

基于用户间接信任及高斯填充的推荐算法

朱佩佩 龙敏

(长沙理工大学计算机与通信工程学院 长沙 410114)

摘要 现有的推荐算法引入用户显式信任,可以有效地提高推荐精度,但没有充分挖掘社交关系,而间接信任在社交信息中具有更加丰富的潜在价值,进一步影响到推荐质量。虽然对于间接信任也存在相关研究,但是计算复杂,采取的信任传递路径不充分。故此,通过信任传递网络图,将各分支节点与总路径节点比例经过逐节点相乘的方式全局获取信任间接值,然后采用信息熵分析用户社交信任关系的实际表现,调整信任,以形成间接信任的计算模型 IpmTrust,并以此模型设计一种考虑用户间接信任的推荐算法 GITCF。该算法利用高斯模型对评分矩阵进行填充,然后采用修正的余弦计算用户相似度。通过 IpmTrust 计算间接信任后,将用户信任与相似度进行一定线性加权融合,最后采用改进的近邻预测进行推荐。实验在 Matlab 仿真平台上进行,对 RMSE, MAE 两个指标评测,将 GITCF 与现有的推荐算法、传统推荐算法做比较。GITCF 的推荐精度比现有推荐的推荐精度提高了近 7%,也高于不含信任的传统推荐的推荐精度。实验结果表明, IpmTrust 模型有一定的有效性,设计的推荐算法可改善推荐结果的质量。

关键词 推荐算法,间接信任,用户偏好,信息熵,高斯填充

中图分类号 TP391 **文献标识码** A

Recommendation Methods Considering User Indirect Trust and Gaussian Filling

ZHU Pei-pei LONG Min

(School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha 410114, China)

Abstract The existing recommendation algorithm introduces the user display trust, which can effectively improve the recommendation accuracy, but does not fully exploit the social relationship, and the indirect trust has richer potential value in the social information, further affecting the recommendation quality. Although there are related studies on indirect trust, the calculation is complicated and the path of trust transmission is not sufficient. Therefore, through the trust transfer network diagram, the ratio of each branch node to the total path node is multiplied by node-by-node to obtain the trust indirect value globally. Secondly, the information entropy is used to analyze the actual performance of the user's social trust relationship, and the trust is adjusted to form the calculation model IpmTrust of indirect trust. And based on this model, a recommendation algorithm GITCF considering user indirect trust is designed. The algorithm uses the Gaussian model to fill the scoring matrix, and then uses the modified cosine to calculate the user similarity. After IpmTrust calculates the indirect trust, the user trust and the similarity are linearly weighted and merged. Finally, the improved neighbor prediction is used for recommendation. The experiment was carried out on the Matlab simulation platform. The RMSE and MAE evaluations were compared. The GITCF was compared with the existing recommendation algorithms and the traditional recommendation algorithms. The GITCF is improved by nearly 7% compared with the existing recommendation recommendation, and is also higher than the trust-free ones. The experimental results show that the IpmTrust model has certain validity, and the recommended algorithm can improve the quality of recommendation results.

Keywords Recommendation algorithm, Indirect trust, User preference, Information entropy, Gaussian filling

1 引言

传统的协同过滤(Collaborative Filtering, CF)推荐算法可以快速而准确地帮助用户获取所需信息^[1-2],但由于用户反馈信息少、数据源稀疏性大,其推荐精度存在局限性。近年

来,为了缓解数据稀疏性、提高推荐算法质量,许多学者在推荐中引入多类辅助信息^[3-6],并通过实验分析证明了这类方案的可行性。其中,信任关系相比其他位置、时间、签到等辅助信息更能挖掘用户潜在价值^[7],通过信任构成的社交网络图,在推荐领域可以关联到其他信任用户,从而进一步完善推荐

效果^[8]。另外,信任一般分为直接信任(Direct Trust)和间接信任(Indirect Trust),直接信任收集方便,而间接信任的采集困难。因此,一些学者进行了有关信任推荐方面的工作。

Kalaï等^[9]将直接信任作为一种检测机制,结合社交和协作,提出一种基于用户信任和专业领域知识的服务推荐。游静等^[10]关注信任数据的可靠度,提出一种对信任可靠性进行动态管理的算法模型;Ma等^[11]将直接信任引入聚类推荐中,结合信任与非信任概念,来挖掘用户信任社区,提出一种将信任聚合的推荐算法;Chen等^[12]利用用户社会关系,并结合信任传播机制及时序消息,提出一种新型的个性化推荐模型TTSMF;朱敬华等^[13]考虑到信任可改进推荐结果,利用信任与不信任之间的关联性质,以计算信任度,提出在基于位置服务上融入信任的推荐算法。以上推荐算法将直接信任融入到推荐中,较好地缓解了数据稀疏性,通过实验分析,这些算法可以提高推荐准确率。

虽然直接信任完善了推荐结果,但推荐算法本身的问题仍然存在,即用户显示反馈少、相互信任的用户无法联系起来。实际中,用户浏览其他用户的停留时间、主页来往次数(未关注、未点赞)、内容留言等以非显示的形式所表现的间接信任数据十分丰富^[14],虽然数据采集存在一定困难,但由间接信任引出的隐式用户增多,在构成的信任路径上,相联用户集合增大。所以,考虑间接信任,探索用户间潜在的关联性,可解决推荐算法数据稀疏问题。一些研究学者做了如下工作。

潘一腾等^[15]利用社交网络思想,估量间接信任值,提出一种信任关系隐含相似度方法,该方法确保相似用户的精度,从而提升推荐准确率。Xu等^[16]将直接信任与间接信任相结合,提出一种基于社会信任行为的矩阵分解推荐算法TBSVD,该算法在推荐准确性方面有着不错的表现。Lingam等^[17]建立社交信任关联模型来评估信任路径的质量效果,通过评估后选择最佳信任路径,以此基础进行目标用户的推荐,实验表明该算法可提高推荐准确性。Azadjalal等^[18]考虑到用户项目评分矩阵的稀疏性,通过特定的可靠程度度量法来识别间接信任,提出一种基于 Pareto 和置信度概念的信任推荐方法。Yin等^[19]将社会学的相似度和信任相结合,把信任传递路径模型、循环信任模型进行混合权重计算,提出一种改进的基于信任社交的推荐算法;Gohari等^[20]利用信任建立一种置信模型,该模型不仅取决于用户间的信任关系,还取决于用户评分偏好,实验结果证明了该算法的有效性。

由此可见,推荐算法中引进间接信任机制,可以进一步提升推荐精确度,但这些研究工作大多未充分挖掘间接信任及其与用户的社交关联性,或其信任传递路径只从单一角度去选择,没有考虑到全局路径下间接信任对整体算法的影响。故此,本文综合以上情况,提出一种基于用户间接信任及高斯填充的推荐算法。本文主要内容如下:1)提出一种新的间接信任计算模型,该模型不仅可计算全局路径下用户间接信任,而且引入信息熵来评价用户信任质量,以整合信任值;2)提出一种新思路,即采用高斯模型对用户初始评分矩阵进行填补,为能获取相似精度高的用户提供条件;3)综合间接信任及高

斯模型,对设计的推荐算法进行实验,获得了更好的推荐效果。

2 传统协同过滤推荐算法的设计流程

2.1 预处理

在推荐算法的预处理阶段,通常把用户历史记录数据构建成评分矩阵 $\mathbf{SM}_{m \times i}$,其中 m 表示用户, i 表示项目, $\mathbf{SM}_{m \times i}$ 表示 m 个用户对 i 个项目的评分集合,其 $\mathbf{SM}_{m \times i}$ 如表1所列(评分制采用1~5分)。

表1 $\mathbf{SM}_{m \times i}$

	Item ₁	Item ₂	Item ₃	...	Item _i
User1	2	3	—	...	3
User2	—	3	2	...	2
...
User _m	—	2	2	...	5

通过二维矩阵的形式来转换用户数据,可以更加直观地了解到用户评分状况,在表1中,“—”表示用户对该项目未进行评分,为空白区域。一般来说,用户不愿花费额外时间对项目进行反馈操作^[14],这也造成推荐算法的数据稀少,出现冷启动问题。

2.2 计算用户相似度

推荐算法将用户记录进行预处理后,结合评分矩阵,根据相似度算法来计算各个用户间的相似程度,为之后给目标对象推荐做准备。传统推荐算法常用的相似度度量法有皮尔逊相关系数度量法、余弦相似度、修正的余弦^[1]等。

本文考虑到需要引入信任等多维数据,且修正的余弦在余弦相似度上减去均值来缩小误差,因此采用修正的余弦函数相似度^[1],其表达式为:

$$\sin(m, k) = \frac{\sum_{i \in U_{m,k}} (C_{m,i} - \bar{C}_m)(C_{k,i} - \bar{C}_k)}{\sqrt{\sum_{i \in U_m} (C_{m,i} - \bar{C}_m)^2} \sqrt{\sum_{i \in U_k} (C_{k,i} - \bar{C}_k)^2}} \quad (1)$$

$U_{m,k}$ 属于用户 m, k 共同评分的项目集合, $C_{m,i}$ 为用户 m 对项目 i 的评分, $C_{k,i}$ 为用户 k 对项目 i 的评分, \bar{C}_m 为用户 m 对所有项目的平均评分, \bar{C}_k 为用户 k 对所有项目的平均评分。

2.3 评分预测

经过相似度算法计算用户相似度后,推荐算法会将用户相似集合进行排序,传统的推荐一般采用 Top-N 排序法,按照相似程度由大到小顺序,将排名靠前的用户通过近邻预测模型来预测目标对象对未评分项目的评分,以达到推荐目的。其中,传统近邻预测模型^[1]如式(2)所示:

$$pre(f, i) = \frac{\sum_{n \in N_{f,i}} \sin(m, n) \times C_{f,i}}{\sum_{n \in N_{f,i}} \sin(m, n)} \quad (2)$$

其中, $pre(f, i)$ 为用户 f 对项目 i 的预测评分, $\sin(m, n)$ 表示用户 m 与用户 n 的相似度, $N_{f,i}$ 表示用户 f 评分的 i 个项目集合, $C_{f,i}$ 表示用户 f 对项目 i 的实际评分。

3 考虑用户间接信任及高斯填充的推荐方法

本文在传统推荐算法的基础上,考虑了用户间接信任,设计了一种融入用户间接信任及高斯填充的推荐算法,其思路如图1所示。

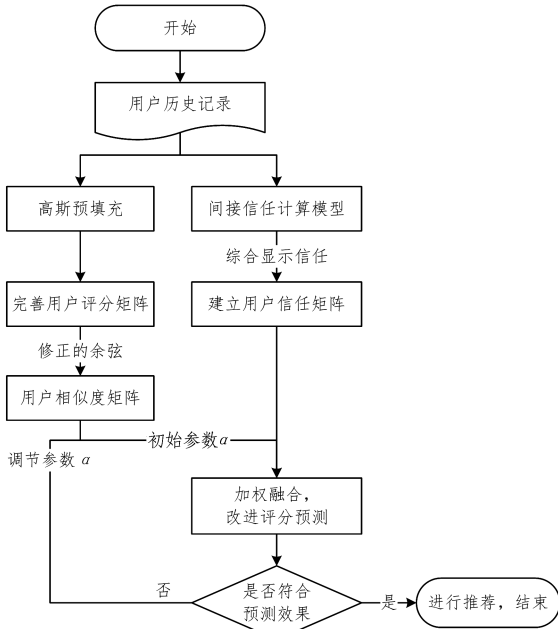


图1 推荐方法示意图

3.1 预处理高斯填充

通过 2.1 节可知,因为推荐系统的数据稀疏性、冷启动问题,在预处理阶段,用户评分矩阵相当稀疏,这给后续计算用户的相似度的精度带来影响;由 2.2 小节可知,相似度的计算需要依赖共同评分的用户,而用户的评分数据若过于稀少,则其相似差异性会增大,造成推荐的不准确性。

因此,针对评分矩阵进行相关填充,缩小相似用户之间的误差,是推荐算法中一个亟需解决的问题。研究学者为此提出一些解决方案:Mazumder 等^[21]通过用户评分均值的方式对初始评分矩阵进行填充,以保证用户相似计算的精度;Sarwar 等^[1]在推荐算法早期提出一种奇异值分解(Singular Value Decomposition, SVD)方法,来分析用户项目矩阵之间的联系,以此填补用户评分;吴宾等^[22]在联合正则化的 SVD 基础上,提出一种矩阵分解模型,并以此推荐,该模型缓解了用户冷启动,同时也得到了有效的用户评分填充。

大多算法通过矩阵分解的方式对评分矩阵填补,但进行矩阵分解需要进行梯度下降^[1,21-22]等方式来获取最优解,这就给推荐计算带来时间和性能的消耗。

因此,本文考虑到实际日常中,每位用户对项目的真实评价都会有自己的一套标准,且每一套标准在一定合理范围内,结合高斯函数(Gaussian Function)的正态分布性质及时间复杂度较低的特点,提出一种对评分矩阵填补的新尝试,即通过结合高斯函数的计算模型对用户历史评分矩阵的空白区域进行填补。其高斯函数表达式为:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (3)$$

其中, $x \in (-\infty, +\infty)$; σ 为高斯半径,且 $\sigma \in [0.1, 250]$; μ 表示均值。在进行预填充时,假设评分矩阵为 $ScoMat_{m \times i}$, 含义为用户 m 对项目 i 的评分矩阵;用户评分集合为 $C_{m,i}$, 评分均值为 $\overline{C_{m,i}}$, 取 σ 为 1, 则符合对 $ScoMat_{m \times i}$ 矩阵填充的高斯公式为:

$$ScoMat_{m \times n} = \frac{1}{\sqrt{2\pi}} e^{-\frac{(C_{m,n} - \overline{C_{m,n}})^2}{2}} \quad (4)$$

假设项目数量为 1000, 用户评分为 0~5, 含 0.5 分制, 由式(4)进行数据填补, 其正态曲线如图 2 所示。

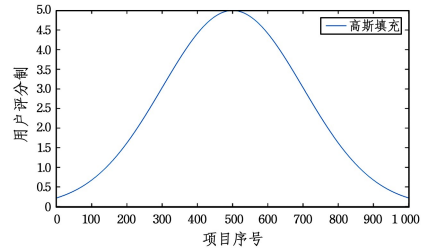


图2 高斯填充正态曲线示意图

由于高斯填充会覆盖原本的评分数据, 因此本文对式(4)添加了判断条件:

$$ScoMat_{m \times i} = \begin{cases} C_{m,i}, & ScoMat(m,i) \in \text{Null} \\ \frac{1}{\sqrt{2\pi}} e^{-\frac{(C_{m,i} - \overline{C_{m,i}})^2}{2}}, & ScoMat(m,i) \notin \text{Null} \end{cases} \quad (5)$$

其中, $ScoMat(m,i)$ 表示在用户评分矩阵中用户 m 对项目 i 的评分值。通过式(5), 对用户评分矩阵中的空缺数据进行补充, 可以较好地缓解数据稀疏。

3.2 间接信任计算模型

信任在现实社会中越来越重要, 体现着一种良好的品质, 推荐算法中融入信任可以提升算法的准确率, 故此度量用户信任也成为目前推荐算法研究的一个热点。通常, 显式信任由用户之间的关注、点赞即可获得, 一般为二值信任, 取值为 0 和 1。但关于间接信任的计算方式各有不同: 由于信任存在传递性, 间接信任的计算利用其特点进行展开, 其信任传递示意图如图 3 所示。

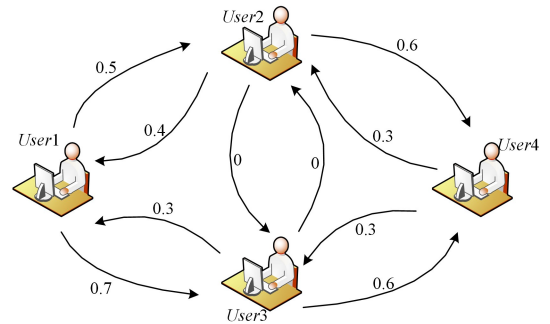


图3 信任网络传递示意图

间接信任的计算采用下式:

$$IndTrust(m,n) = Tru_{m,k} \times Tru_{k,n} \quad (6)$$

该计算方式为用户 m 与用户 n 的间接信任等于由用户 m 至 k 路径信任值乘以用户 k 到 n 的路径信任, 如计算图 3 中 $User1$ 对 $User4$ 的间接信任, 即: $IndTrust(User1, User4) = 0.5 \times 0.6 = 0.3$ 。但通过图 3 可知, $User1$ 对 $User4$ 的间接信任不仅包含路径 $User1 \rightarrow User2 \rightarrow User4$, 还包括 $User1 \rightarrow User2 \rightarrow User3 \rightarrow User4$ 。可见式(6)并未完全考虑到这种情况。

另外, 傅敏^[23]提出的信任递归计算模型为:

$$IndTrust(m,n) = \sum_{v \in R} T_{m,v} \times T_{v,n} \quad (7)$$

其中, R 表示信任用户集合, $T_{m,v}$ 表示用户 m 到最短路线节点 v 的信任值, $T_{m,v}$ 表示最短路线节点 v 到用户 n 的信任值。

式(7)考虑使用最短路线的形式限制多条信任传递路径,从而只需计算特定路径。然而,此模型只计算特殊路径,忽略了多条路径给间接信任带来的影响,并设置阈值来去除小部分的信任用户,这样多次调节阈值会带来误差,也会使计算复杂,影响性能。

综合以上分析,本文整合信任传递性,以传递路径为依据,利用信息熵思想评估用户信任质量,同时将其作为惩罚权重的平衡因子来计算各传递路径的间接信任,其计算模型为:

$$IpmEn(m,n) = \sum_{v \in R_{m,n}} \frac{\sum_{i \in m}^{k=n} Brch_{i,k} \times Tru_{i,k}}{Path_{m,v}} \quad (8)$$

式(8)与式(7)的不同之处在于:式(8)需要计算用户 m 到用户 n 所有传递路径的信任值,而式(7)只是计算所有路径中最短路径,虽然式(8)遍历时间更长,但在综合用户信任准确率方面,要比式(7)好。在式(8)中, $R_{m,n}$ 为用户 m 对用户 v 信任传递经过的用户集合; $Path_{m,v}$ 表示 m 至 v 传递的总路径长, $Brth_{m,v}$ 表示 m 至 v 传递的各分支路径长, $Tru_{i,k}$ 为用户 i 对用户 k 的显式信任。在本文中,显式信任为二值信任,每条路径长度为 1,为了以下描述方便,这里将式(8)间接信任的计算表达式简称为 IpmEn 模型。

比如,利用 IpmEn 模型来计算图 3 中 $User1$ 对 $User4$ 的间接信任: $User1$ 与 $User4$ 之间共含有 4 条信任传递路径,故其计算过程如下:

$$\begin{aligned} IpmTrust(User1,User4) &= \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} + \frac{1}{3} \times \frac{2}{3} \times \frac{3}{3} + \frac{1}{3} \times \frac{2}{3} \times \frac{3}{3} \\ &= \frac{1}{2} + \frac{4}{9} = \frac{17}{18} \end{aligned}$$

该 IpmEn 模型可以很全面地计算间接信任,但存在一个问题:用户在社交上网时并非实际信任用户,而有可能出于从众心理,比如拥有大批粉丝的明星;还有可能是被雇佣刷关注的用户等。因此,本文对模型进行一定的权重惩罚,避免信任数值伪偏高现象。

为了评估用户信任质量的好坏,采用信息熵^[24]来分析用户行为,其信息熵的表达式为:

$$H(x) = - \sum_{x \in X} p(x) \log p(x) \quad (9)$$

其中, X 为随机变量。信息熵值越高,表明用户信任质量越高;相反,表明用户信任质量越低,有随众等现象的可能性。本文结合信息熵思想,定义信任用户 m 的信任概率 $p(x)$ 的取值为:该用户显式信任的总次数与所有用户个数的所占比例,其表达式为:

$$p(m) = \frac{|\sum_{ValTrs_i \in U_m} f(ValTrs_i = 1)|}{|U_m|} \quad (10)$$

其中, U_m 表示用户 m 的历史数据中的总用户集合; $ValTrs_i$ 为在 U_m 集合中信任用户所对应的二值信任(包含 0 和 1); $f(ValTrs_i = 1)$ 函数当满足 $ValTrs_i = 1$ 时,返回值为 1,相反为 0。故此,其用户信任信息熵的表达式为:

$$H(m) = - \sum_{m \in U_m} p(m) \log p(m) \quad (11)$$

建立信任信息熵的表达式后,本文对 IpmEn 模型进行一定的权重惩罚,所得完善后的间接信任计算模型 IpmTrust 为:

$$pre(f,m) = \bar{C} + \frac{\sum_{n \in N_{f,m}} ((1-\alpha)SinMat(m,n) + \alpha TruMat(m,n)) \times (C_{f,m} - \bar{C})}{\sum_{n \in N_{f,m}} (1-\alpha)SinMat(m,n) + \alpha TruMat(m,n)} \quad (16)$$

$$IpmTrust(m,n) = |H(m)| \times \sum_{v \in R_{m,n}} \frac{\sum_{i \in m}^{k=n} Brch_{i,k} \times Tru_{i,k}}{Path_{m,v}} \quad (12)$$

增加一定权重后的 IpmTrust 模型可以提升计算用户间接信任的精度,同时省去手动调节阈值的麻烦。

3.3 构建信任矩阵及相似度计算

3.3.1 构建信任矩阵

由 2.2 小节得出的间接信任计算模型 IpmTrust,通过综合显式信任,搭建用户信任矩阵,为后面的预测评分的改进提供额外的数据库,其建立形式如式(13)所示:

$$TruMat_{m \times n} = \begin{cases} IpmTrust(m,n), & TruMat(m,n) \in Null \\ 1, & TruMat(m,n) \notin Null \end{cases} \quad (13)$$

在式(13)中, $TruMat_{m \times n}$ 为用户信任矩阵, $TruMat(m,n)$ 表示用户信任矩阵中用户 m 对用户 n 的信任值,未进行式(13)操作的信任矩阵如表 2 所列。

表 2 $TruMat_{m \times n}$

	User1	User2	User3	...	Usern
User1	1	0	1	...	1
User2	0	0	1	...	0
...
Userm	1	1	0	...	0

结合式(12)与式(13), $TruMat_{m \times n}$ 进行了显式与间接信任的综合计算,如表 3 所列。

表 3 $TruMat_{m \times n}$ 综合计算结果

	User1	User2	User3	...	Usern
User1	1	0.23	1	...	1
User2	0.30	0.35	1	...	0.12
...
Userm	1	1	0.52	...	0.45

3.3.2 相似度计算

本文将由式(5)高斯预填充后的用户评分矩阵,通过修正的余弦相似度式(1)计算用户相似性 $SinMat(m,n)$,并构建用户相似度矩阵 $SinMat_{m \times n}$ 。

3.4 评分预测模型的改进

传统的评分预测模型如式(2)所示,通常在设计推荐算法时,会考虑到用户评分存在一定个人倾向,为了减少这类误差,评分预测模型会在传统模型上加入评分均值进行权重惩罚,即现有评分预测模型^[25]为:

$$pre(f,m) = \bar{C} + \frac{\sum_{n \in N_{f,m}} \sin(m,n) \times (C_{f,m} - \bar{C})}{\sum_{n \in N_{f,m}} \sin(m,n)} \quad (14)$$

其中, \bar{C} 为所有用户评分的均值。但式(14)只考虑到用户评分相似度,不太适合信任数据。因此,本文通过设置参数 α ,对用户相似度与用户信任矩阵进行线性加权融合,融合后的相似度表达式为:

$$SinMat(m,n) = \alpha \times SinMat(m,n) + (1-\alpha) \times TruMat(m,n) \quad (15)$$

式(15)既考虑到用户评分,又容纳了用户信任,则在式(14)的基础上改进评分预测:

其中, $\alpha \in (0, 1)$ 。由式(16)进行推荐算法预测, 同时可调节 α , 以获得较好的推荐效果。

4 实验结果与分析

4.1 实验环境与数据集

本文使用 MATLAB-R2016a 对设计的新算法进行实验, 并使用公认的电影评分数据集 FilmTrust^{[26]1)} 对其进行验证。FilmTrust 数据集包含 35497 条历史记录, 1508 个用户, 2070 部电影, 评分制为 0.5~4.5 分, 以及 1853 条信任数据, 信任值为二值信任, “1”表示信任, “0”表示不信任。为了保障新算法的有效性, 文中数据集将被随机划分成 80% 的训练集和 20% 的测试集, 并进行 10 次交叉验证, 以达到实验验证效果。

4.2 评价指标

关于评价指标, 本文使用推荐算法中常用的 RMSE、MAE。假设推荐算法预测用户的评分集为 $p_i = \{p_1, p_2, \dots, p_n, i \in [1, n]\}$, 用户真实评分集为 $q_i = \{q_1, q_2, \dots, q_n, i \in [1, n]\}$, 则两评价指标的定义表达式为:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n |p_i - q_i|^2}{n}} \quad (17)$$

$$MAE = \frac{\sum_{i=1}^n |p_i - q_i|}{n} \quad (18)$$

4.3 结果及分析

出于实验描述方便, 将用户间接信任及高斯预填充的推荐算法 (Recommendation algorithm considering user indirect trust and Gaussian pre-filling) 记为 GITCF; 传统的协同过滤推荐算法 (Collaborative filtering recommendation)^[1] 记为 CF; Salakhutdinov 等^[27] 提出的基于概率矩阵分解的推荐算法记为 PMF; 利用信任传递性质的推荐算法^[28] 记为 SocialMF; 陈婷等^[29] 在矩阵分解基础上结合信任的推荐算法记为 Trust-PMF; 采用动态信任的推荐算法^[30] 记为 Proposed-CF。将传统的评分预测模型 (Traditional Scoring Prediction Model) 记为 TSPM, 现有的评分预测记为 MSPM, 改进的评分预测记为 ISPM。

为了验证 GITCF 算法的有效性, 本文进行了两类实验: 第一类实验主要为参数 α 、迭代次数对 GITCF 算法的影响, 以及采用不同评分预测与改进的评分预测模型对 GITCF 算法的影响; 第二类实验主要为 CF、PMF、SocialMF、Trust-PMF、Proposed-CF 算法与 GITCF 算法之间的准确度对比。

4.3.1 自身参数实验

自身参数实验共分为 3 组: 第 1 组设置不同参数 α , 获取其合适值, 以达到预测评分最佳效果; 第 2 组将改进后的评分预测模型与传统预测及现有预测进行对比; 第 3 组设置不同迭代次数, 取合适的迭代数, 以达到最佳训练和测试效果。

第 1 组实验: 分别设置 5 组不同参数 α , 每组参数值间隔 0.2, 即 0.0、0.2、0.4、0.6、0.8; 邻居个数选择范围为 5~60。由 GITCF 算法在 FilmTrust 数据集上进行实验, 其实验效果如图 4 所示。

由图 4 可知, GITCF 在不同 α 调节下, 实验结果也不同。一般来说, RMSE、MAE 值越低, 推荐准确率越高, 当图 4 邻

居个数为 25~30 时, 各 α 所处的 RMSE、MAE 值均最小, 其中当 $\alpha=0.4$ 时, 整个 GITCF 算法推荐过程准确度高, 故本文选取 $\alpha=0.4$ 为合适值。

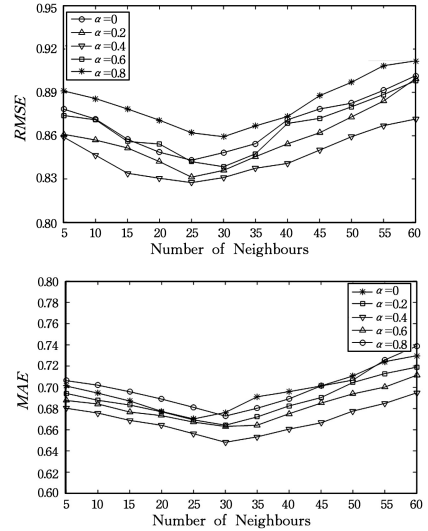


图 4 不同 α 值的实验效果

另外, 当 $\alpha=0.0$ 时, 预测评分模型只引入信任数据, 没有调节用户相似度与信任之间的关系。故此, 为了验证引入辅助数据对评分预测模型是否有效, 进行第二组实验。

第二组实验环境、数据集均与第一组的保持不变, 当 $\alpha=0.0$ 时, 式(16)的 ISPM、式(14)MSPM 与式(2)TSPM 之间评分预测模型效果的比较结果如 5 所示。

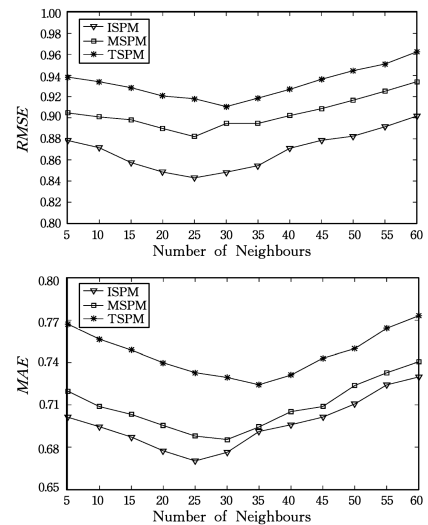


图 5 评分预测模型的对比实验效果

通过图 5, 在 RMSE 和 MAE 指标方面, 改进后的评分预测模型考虑了用户信任数据, 在推荐准确率上均要比单独考虑用户评分数据的评分预测模型好, 这主要是因为引入间接信任不仅缓解了数据稀疏, 而且提升了计算相似用户的精度, 给最终的推荐结果带来正反馈。

第三组实验: 设置迭代次数为 10~120, 间隔 10 次。GITCF 算法的效果如图 6 所示。

通过图 6 可知, 随着迭代次数增加, GITCF 算法的 RMSE、MAE 值下降, 准确率提升, 当迭代次数在 70 时, 推荐效果最佳。

¹⁾ <https://www.librec.net/datasets.html>

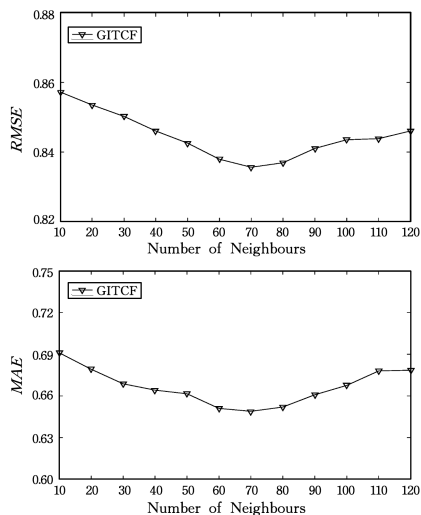


图 6 迭代次数的实验效果

4.3.2 算法对比实验

本文的算法对比实验主要将 GITCF (参数设置 $\alpha=0.4$, 迭代次数为 70, 采用改进的评分预测模型) 与 CF, PMF, SocialMF, Trust-PMF, Proposed-CF 算法, 在数据集、实验环境相同的条件下进行比较, 其实验效果如表 4 所列。

表 4 各算法在最佳 RMSE, MAE 的对比

算法	CF	Proposed-CF	PMF	SocailMF	Trust-PMF	GITCF
RMSE	1.1025	1.0532	1.0618	0.8535	0.8489	0.8375
MAE	0.8320	0.7858	0.7700	0.6708	0.6556	0.6482

结合表 4, 在 RMSE、MAE 不同指标下, 算法 Proposed-CF 在准确率方面要比传统推荐算法 CF 要好, 这是由于 Proposed-CF 将信任引入到推荐中, 并利用信任的动态性不定时更新数据, 辅助数据的加入缓解了数据稀疏, 从而提高了推荐精度; PMF 算法要好于 CF 算法, 主要在于 PMF 从数据稀疏的角度出发, 对用户项目特征矩阵采用矩阵分解, 由梯度下降法等方式进行求解, 来弥补用户项目矩阵中空缺的数据, 而 CF 算法只是通初始矩阵来计算推荐, 故准确率低; 同样, SocialMF 算法重点则是利用信任的传递性质来推荐, 推荐效果要好于以上三类算法; Trust-PMF 算法综合了矩阵分解和信任网络, 通过实验对比, 相比 CF, PMF, SocialMF, Proposed-CF 算法, 推荐效果更显著。故此, 在推荐算法中引入信任可以改善推荐质量。

然而, 本文中的 GITCF 算法相比其他五类算法, RMSE、MAE 指标均偏低, 效果更好, 推荐准确性更高。其主要原因是 GITCF 一方面在预处理阶段, 对用户评分矩阵进行高斯填充, 减少大量空白区域, 给后续计算相似用户的精度带来保障; 另一方面, GITCF 考虑到了全局路径下的间接信任, 并设计了间接信任的计算模型 IpmTrust, 由于间接信任要比显式信任隐含的数据量多, 将其融入到推荐中, 可以更好地缓解数据稀疏性。此外, 通过表 4 可知, 与同样考虑信任传递性的 SocialMF 算法相比, GITCF 在两指标上的推荐性高, 这是因为 SocialMF 采用信任传递性计算用户特征, 忽略了多条路径下信任用户的相互影响, 而 GITCF 则考虑每条路径下的可能性, 进一步挖掘用户潜在价值。最后, 结合以上实验, 验证了 GITCF 算法的有效性。

结束语 本文主要考虑到全局路径下, 用户间接信任富

含大量的数据源, 通过设计一种间接信任的计算模型 IpmTrust 来挖掘其潜在的价值, 并采用高斯函数对评分矩阵进行预处理, 以此确保计算相似用户的精度, 再结合 IpmTrust 设计一种考虑用户间接信任的推荐算法 GITCF。最后通过实验分析, 其推荐质量表现良好。但 GITCF 算法仍有继续深入的空间, 比如, 在邻居个数增加、新用户大量引进的情况下, 若推荐内容过于老旧、没有新颖性, 则推荐结果会持续下降。对此类问题, 本文的进一步研究工作将考虑结合线上及时更新等方式, 进行内容更替。

参考文献

- [1] SARWAR B, KARYPIS G, KONSTAN J, et al. Item-based collaborative filtering recommendation algorithms[C] // International Conference on World Wide Web. 2001.
- [2] 邓爱林, 朱扬勇, 施伯乐. 基于项目评分预测的协同过滤推荐算法[J]. 软件学报, 2003(09): 1621-1628.
- [3] YU C, XU J, DU X. Recommendation algorithm combining the user-based classified regression and the item-based filtering[C] // International Conference on Electronic Commerce: the New E-commerce: Innovations for Conquering Current Barriers. 2006.
- [4] LIM K H, CHAN J, LECKIE C, et al. Personalized tour recommendation based on user interests and points of interest visit durations[C] // International Conference on Artificial Intelligence. 2015.
- [5] 余永红, 高阳, 王皓, 等. 融合用户社会地位和矩阵分解的推荐算法[J]. 计算机研究与发展, 2018, 55: 113-124.
- [6] 焦旭, 肖迎元, 郑文广, 等. 基于位置的社会化网络推荐技术研究进展[J]. 计算机研究与发展, 2018, 55(10): 2291-2306.
- [7] ZHANG F, LONG B, FENG G. A User Trust-Based Collaborative Filtering Recommendation Algorithm[C] // International Conference on Information & Communications Security. 2009.
- [8] 王珂, 高琳. 基于社交圈的在线社交网络朋友推荐算法[J]. 计算机学报, 2014, 37(4): 801-808.
- [9] KALAI A, ZAYANI C A, AMOUS I, et al. Social collaborative service recommendation approach based on user's trust and domain specific expertise[J]. Future Generation Computer Systems, 2018(80): 355-367.
- [10] 游静, 上官经伦, 徐守坤, 等. 考虑信任可靠度的分布式动态信任管理模型[J]. 软件学报, 2017, 28(9): 2354-2369.
- [11] MA X, LU H W, GAN Z B, et al. An explicit trust and distrust clustering based collaborative filtering recommendation approach [J]. Electronic Commerce Research and Applications, 2017 (25): 29-39.
- [12] CHEN H L. A Personalized Recommendation Algorithm Based on the Fusion of Trust Relation and Time Series[C] // IEEE International Conference on Computational Science & Engineering. IEEE, 2017.
- [13] 朱敬华, 明睿. LBSN 中融合信任与不信任关系的兴趣点推荐 [J]. 通信学报, 2018, 39(7): 157-165.
- [14] 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.
- [15] 潘一腾, 何发智, 于海平. 一种基于信任关系隐含相似度的社会化推荐算法[J]. 计算机学报, 2018, 41(1): 65-81.
- [16] XU X, YUAN D. A novel matrix factorization recommendation algorithm fusing social Trust and Behaviors in micro-blogs[C] //

- IEEE International Conference on Cloud Computing & Big Data Analysis. IEEE, 2017.
- [17] LINGAM G, RANJAN R R, DVLN S. Learning automata-based trust model for user recommendations in online social networks [J]. Computers and Electrical Engineering, 2018, 66.
- [18] AZADJALAL M, MORADI P, ABDOLLAHPOURI A, et al. A trust-aware recommendation method based on Pareto dominance and confidence concepts [J]. Knowledge-Based Systems, 2017 (116): 130-143.
- [19] YIN C Y, WANG J, PARK J H. An Improved Recommendation Algorithm for Big data Cloud Service based on the Trust in Sociology [J]. Neurocomputing, 2017 (256): 49-55.
- [20] GOHARI F S, ALIEE F S, HAGHIGHI H A. A new confidence-based recommendation approach: Combining trust and certainty [J]. Information Sciences, 2018 (422): 21-50.
- [21] MAZUMDER R, HASTIE T, TIBSHIRANI R. Spectral Regularization Algorithms for Learning Large Incomplete Matrices [J]. Journal of Machine Learning Research, 2010, 11(11): 2287-2322.
- [22] 吴宾, 娄铮铮, 叶阳东. 联合正则化的矩阵分解推荐算法 [J]. 软件学报, 2018, 29(9): 2681-2696.
- [23] 傅敏. 基于信任和不信任的协同过滤推荐模型研究 [D]. 燕山: 燕山大学, 2012.
- [24] 王国胤, 于洪, 杨大春. 基于条件信息熵的决策表约简 [J]. 计算机学报, 2002(7): 759-766.
- [25] 毛一凡, 饶世钧. 对修正的 K 近邻域关联算法的仿真与可信性评估 [J]. 计算机仿真, 2004, 21(7): 11-13.
- [26] GOLBECK J. Filmtrust: movie recommendations from semantic web-based social networks [C] // Consumer Communications & Networking Conference. IEEE, 2006.
- [27] SALAKHUTDINOV R, MNIH A. Probabilistic Matrix Factorization [C] // International Conference on Neural Information Processing Systems. Vancouver, Canada, 2007.
- [28] JAMALI M, ESTER M. A matrix factorization technique with trust propagation for recommendation in social networks [C] // Proceedings of the 2010 ACM Conference on Recommender Systems (RecSys 2010). Barcelona, Spain: ACM, 2010: 26-30.
- [29] 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.
- [30] ZHANG Z, LIU Y, JIN Z, et al. A Dynamic Trust based two-layer Neighbor Selection Scheme towards Online Recommender Systems [J]. Neurocomputing, 2018, 285: 94-103.

(上接第 158 页)

- [46] ZHAO Y, XU C, CAO Y. Research on query-by-committee method of active learning and application [C] // International Conference on Advanced Data Mining and Applications. Springer-Verlag, 2006: 985-991.
- [47] RASHID A M, KARYPIS G, RIEDL J. Learning preferences of new users in recommender systems: an information theoretic approach [J]. ACM Sigkdd Explorations Newsletter, 2008, 10(2): 90-100.
- [48] CARENINI G, SMITH J, POOLE D. Towards more conversational and collaborative recommender systems [C] // International Conference on Intelligent User Interfaces. ACM, 2003: 12-18.
- [49] PENNOCK D M, HORVITZ E, LAWRENCE S, et al. Collaborative Filtering by Personality Diagnosis: A Hybrid Memory-and Model-Based Approach [C] // Proceedings of IJCAI-99. 2013: 473-480.
- [50] SETTLES B, CRAVEN M. An analysis of active learning strategies for sequence labeling tasks [C] // Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2008: 1070-1079.
- [51] IKHWANTRI F, LOUVAN S, KURNIAWAN K, et al. Multi-Task Active Learning for Neural Semantic Role Labeling on Low Resource Conversational Corpus [C] // Proceedings of the 2007 ACM Conference on Recommender Systems. 2018: 37-49.
- [52] BRIDGE D, RICCI F. Supporting product selection with query editing recommendations [C] // ACM Conference on Recommender Systems. ACM, 2007: 65-72.
- [53] RICCI F, NGUYEN Q N. Acquiring and Revising Preferences in a Critique-Based Mobile Recommender System [J]. IEEE Intelligent Systems, 2007, 22(3): 22-29.
- [54] MCNEE S M, LAM S K, KONSTAN J A, et al. Interfaces for eliciting new user preferences in recommender systems [C] // International Conference on User Modeling. Springer-Verlag, 2003: 178-187.
- [55] GOLBANDI N, KOREN Y, LEMPEL R. Adaptive bootstrapping of recommender systems using decision trees [C] // ACM International Conference on Web Search and Data Mining. ACM, 2011: 595-604.
- [56] KARIMI R, WISTUBA M, NANOPOULOS A, et al. Factorized Decision Trees for Active Learning in Recommender Systems [C] // IEEE, International Conference on TOOLS with Artificial Intelligence. IEEE Computer Society, 2013: 404-411.
- [57] HERLOCKER J L, KONSTAN J A, TERVEEN L G, et al. Evaluating collaborative filtering recommender systems [J]. ACM Transactions on Information Systems, 2004, 22(1): 5-53.
- [58] HARPALE A S, YANG Y. Personalized active learning for collaborative filtering [C] // International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2008: 91-98.
- [59] SHANI G, GUNAWARDANA A. Evaluating Recommendation Systems [M] // Recommender Systems Handbook. Springer, 2011: 257-297.
- [60] LIU N N, YANG Q. EigenRank: a ranking-oriented approach to collaborative filtering [C] // International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2008: 83-90.