用户在系统中的所有操作行为数据^[11],数据量大,数据更稠密、更稳定,通常可以真实反映用户态度,具有很好的应用价值。然而,隐式反馈推荐也存在一些常见的问题,例如缺少正负反馈维度的区分、用户偏好程度不明确、隐式反馈数据样本利用度低等。

根据上述隐式反馈推荐存在的问题,现有的隐式反馈推荐算法主要分为3类:面向单类协同过滤的推荐、引入辅助信息的推荐以及基于排序的推荐。

(1)单类协同过滤推荐。Pan等^[12]于2008年提出了单类协同过滤问题(One-Class Collaborative Filtering, OCCF),主要是指用户所产生的隐式反馈行为仅能反映用户对当前项目进行过一定的操作,但未被用户进行操作过的项目,不能直接反映用户对其没有偏好,也有可能是用户没有接触到该项目。不少学者从权重、置信度加权及矩阵分解等角度对OCCF问题进行研究。Pan等人^[13]提出利用wALS算法(weighted Alternating Least Squares)和权重来缓解负样本问题,其基本思想是:将所有的缺失数据作为负样本,对所有样本进行加权。用户对项目的操作行为代表了用户偏好,具有较高的可行性,将其权重设置为1。相反若用户对其没有操作,即将其相关权重设置为0或1。Hu等人^[14]提出了改进的Implicit ALS算法,考虑了用户偏好和置信度之间的联系,将是否有过用户操作行为的项目进行二进制量化,并给数据集中的正负样本分别分配一个变化的信任权值,在经典的矩阵分解模型基础上进行优化。但都仅根据用户是否有过操作对样本进行正负划分,而忽略用户操作频次带来的影响,且将缺失数据作为负反馈的推荐算法复杂度高,数据集较大。

(2)引入辅助信息的推荐。在单类协同过滤方法的基础上引入辅助信息,便可以挖掘用户潜在兴趣偏好,进而提高推荐效果,引入的辅助信息包含用户或项

目内容、时间或位置的上下文信息、用户的信任网络或社交关系等。Chen等人^[15]引入项目信息,提出一种基于项目的相似性的模型(HCoM)来处理用户购买行为等隐式反馈数据的稀疏问题,在提高OCCF的准确性的基础上还缓解了数据稀疏问题,但引入项目信息的算法针对的特征需求不同,需要根据用户偏好合理选取,所以缺乏一定的通用性。俞东进等人^[16]基于玩家操作次数和操作时常等隐式反馈数据及其时效性,提出了一种基于隐式反馈数据的网吧游戏推荐方法(DIIF-SVD++),克服了用户偏好不完整性问题,实现了游戏的实时个性化推荐。陈婷等人^[17]通过计算用户间的相似度和信任度构建用户之间的偏好关系,提出了一种基于信任的社交推荐算法(Trust-PMF),但对于隐式反馈推荐中负样本的缺失问题依然没有解决。

(3)基于排序的隐式反馈推荐。Rendle等人^[18]于2009年围绕隐式反馈提出了贝叶斯个性化排序(BPR),其关键思想是实现最大后验概率估计规则,以确保观察项目的排名应高于未知项目。Guo等人^[19]利用项目内容信息和隐式反馈进行建模,提出一种自适应个性化排序方法(CA-BPR),加快了个性化排序的成对学习,提高了个性化排序的准确性。申艳梅等人^[20]针对BPR未能充分利用用户行为信息导致数据稀疏的问题,提出了均值贝叶斯个性化排序算法(MBPR),并通过引入遗忘函数,根据用户评分信息对项目进行了正负反馈划分,进一步挖掘了用户对未知项目的偏好,实验结果表明该算法的推荐性能及鲁棒性均有显著提高。

综上所述,现有的单类协同过滤以及引入辅助信息推荐方法中存在正负样本划分不合理、忽略用户操作频次、无法准确建模用户偏好等问题。针对这些问题,该文引入辅助信息并提出了一种基于隐式反馈和加权用户偏好的推荐算法(IFW-LFM)。该算法首先对wALS算法进行改进,考虑用户操作频次与正负样本划分间的关系,从缺失的数据集中划分正负样本,并不再需要人为引入负样本;其次考虑了用户操作频次对用户偏好程度的影响,定义了置信度这一概念,明确了用户偏好程度,并将其应用在隐因子模型的框架中;对于隐式反馈数据利用度低的问题,通过动态考虑时间问题,将用户收听歌曲起止时间、收听时长等隐式反馈样本利用进来,提高了隐式反馈数据利用价值;最后,在两组真实数据集上对算法进行了测试,从两个不同的数据集得出结论,比一个数据集更为严格可信。

1 基于隐式反馈和加权用户偏好的推荐算法

为了便于形式化描述,符号定义见表1。

表 1 符号定义

符号	定义
u	用户
i	项目
U	所有用户集合
I	所有项目集合
A_{ui}	用户 u 对项目 i 的操作频次
p_{ui}	用户 u 对项目 i 的偏好值
E_{ui}	用户 u 对项目 i 的偏好程度(置信度)
d_{ui}^j	用户 u 对歌曲 i 的第 j 次收听时长
t_u^T	用户 u 在时间跨度 T 内最后一次收听歌曲的开始时间
t_{ui}^T	用户 u 在时间跨度 T 内最后一次收听歌曲 i 的开始时间
p_{uk}	用户 u 对第 k 个隐因子类别的兴趣
q_{ik}	项目 i 与第 k 个隐因子类别之间的权值

1.1 隐式反馈中正负反馈维度区分

针对隐式反馈数据缺少正负反馈维度的区分问题,即隐式反馈一般只有正反馈而缺失负样本的问题,常见的思路是人为引入负样本或者将所有的缺失值均作为负样本来使用,这两种解决策略都与实际数据有较大偏差。Pan 等人^[13]提出利用 wALS 算法(weighted Alternating Least Squares)和权重来缓解负样本问题,仅仅只考虑了用户是否进行过操作,但将所有缺失数据都当成负反馈其实是不合理的,忽略了用户频次对正负反馈划分的影响。因此,该文学习 wALS 算法的思想从缺失值中划分潜在的正负样本。首先,定义用户 u 和项目 i ,设置 $A_{ui} > 1$,这与 wALS 中将 $A_{ui} = 0$ 设为正样本不同,原因在于表示用户 u 对项目 i 进行了不止一次的操作,说明用户对该项目有一定的偏好,且这种推断具有较强的可信度,即设置该项目的偏好值 $p_{ui} = 1$ 。对于缺失值来说 $A_{ui} = 0$,表示用户 u 对项目 i 没有进行过操作,但只能推断这些项目可能并不一定为负样本,因此设置其偏好 $p_{ui} = \delta \in [0, 1]$ 。对于临界值 $A_{ui} = 1$,即用户对推荐项目 i 只进行过一次操作,该文也将其偏好设定为与 $A_{ui} = 0$ 时取值一致,这是因为一次点击并不能推断用户偏好,可能为正样本,也可能为负样本,如公式 1 所示:

$$p_{ui} = \begin{cases} 1, & A_{ui} > 1 \\ \delta \in [0, 1], & A_{ui} = 0, 1 \end{cases} \quad (1)$$

其中,当 $\delta = 0$ 时,说明所有缺失值都无法划分正负样本,即将所有样本都作为正样本,当 $\delta = 1$ 时,即将所有缺失值都当做负反馈样本。

1.2 引入偏好程度

单纯将用户有无隐式行为作为区分用户偏好与否的标准,这样简单的判断是有问题的。比如当用户点击某一首歌曲时,用户可能对这首歌曲没有偏好,可能

只是分享给朋友收听,也可能用户不知道该首歌曲的存在。而当用户某种隐式行为的操作频次增加时,用户对该项目的喜好程度随之增加。因此,结合音乐收听背景,文中模型引入一个新的概念偏好程度 E_{ui} ,它表示用户 u 在某一操作频次下对该项目 i 具有偏好的置信度,如公式 2 所示:

$$E_{ui} = 1 + \beta A_{ui} \quad (2)$$

其中, A_{ui} 为用户收听歌曲操作频次; β 为偏好程度的计算系数,通过控制其大小可以为每一个用户-项目对提供最小置信度。由式 2 可以看出,当用户隐式反馈中收听歌曲的频次数值越大,其偏好程度 E_{ui} 越大。

1.3 提高隐式反馈样本利用率

如上所述,通过划分正负反馈以及引入用户偏好程度只能解决隐式反馈数据中的前两个问题,并没有解决传统隐式反馈推荐算法样本利用率低的问题。例如,音乐推荐背景中的用户收听歌曲的起止时间、收听时长等数据在传统的 LFM 模型中并没有加以利用,因而不能充分发挥隐式反馈推荐作用。时间敏感性对于隐式反馈推荐系统同样至关重要。对此, Li 等人^[21]提出用户偏好呈指数衰减,这一观点与艾宾浩斯遗忘曲线相吻合,即人们忘记事情的速度先快后慢。申艳梅^[20]也将遗忘函数加入到了 MBPR 算法中,考虑了用户兴趣随时间变化的特征。但统一的遗忘函数并未实现用户遗忘规律的个性化,对用户收听时长数据也没有加以利用。因此,该文结合音乐推荐背景,用户收听歌曲起止时间,对歌曲收听时长等数据,利用经典的矩阵分解方法,提出了如下优化目标函数:

$$r_{ui} = \min \sum_{u=1}^m \sum_{i=1}^n E_{ui} (p_{ui} - x_u y_i^T)^2 + d_{ui}^j (t_u^T - t_{ui}^T) \quad (3)$$

1.4 用户兴趣预测和推荐算法

在实际应用中,较多显式反馈中的推荐算法并不能直接应用于隐式反馈推荐中。该文选取 LFM 方法实现用户对某首歌曲的兴趣预测,通过将用户和项目都投射至隐因子空间得到隐式特征,分别计算用户以及项目每一个隐因子类别之间的关系,如式 4 所示:

$$r'_{ui} = p_u^T q_i = \sum_{a=1}^s p_{ua} q_{ia} \quad (4)$$

综上,需要对优化的如下损失函数找到合适的参数 p_{uk}, q_{ik} :

$$\text{Loss} = \sum_{ui \in R} (r_{ui} - r'_{ui})^2 = \sum_{ui \in R} \left[(r_{ui} - \sum_{a=1}^s p_{ua} q_{ia})^2 + \lambda \|p_u\|^2 + \lambda \|q_i\|^2 \right] \quad (5)$$

其中,正则化参数 $\lambda \|p_u\|^2 + \lambda \|q_i\|^2$ 是为了避免模型过度拟合,即利用随机梯度下降法求解 p_{uk} 和 q_{ik} 。主要步骤如下:

(1) 通过反复实验设定合适的学习率 Learning rate 和正则化参数 λ 。

(2) 计算 p_{uk} 和 q_{ik} 的偏导数, 找到最速下降方向:

$$\frac{\partial L}{\partial p_{uk}} = -2q_{ik} \cdot (r_{ui} - \sum_{a=1}^s p_{uk} q_{ik}) + 2\lambda p_{uk} \quad (6)$$

$$\frac{\partial L}{\partial q_{ik}} = -2p_{uk} \cdot (r_{ui} - \sum_{a=1}^s p_{uk} q_{ik}) + 2\lambda q_{ik} \quad (7)$$

(3) 根据最速下降方向, 不断更新 p_{uk} 和 q_{ik} , 反复迭代优化参数:

$$p_{uk} \leftarrow p_{uk} + \alpha [q_{ik} \cdot (r_{ui} - \sum_{a=1}^s p_{uk} q_{ik}) - \lambda p_{uk}] \quad (8)$$

$$q_{ik} \leftarrow q_{ik} + \alpha [p_{uk} \cdot (r_{ui} - \sum_{a=1}^s p_{uk} q_{ik}) - \lambda q_{ik}] \quad (9)$$

具体算法描述如下:

算法: 基于隐式反馈和加权用户偏好的推荐算法 (IFW-LFM)

输入: 用户-项目操作矩阵 A , 偏好程度计算系数 β , 偏好值 $\delta \in [0, 1]$

输出: 用户-项目预测偏好程度矩阵 R'

Step 1: 读取数据集, 初始化用户-项目操作矩阵 A ;

Step 2: 根据式 1 计算用户偏好 p_{ui} , 根据偏好程度计算系数 β 计算式 2 用户偏好程度;

Step 3: 根据用户收听歌曲起止时间, 对歌曲收听时长等信息, 得到式 3 的优化函数;

Step 4: 根据式 6~9 反复迭代更新计算 p_{uk} , q_{ik} , 使得式 5 的损失函数最小;

Step 5: 根据式 4 预测用户兴趣, 得到用户兴趣预测值 r'_{ui} , 根据用户兴趣预测值返回用户-项目预测偏好程度矩阵 R' ;

Step 6: 设定 Top-N 推荐中 N 的数值, 根据步骤 5 中得到的 R' , 选取前 N 个产生推荐列表。

2 实验评估

2.1 实验数据集

实验利用了两组国外大型音乐数据集, 以便更好地呈现出算法的性能。数据集 1 为 last.fm-1k, 包含两个表, 其中用户历史交互记录表记录了从 2005 年 7 月到 2009 年 5 月 992 个用户的全部收听历史记录 (约 2 000 万条), 包括 userID, event timestamp, artistID, artist_name, songID, song_name 等信息, 用户特征表包含性别、年龄、国家、注册时间等。另一组数据是来自数据集 last.fm-360k, 包含六个表, 记录了 1 892 个用户的全部收听历史记录, last.fm 360k 数据的格式与 1k 数据大致相同, 增加了 user 对 artist 收听次数的记录。

2.2 评价指标

在一般的推荐算法指标评价中, 评分预测通常使用平均绝对误差和均方根误差来度量, 但在隐式反馈中不存在用户“实际”评分, 因此, 上述两个由用户实

际评分与预测评分之间的误差来评估推荐性能的指标不适合计算。该文的目的是对用户进行个性化的音乐推荐, 注重的是最终的推荐效果, 而不是预测评分与实际评分间的误差。因此, 选取准确率、召回率和 NDCG 作为实验的评估标准。准确率为相关的推荐物品数占推荐物品总数的比率, 召回率为相关的推荐数占实际相关物品总数的比率, NDCG 表示归一化折损累计概率, 与 DCG 相比, 可以更准确地评估排序性能。其计算方法如式 10~12 所示:

$$\text{Precision} = \frac{t_{pu}}{t_{pu} + f_{pu}} \quad (10)$$

$$\text{Recall} = \frac{t_{pu}}{t_{pu} + f_{nu}} \quad (11)$$

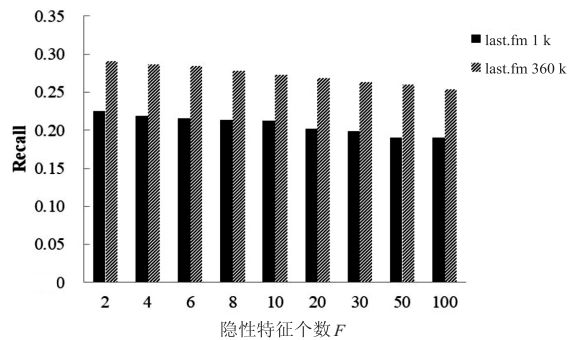
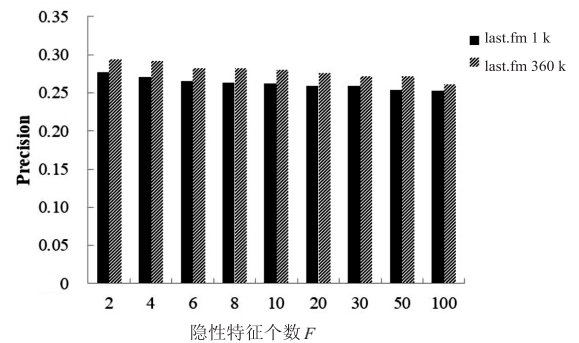
$$\text{NDCG}_N = \frac{\text{DCG}_N}{\text{IDCG}_N}, \text{DCG}_N = \sum_{i=1}^N \frac{2^{\text{rel}_i} - 1}{\log_2(i + 1)} \quad (12)$$

其中, N 为推荐列表中的项目数量, rel_i 表示推荐列表中第 i 个位置的项目的相关分数。实验中, 如果结果为正, 则 $\text{rel}_i = 1$, 否则, $\text{rel}_i = 0$ 。IDCG (理想折损累计增益) 表示可得到的 DCG 的最大值。

2.3 参数影响分析

在文中算法中, 4 个重要参数分别为隐性特征个数 F , 负样本比例 ratio, 正则化参数 λ , 以及学习速率 Learning rate, 实验研究了这 4 个参数对推荐结果的影响。

由图 1 可以看出, 随着隐性特征数目的增多, 算法的 Precision、Recall 和 NDCG 值均成下降趋势。这是由于在音乐推荐背景中, 隐性特征个数相对来说比较单一, 用户的兴趣偏好主要受几个关键的隐性特征影响, 并不是隐性特征越多越好。



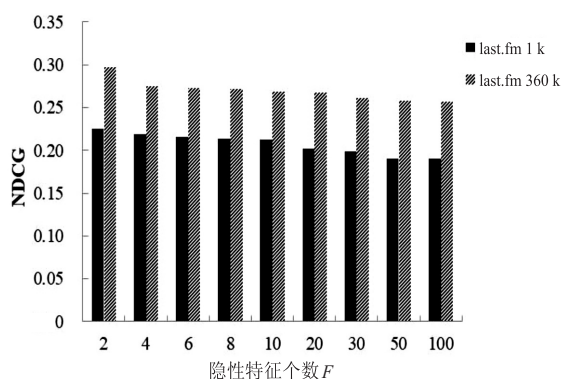


图1 隐性特征个数对性能的影响

如图2所示,负样本比例 ratio 对算法的性能也有一定影响。实验发现,当 ratio 小于等于 5 时,准确率、召回率和 NDCG 值均随 ratio 的增加而提高,当 ratio 大于 5 后,图像趋于平缓,三类评价指标基本保持稳定。

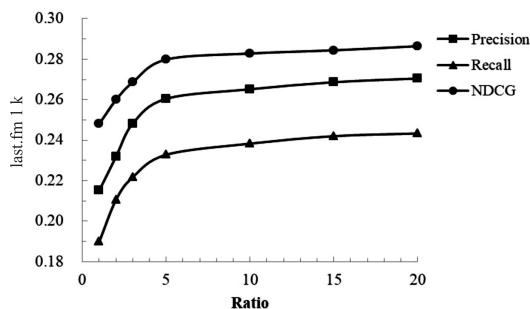


图2 负样本比例对性能的影响

图3展示了数据集1中准确率等指标在不同正则化参数 λ 值下的变化。当 λ 在 0.07 至 0.4 之间时推荐效果最好,当 $\lambda < 0.07$ 时,推荐效果随 λ 值的增大而提升,当 $\lambda > 0.4$ 时,模型不能很好拟合,经过实验可以选取合适的 λ 值,使得算法的推荐效果更好。

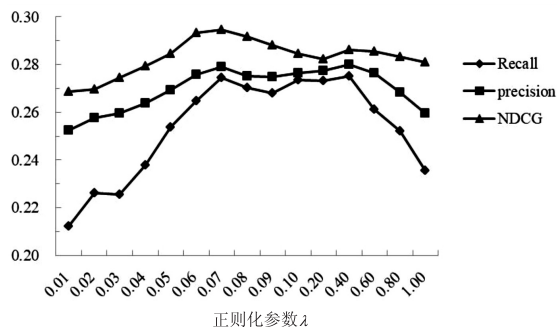


图3 正则化参数对性能的影响

在梯度下降法中,学习率 Learning rate 会直接影响算法的收敛速度和最终优化效果。由图4可知,推荐效果随迭代次数的增加会先有一个增强趋势至峰值,而后逐渐下降。在数据集1中,当初始学习率为 0.002 时,推荐效果最为理想,但需要 30 次左右的迭代才能达到,当学习率为 0.001 时,推荐效果反而下降,这说明当学习率过小时,收敛速度变慢,推荐效果

也随之下降,这要进一步地动态调整学习率,使得迭代次数也调整为合适的值。

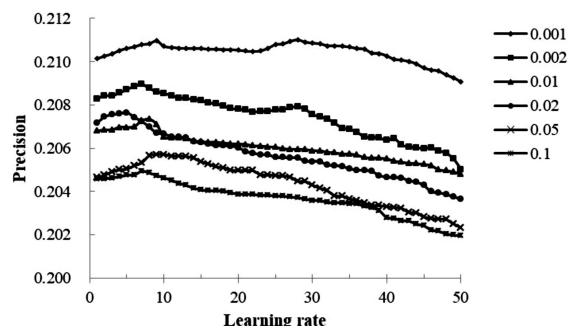


图4 学习率对性能的影响

2.4 实验结果与分析

基于隐式反馈的音乐推荐算法较少,同时考虑到隐式反馈应用到音乐推荐场景本身的特点,一些已有的隐式反馈推荐算法不适用于音乐推荐场景。该文选择以下 5 种算法作为对比算法。

(1) UserCF: 基于用户的协同过滤推荐算法。

(2) ItemCF: 基于项目的协同过滤推荐算法。

(3) BPR^[18]: 将推荐问题转化为成对排序问题,将用户参与项与未参与项分别作为正负反馈,构造矩阵参数,从而对每名用户产生推荐列表。

(4) SVD^[21]: 采用随机梯度下降法优化用户及物品特征矩阵,使其接近原始的评分矩阵,常应用于隐式反馈推荐中。

(5) LFM^[22]: 通过隐含特征联系用户兴趣和物品,是经典的隐式反馈推荐方法之一。

为达到理想的推荐效果,实验选取了 80% 的数据作为训练集,20% 作为测试集,隐性特征个数 $F = 10$,负样本比例 Ratio = 5,正则化参数 $\lambda = 0.07$,学习率 Learning rate = 0.002,模型迭代次数为 30,实验结果如表 2 和表 3 所示。

由表 2 和表 3 可以观察到,文中算法在两个数据集下明显优于 5 种对比算法,其中 Item CF 算法的召回率、准确率和 NDCG 值最低,比 UserCF 的值都要低,这说明基于项目的协同过滤算法没有利用用户行为数据,其推荐效果最差。而相较于 SVD 仅考虑了隐式特征,并未涉及隐式正负反馈样本的划分,BPR 算法将所有缺失值都当做负样本,LFM 缺乏时间敏感性的问题,文中提出的基于隐式反馈和加权用户偏好的推荐算法对隐式反馈正负样本进行了合理的划分,同时考虑了收听时长等时效性隐式反馈数据,推荐性能均有显著提升。数据集 2 中,IFW-LFM 算法在时间跨度为 180 天时,其召回率、准确率和 NDCG 值较 UserCF、ItemCF、LFM、BPR、SVD 分别最大平均提升了 45.81%, 83.83% 和 60.33%,这说明考虑用户操作

频次从缺失值中划分正负样本与考虑其和用户偏好间的关系并引入时间辅助信息的思路是行之有效的。

表 2 last. fm 1k 数据集实验结果

d	评价指标	IFW-LFM	UserCF	ItemCF	BPR	SVD	LFM	最小提升/%
14	Recall	0.168 3	0.119 0	0.078 4	0.135 2	0.145 3	0.162 1	3.82
	Precision	0.221 0	0.087 6	0.072 4	0.162 0	0.182 8	0.211 9	4.29
	NDCG	0.222 4	0.115 3	0.087 1	0.171 3	0.188 2	0.211 0	5.40
30	Recall	0.185 5	0.120 9	0.099 9	0.147 9	0.162 3	0.173 5	6.92
	Precision	0.251 6	0.110 3	0.093 6	0.186 5	0.223 6	0.240 2	4.75
	NDCG	0.239 2	0.117 9	0.105 5	0.180 7	0.217 7	0.236 0	1.36
60	Recall	0.208 5	0.147 3	0.109 0	0.183 8	0.192 9	0.194 4	7.25
	Precision	0.278 5	0.133 4	0.115 8	0.195 3	0.240 5	0.258 8	7.61
	NDCG	0.262 6	0.128 7	0.115 4	0.203 6	0.249 4	0.253 2	3.71
90	Recall	0.225 8	0.157 3	0.127 7	0.201 1	0.204 0	0.212 0	6.51
	Precision	0.293 9	0.166 4	0.156 7	0.218 9	0.251 3	0.262 4	12
	NDCG	0.288 5	0.135 2	0.127 3	0.225 8	0.260 8	0.273 9	5.33
180	Recall	0.262 9	0.181 8	0.158 3	0.211 8	0.225 7	0.232 6	13.03
	Precision	0.314 1	0.189 3	0.179 1	0.234 4	0.272 0	0.276 6	13.56
	NDCG	0.350 0	0.190 9	0.149 1	0.261 2	0.302 9	0.311 5	12.36

表 3 last. fm 360k 数据集实验结果

d	评价指标	IFW-LFM	UserCF	ItemCF	BPR	SVD	LFM	最小提升/%
14	Recall	0.229 6	0.135 0	0.111 9	0.179 8	0.190 5	0.225 4	1.86
	Precision	0.249 7	0.152 9	0.142 2	0.181 1	0.204 9	0.241 0	3.61
	NDCG	0.224 7	0.145 1	0.138 2	0.183 7	0.192 3	0.215 0	4.17
30	Recall	0.239 5	0.158 0	0.137 3	0.187 1	0.201 8	0.233 0	2.79
	Precision	0.260 7	0.163 6	0.150 3	0.203 2	0.211 2	0.256 3	1.72
	NDCG	0.242 1	0.179 5	0.165 4	0.209 3	0.216 1	0.238 0	1.72
60	Recall	0.277 2	0.166 5	0.143 0	0.212 2	0.216 9	0.251 3	10.31
	Precision	0.275 4	0.186 4	0.177 4	0.224 2	0.237 4	0.263 6	4.48
	NDCGP	0.265 9	0.195 8	0.180 1	0.223 6	0.238 0	0.249 8	6.45
90	Recall	0.294 7	0.173 5	0.159 2	0.224 4	0.235 3	0.257 4	14.49
	Precision	0.299 7	0.207 2	0.181 0	0.225 9	0.251 7	0.272 8	9.86
	NDCG	0.286 8	0.211 6	0.205 6	0.238 7	0.250 0	0.262 0	9.47
180	Recall	0.319 6	0.198 4	0.173 8	0.244 8	0.252 3	0.269 0	18.81
	Precision	0.319 5	0.219 2	0.208 2	0.245 3	0.270 9	0.283 0	12.90
	NDCG	0.312 7	0.229 9	0.207 3	0.268 0	0.272 8	0.282 4	10.73

3 结束语

针对现有隐式反馈中存正负样本划分不合理、忽略用户操作频次、无法准确建模用户偏好的问题,提出了基于隐式反馈和加权用户偏好的推荐算法(IFW-

LFM)。该算法借助 wALS 思想,考虑用户操作频次与正负样本划分间的关系,并对临界值的用户偏好进行了讨论;接着考虑了用户收听歌曲频次对用户偏好程度的影响,根据用户操作频次定义了置信度,明确了用户偏好程度这一重要信息;最后提高了隐式反馈样

本中关于时效性信息的利用,根据用户收听歌曲起止时间与收听时长等样本数据,构建了隐式反馈推荐模型,并利用 LFM 算法预测用户兴趣,提高了推荐性能。然而,在考虑样本利用率低的问题时,仅加入了时效性信息而未考虑其他辅助信息,未来将在此模型的基础上考虑加入更多的辅助信息,以进一步提高样本利用率。此外,该文是基于纯粹隐式反馈数据上的推荐,对显式反馈数据没有加以利用,下一步也可以考虑这两种反馈数据如何进行有效的结合以提高推荐性能。

参考文献:

- [1] BORGS C, CHAYES J, KARRER B, et al. Game-theoretic models of information overload in social networks[C]//International workshop on algorithms and models for the web-graph. Stanford: Springer, 2010: 146–161.
- [2] BELKIN N J, CROFT W B. Information filtering and information retrieval: two sides of the same coin? [J]. Communications of the ACM, 1992, 35(12): 29–38.
- [3] PAZZANI M J, BILLSUS D. Content-based recommendation systems[M]//The adaptive web: methods and strategies of web personalization. Berlin: Springer, 2007: 325–341.
- [4] LOPS P, JANNACH D, MUSTO C, et al. Trends in content-based recommendation[J]. User Modeling and User-Adapted Interaction, 2019, 29(2): 239–249.
- [5] FKIHF. Similarity measures for collaborative filtering-based recommender systems: review and experimental comparison [J]. Journal of King Saud University – Computer and Information Sciences, 2022, 34(9): 7645–7669.
- [6] KIM B M, LI Q, PARK C S, et al. A new approach for combining content-based and collaborative filters[J]. Journal of Intelligent Information Systems, 2006, 27(1): 79–91.
- [7] 黄小斌. 网络信息过滤原理与应用[M]. 北京: 北京图书馆出版社, 2005.
- [8] PEREIRA M, ETEMAD E, PAULOVICH F. Iterative learning to rank from explicit relevance feedback[C]//Proceedings of the 35th annual ACM symposium on applied computing. Brno: ACM, 2020: 698–705.
- [9] PRATHAMA F, SENJAYA W F, YAHYA B N, et al. Personalized recommendation by matrix co-factorization with multiple implicit feedback on pairwise comparison[J]. Computers & Industrial Engineering, 2021, 152: 107033.
- [10] 张亚明, 高忠萍, 高祎晴, 等. 融合显隐式反馈的协同过滤推荐算法研究[J]. 小型微型计算机系统, 2022, 43(4): 731–740.
- [11] YIN J, WANG Z S, LI Q, et al. Personalized recommendation based on large-scale implicit feedback[J]. Journal of Software, 2014, 25(9): 1953–1966.
- [12] PAN R, ZHOU Y, CAO B, et al. One-class collaborative filtering[C]//Eighth IEEE international conference on data mining. Pisa: IEEE, 2008: 502–511.
- [13] PAN R, SCHOLZ M. Mind the gaps: weighting the unknown in large-scale one-class collaborative filtering[C]//Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining. New York: ACM, 2009: 667–676.
- [14] HU Y, KOREN Y, VOLINSKY C. Collaborative filtering for implicit feedback datasets[C]//Eighth IEEE international conference on data mining. Pisa: IEEE, 2008: 263–272.
- [15] CHEN L, XIN X, WONG D, et al. HCoM: item-based similarity model for heterogeneous implicit feedback[C]//18th IEEE international conference on mobile data management (MDM). Daejeon: IEEE, 2017: 40–49.
- [16] 俞东进, 陈 聪, 吴建华, 等. 基于隐式反馈数据的个性化游戏推荐[J]. 电子学报, 2018, 46(11): 2626–2632.
- [17] 陈 婷, 朱 青, 周梦溪, 等. 社交网络环境下基于信任的推荐算法[J]. 软件学报, 2017, 28(3): 721–731.
- [18] RENDLE S, FREUDENTHALER C, GANTNER Z, et al. BPR: Bayesian personalized ranking from implicit feedback [C]//Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence. Arlington: AUAI, 2009: 452–461.
- [19] GUO W, WU S, WANG L, et al. Adaptive pairwise learning for personalized ranking with content and implicit feedback [C]//IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology (WI-IAT). Singapore: IEEE, 2015: 369–376.
- [20] 申艳梅, 姜冰倩, 敖 山, 等. 基于遗忘函数的均值贝叶斯个性化排序算法研究[J]. 计算机应用研究, 2021, 38(5): 1350–1354.
- [21] LI G, CHEN Q. Exploiting explicit and implicit feedback for personalized ranking [J]. Mathematical Problems in Engineering, 2016(1): 1–11.
- [22] 陈 晔, 刘志强. 基于 LFM 矩阵分解的推荐算法优化研究[J]. 计算机工程与应用, 2019, 55(2): 116–120.