

基于社交网络信任关系的服务推荐方法

王佳蕾 郭耀 刘志宏

(西安电子科技大学网络与信息安全学院 西安 710071)

摘要 随着服务型计算的兴起,大量跨领域电子服务应运而生。用户要从众多服务中挑选出适合自己且可信的服务十分困难,因而提出高效的服务推荐算法十分必要。传统的协同推荐方法存在冷启动、数据稀疏以及实时性不好等问题,在评分数据较少时推荐效果不佳。为获得更好的推荐结果,文中在社交网络中使用信任传递机制,建立信任传递模型,由此获取任意用户间的信任度。另一方面,设计了相似性判定指标,凭借系统评分数据,求得用户间的偏好相似度。在得到用户间信任度和偏好相似度的基础上,根据社交网络的特性,动态结合两部分指标以获得综合推荐权重,再以此权重替代传统相似度衡量标准进行基于用户的协同过滤推荐。所提方法能在解决传统推荐算法问题的基础上进一步提升推荐效果,并以准确率、覆盖率为标准在 Epinions 数据集上进行验证,获得了较好的效果。

关键词 推荐系统,信任网络,协同过滤,冷启动

中图分类号 TP393 **文献标识码** A

Service Recommendation Method Based on Social Network Trust Relationships

WANG Jia-lei GUO Yao LIU Zhi-hong

(School of Cyber Engineering, Xidian University, Xi'an 710071, China)

Abstract With the advent of service computing, many different electronic services have emerged. Users often have to find what they need from a large number of services, which is a formidable task. Hence, it is necessary to put forward an efficient recommendation algorithm. The traditional cooperative recommendation system has some problems, such as cold start, sparsity of data and poor real-time performance, which lead to poor recommendation results under the circumstances with less scoring data. In order to get a better recommendation result, this paper introduced trust transfer in social networks and utilized it to establish a trust transfer model to obtain trust among users. On the other hand, based on the score data, the similarity between users in the system is calculated. On the basis of similarity between users' trust and preference, according to the characteristics of social networks, users' trust and preference are dynamically combined to obtain comprehensive recommendation weights. The comprehensive recommendation weights can replace the traditional similarity measurement standards for user-based collaborative filtering recommendation. This method was verified through the Epinions data set and can further improve the recommendation effect and.

Keywords Recommendation system, Trust network, Collaborative filtering, Cold start

1 引言

在互联网时代,线上资源信息的数据量急速膨胀。海量信息及资源为用户提供丰富服务的同时,也为用户便利地选择符合自身需求的服务带来了挑战。针对数据超载问题,做到安全、快速、准确地为用户进行服务推荐已经成为目前网络服务研究的热点之一。

传统推荐系统多是基于协同过滤算法、内容数据过滤算法以及混合推荐算法等^[1]进行实现的。其中,协同过滤算法较为直观、准确,也是目前使用较为广泛的一种算法。其特点是只需处理用户对项目的评分数据,不需深入考虑用户及项目特性就可进行推荐。然而,协同过滤算法存在数据稀疏、可用性差以及冷启动等问题。为提高服务推荐质量并解决传统

算法出现的问题,国内外研究人员做了大量的研究工作。Liu等^[2]通过使用社交亲密度和个体用户影响力建立模型,设定多个端对端的性能属性约束,从而寻找复杂社交网络中的最佳信任路径,得到最佳推荐。Gartrell等^[3]通过一个共识函数为每个组员评分,根据组内的属性及关系来获取组预测评分,以建立推荐系统。Tang等^[4]提出了结合全局信任和局部信任关系从而给出信任值的信任度衡量标准。Zhao等^[5]提出通过加入社交网络连接的SBPR(社交贝叶斯个性化排名)模型来缓解冷启动等问题。Golbeck等^[6]提出了采用扩展广度优先搜索算法获得推荐者列表的TidalTrust算法,以最短路径距离寻找推荐者,并以活跃用户与推荐者之间的信任值为权重预测目标项目评分。Hang^[7]通过准确惩罚或奖励服务来提供服务质量,从而建立信任模型。Jamali等^[8]提出了

本文受国家自然科学基金(U1405255)资助。

王佳蕾(1992—),女,硕士生,主要研究方向为信任管理、推荐系统、物理层安全,E-mail:wangjialei92@163.com;郭耀(1994—),女,硕士生,主要研究方向为信任管理、推荐系统、物理层安全,E-mail:958487621@qq.com;刘志宏(1968—),男,博士,副教授,主要研究方向为密码学、信息安全、网络编码、复杂网络、传感器网络等,E-mail:liuzhihong@mail.xidian.edu.cn(通信作者)。

TrustWalker 方法,其以随机漫游搜索信任网络,利用项目间的相似性来降低搜索算法的时间复杂度。Ziegler 等^[9]在语义网络上使用本地组信任管理的方式,结合用户信任与不信任关系获取社交网络的信任关系。Tang 等^[10]分析了在线网络上的信任演化过程,并提出 eTrust 模型。Deng 等^[11]提出了一种分为两个阶段的推荐流程,在初始化过程中利用深度学习来确定矩阵分解中的初始化社交信任度,将用户的兴趣及其可信赖的朋友的兴趣与社交效应的影响一起用于推荐。Yang 等^[12]提出了集成用户提供的稀疏评级数据和社交信任网络来提高协作过滤性能的方法。采用矩阵分解技术,根据用户的信任关系将用户映射到低维潜在特征空间,旨在更准确地反映用户对自己意见形成的相互影响,并更好地学习以为用户提供优质模式的高质量建议。Xiong 等^[13]提出了一种新的基于信任的 Top-k 的推荐算法 BTRank,其集成了评分和信任信息,构建了评分排序模型,有效地提高了所有用户的 Top-k 项目列表的质量。Guo 等^[14]提出了一种基于信任的在线社交网络隐私保护推荐方案,其中在线社交网络用户通过多跳信任链,应用其属性来寻找匹配朋友,并与陌生人建立社交关系。

通过用户社交网络的信任传播能够有效解决数据稀疏和冷启动问题,但目前流行的算法在推荐的准确率和可行性上还有待提高。本文提出了一种基于信任和偏好的协同过滤的算法——TPCF (Collaborative Filtering Algorithm Based on Trust and Preferences)算法。TPCF 算法的基本思想是:在局部网络中使用传播激活模型建立信任网络,根据各用户间的信任关系获得用户间的信任值;提出偏好相似度、产品体验相似度及兴趣相似度的概念,设计偏好相似度计算指标并获得用户偏好相似度;动态结合信任度和偏好相似度权重得到用户推荐权值,在已知用户评分及推荐权值的情况下,使用最近邻协同过滤算法计算出预测评分,并给出相应的推荐列表。

本文第 2 节介绍信任的定义、传递、动态激活模型;第 3 节介绍本文核心即 TPCF 算法,主要涉及信任模型建立算法、用户偏好相似度算法、基于用户推荐权值的协同过滤算法;第 4 节为实验部分;最后总结全文。

2 基于社交网络的信任模型

2.1 信任评分场景

推荐场景可以定义为社交网络下的服务推荐 SOE-SN (Service-Oriented Environments with Social Network)^[4]。在 SOE-SN 中存在用户间社交网络 G 以及用户对服务的评分网络 $Rating$ 两部分。定义 $G(V, E, W)$ 表示社交网络,其中 V 表示网络中用户组成的集合, E 表示用户间信任关系的集合, W 表示信任边上的权值。评价网络 $Rating(V', S, R)$ 中, V' 表示所有参与评价服务的用户集合,且 $V' \subseteq V$; S 表示在服务推荐环境下所有可能获得的不同服务集合; R 表示用户对服务的评价价值集合。

信任模型如图 1 所示。用户间的有向边表示信任关系,边上的权值 $w(a, b)$ 表示用户 a 与用户 b 间信任程度的强弱,是一个介于 0 到 1 之间的实数。用户与服务间的有向边表示用户对服务的评价,边上的权值代表用户对服务的评价分数,本文定义用户的评价分数为 $R(a, b)$ 。其中 $R(a, b) \in \{1, 2, 3, 4, 5\}$ 。

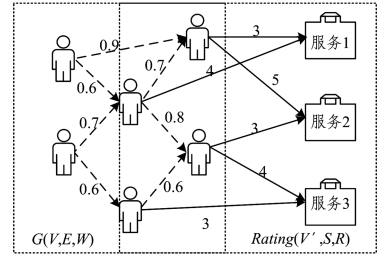


图 1 信任评分场景

2.2 信任特性

在信任传播和信任聚合的过程中,需综合考虑信任的以下特性。

1) 信任关系的传递性。如图 2 所示,用户 a 直接信任用户 b , 用户 b 直接信任用户 c , 于是 a 间接信任用户 c , 即 $\exists w(a, b) \wedge w(b, c) \Rightarrow w(a, c)$ 。用户 a 对用户 c 的信任程度可以根据网络的连接情况和传播激活模型获得。

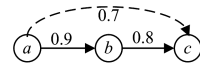


图 2 信任传递示例

2) 信任关系的不对称性。网络中有很大概率出现 $w(a, b) \neq w(b, a)$, 整个用户信任网络是加权有向网络。

3) 信任关系的衰减性。随着链路增长及信任程度的逐渐衰减,无直接联系的用户间间隔增多,两者间的信任程度降低。

4) 信任关系的多路性。在实际信任网络中,两个节点间的信任传递路径通常不止一条。关系链路越多,两个用户间的信任度就越大。

2.3 偏好相似度

文中偏好相似度被定义为为用户间评分服务类型和评分值的综合相似程度。用户会对感兴趣的服务进行评价,而评分结果高也代表用户对此类项目的偏好程度大。因此,可通过用户对服务的评价情况获得用户之间的偏好相似度。本文使用 $sim(a, b)$ 表示用户 a 和用户 b 之间偏好的相似度,其值是介于 0 到 1 之间的实数。

3 基于信任关系与用户偏好的服务推荐算法

基于信任关系与用户偏好的服务推荐算法 TPCF 的框架如图 3 所示。给定评分矩阵和信任矩阵,算法的具体步骤如下:

- 1) 由评分矩阵得到任意两个用户之间偏好的相似度;
- 2) 由用户信任关系矩阵构建信任网络,建立信任传播模型,获得信任度矩阵;
- 3) 动态结合信任度和偏好相似度,得到推荐权值;
- 4) 采用协同过滤推荐,得到最终评分预测并获取推荐列表。

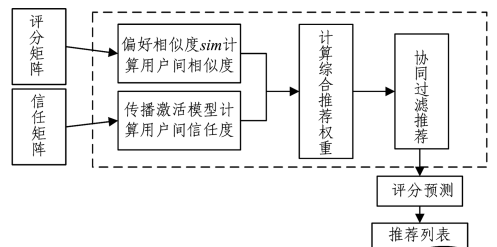


图 3 TPCF 框架

3.1 信任传递模型

使用传播激活模型^[15]计算信任网络的传递过程和信任值。在传播激活模型中,如图4所示,源节点 a 注入的总信任能量为 e ,能量沿着有向边传播到后续节点。能量可以按照边上的权重进行分流,权重大的边所传播能量的比例大。整个信任传播过程中网络的能量 in^0 保持守恒。

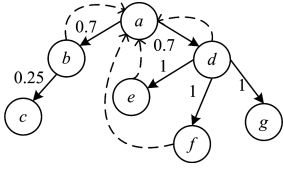


图4 能量的传播过程

每个终叶节点都有一个回溯到源点 a 的边且权值为1,因此能够保证能量在整个网络中守恒且不进入死循环。

流入后继节点的能量由权重 $w(x,y)$ 决定,文中采用平方归一化的方法标准化边的权重。使用 $e_{x \rightarrow y}$ 表示通过用户 x 流向后继节点 y 的能量,公式如下:

$$e_{x \rightarrow y} = d * in(x) * \frac{w(x,y)^2}{\sum_{(x,s) \in E} w(x,s)^2} \quad (1)$$

其中, $in(x)$ 表示流入 x 的信任能量, $trust(x)$ 源点对 x 累积的信任度, d 表示传播因子。当能量流入节点 x 时,有 $d * in(x)$ 的能量流出到后继节点,剩余能量留在当前节点。传播因子 d 使能量不会陷入死循环而导致不守恒,并且保证距离源点近的节点被信任的程度大。如图4所示,其中 $e_{d \rightarrow e} = d * in(d) * \frac{1^2}{1^2 + 1^2 + 1^2 + 1^2}$,而 $e_{b \rightarrow c} = d * in(b) * \frac{0.25^2}{0.25^2 + 1^2}$,这说明后继节点与前驱节点间的信任值大时,流入后继节点的能量占比大。采用平方标准化方法能避免由于前驱节点出度小所造成的连接边上信任度较低但后继节点却被授予高信任等级的情况。在每次循环过程中,采用式(2)可以更新当前节点 x 的信任值 $trust(x)$:

$$trust(x) \leftarrow trust(x) + (1-d) * in(x) \quad (2)$$

传播激活模型能求出任意源点对其他可达节点的信任值,满足链路上权重大的节点信任度大、源点与节点间存在链路多时信任度大、信任随着链路增长逐渐衰减等相关的信任特性。

建立信任传播模型的算法如算法1所示。在算法开始时,初始化 s 流入网络中各节点 x 的能量 $in(x)$,并向源点初始注入能量 $in(s)$ 。能量在整个网络中沿着信任边流动。算法采用宽度优先遍历更新用户 x 积累的 $trust(x)$ 以及即将流入节点 x 的 $in(x)$ 能量值。两次迭代过程中, $trust(x)$ 的差值若满足规定阈值 T ,则信任度达到收敛,从而获取源点用户对所有其他用户的信任度向量 $trust$,否则继续迭代。

算法1 建立信任传播模型

Input:信任网络关系矩阵 W ,当前源点 s ,信任阈值 T

Output:信任向量 $trust$

1. $NW = compressNetwork(W, s)$; // 求缩小范围后的新的社交网络关系矩阵
2. repeat
3. $i = i + 1$;
4. $V_i = V_{i-1}$;
5. $\forall x \in V_i, in_i(x) = 0$;

6. forall $x \in V_{i-1}$ do

7. $trust_i(x) = trust_{i-1}(x) + (1-d) * in_{i-1}(x)$;

8. forall $(x, u) \in edg(NW)$ do // 遍历所有压缩后矩阵中权重大于0的边

9. if $u \notin V_i$ then

10. 将 u 加入集合 V_i

11. 初始化 $trust_i(u) = 0$ 且 $in_i(u) = 0$;

12. 增加回溯到原点的边 $W(x, s) = 1$;

13. end

14. $in_i(u) = in_i(u) + \frac{W(x, s)^2}{\sum_{(x, s) \in E} W(x, s)^2} * d * in_{i-1}(x)$;

15. end

16. end

17. until 每个用户节点两次遍历的信任差值小于阈值 T

根据六度分离理论可知,最大信任路径长度一般在3到6之间,而社交网络本身也具有小世界特性,因此去除最短路径为6跳以上的用户节点,能大范围地提高计算效率。TCPF算法在实现传播激活模型开始处将初始网络进行了压缩处理。算法1中的 $compressNetwork$ 为此缩小社交矩阵范围的运算函数。

3.2 用户偏好相似度量

在推荐系统中,用户更倾向于选择与自身偏好相似的用户所选择的服务。本文定义两用户评分数据的相似程度为偏好相似度。不妨从两方面考虑用户的偏好相似度,一方面考虑用户共同评价过服务的数量,另一方面考虑两用户间服务评分的差异。定义 $fs(x, y)$ 为用户 x 与用户 y 之间的兴趣相似度(favor similarity), $es(x, y)$ 为用户 x 与用户 y 之间的服务体验相似度(experience similarity),整体偏好相似度为 $sim(x, y)$ 。若两个用户间共同评价的服务数量大且对共同评价过的服务有相似的体验,则可以说明两用户间的偏好相似度高。其中,直接采用Pearson相关系数^[16]计算两个用户间的服务体验相似度 $es(x, y)$,计算时只考虑用户 x 和用户 y 间共同评价过服务的评分相似性。

若用户评价过的服务较少,则用户评分向量高度稀疏,采用Pearson相关系数计算两个用户的相似度的准确率低。于是考虑加入用户兴趣相似度衡量标准 $fs(x, y)$,定义如下:

$$fs(x, y) = \frac{\sum_{i \in N(x) \cap N(y)} 1 / (1 + |I(i)|)}{\sqrt{|N(x)| * |N(y)|}} \quad (3)$$

其中, $|N(x)|$ 表示用户 x 评价过的服务的数量。任意属于用户 x 和用户 y 共同评价服务集合的用户 i 都使得 $fs(x, y)$ 分子部分累加 $1 / (1 + |I(i)|)$ 。 $|I(i)|$ 表示数据集中评价过共同服务 i 的用户数量,数量越大说明服务 i 的代表性越差,不能很好地反馈用户偏好。而评价过共同服务 i 的用户 $|N(x) \cap N(y)|$ 越多, $fs(x, y)$ 越大。该算法在传统Jaccard度量^[17]的基础上增加了服务是否具有代表性这一因素。

结合以上两个标准可以定义用户偏好相似度 $sim(x, y)$ 如下:

$$sim(x, y) = fs(x, y) * es(x, y) \quad (4)$$

用户间的偏好相似度计算算法如算法2所示。其中, $es(x, y)$ 为产品体验相似度,使用Pearson相似度衡量方法获取;而 $fs(x, y)$ 为兴趣相似度,使用考虑服务是否有代表性因素的Jaccard相似度衡量,通过式(3)获取。

算法 2 用户间偏好相似度计算

Input: 用户评分矩阵 R 及当前源点 s

Output: 网络中任意用户与源点用户 s 的相似度向量 sim

1. foreach 遍历评分矩阵 R 中所有用户 x do

2. common = $N(s) \cap N(x)$;

3. foreach $i \in$ common do

4. temp = temp + $\frac{1}{1 + |I(i)|}$;

5. end

6. fs(s, x) = $\frac{temp}{\sqrt{|N(s)| |N(x)|}}$;

7. sim(x) = fs(s, x) * Pearson(s, x);

8. end

3.3 动态结合信任关系与用户偏好的服务推荐算法 TPCF

加入用户间的社交关系,并把社交关系量化成信任度,这使得在获取较少历史数据的情景下依靠用户间的信任关系也能获得较为准确的推荐结果。然而,如果只考虑用户间的社交关系而忽略用户间的偏好相似度,则会导致在社交关系稀疏的情况下推荐结果偏差较大。于是将偏好相似度和信任度进行结合,有助于得到更加准确且合理的推荐结果。本文定义服务推荐权值 TP (Trust and Preference similarity) 如下:

$$TP(x, y) = (1 - r) * sim(x, y) + r * trust(x, y) \quad (5)$$

其中, sim 是用户 x 与用户 y 之间的偏好相似度, $\forall sim(x, y) \in [0, 1]$, $trust(x, y)$ 是用户 x 与用户 y 之间的信任度, $\forall trust(x, y) \in [0, 1]$, $TP(x, y)$ 是介于 0 到 1 之间的实数。

r 表示信任度对整体评分的贡献程度,定义如下:

$$r = \frac{\alpha}{2 + e^{t(s)}} \quad (6)$$

其中, $p(s)$ 表示与用户 s 存在直接信任关系的用户数量, $t(s)$ 表示用户 s 评分数据的数量。当 $p(s)$ 与 $t(s)$ 比值较大时,用户间的信任关系多而评分数据数量少,此时 r 值较大,反之 r 值较小。公式中的 α 是一个常数,用于控制信任度和偏好相似度在整个推荐权值指标中的占比,其取值范围为 0 到 3 之间。实验证明,取 $\alpha=1$ 时推荐效果较好。使用服务推荐权值指标 TP 的优势在于,能够动态调节信任度占整个推荐权重的比例,在信任网络稀疏时减小其占比,在评分网络稀疏时增大其占比。

在获取推荐权值矩阵的前提下,结合传统协同过滤算法,就可以获得最终的推荐列表。

4 实验及分析

实验数据集:文中采用包含 22166 个用户、296277 个产品、922267 条用户评分数据的 Epinions 数据集¹⁾进行实验,数据集中存在一半冷启动用户(评分次数少于 5 次),该数据集是一个非常稀疏的数据集。

实验环境:OSX 10.13.3, 2.2 GHz Intel Core i7/6 GB/1600 MHz DDR3。使用 matlab_R2017b,根据关系矩阵创建社交网络,设计算法完成信任传播模型。使用偏好相似度和信任度综合衡量标准进行协同过滤,并最终得到预测评分推荐列表。

评估标准:

1) 平均绝对误差 MAE (Mean Absolute Error),能反映推荐算法预测评分与反馈评分的相似程度,是衡量算法准确性的指标。计算方式如下:

$$MAE = \frac{1}{|E^p|} \sum_{(i,j) \in E^p} |r_{i,j} - r_{i,j}^A| \quad (7)$$

其中, $r_{i,j}$ 表示用户 i 对产品 j 的实际评分值, $r_{i,j}^A$ 表示用户 i 对产品 j 的预测评分值, $|E^p|$ 表示预测评分总数。MAE 是求所有预测评分偏差绝对值的算术平均值。

2) 均方根误差 RMSE (Root Mean Squared Error),偏向预测值误差较大的情况,更能反映预测评分与反馈评分相比的波动情况。误差越小,说明预测越准确。

$$RMSE = \sqrt{\frac{\sum_{(i,j) \in E^p} (r_{i,j} - r_{i,j}^A)^2}{|E^p|}} \quad (8)$$

3) 搜索引擎算法指标 DCG (Discounted Cumulative Gain),可翻译为“折扣增益值”。将每条结果的相关性分等级来衡量,等级高的结果且位置越靠前,则 DCG 越大。

$$DCG = \sum \frac{(2^{r(s,u)} - 1)}{\log_2(1 + p_u)} \quad (9)$$

其中, $r(s, u)$ 表示用户 s 对产品 u 的预测评分值, p_u 表示在实际推荐列表中 u 所处的位置。

4) 评分覆盖率 COV (Coverage Ratio),用来描述算法的覆盖率,即推荐系统能够预测的评分数量占整个测试集合评分数量的百分比。

$$COV = \frac{\text{预测成功的评分数量}}{\text{所有评分数量}} \quad (10)$$

实验步骤:

1) 建立信任传播模型实验

本文使用传播激活模型建立信任网络,由此获取源点用户对其他用户节点的信任值。图 5 表示在初始注入能量不同时信任值的收敛情况。图中横坐标代表循环次数,纵坐标代表当前循环所有用户节点信任值与上次循环信任值的差值。随着循环次数的增加,信任差值缩减速度加快。

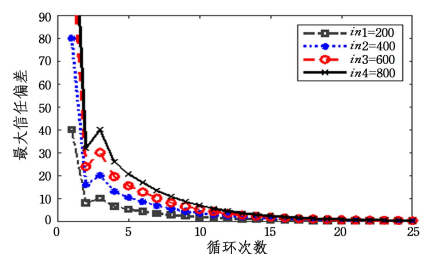


图 5 初始激活能量对信任值收敛的影响

从图 5 可知,不论初始注入网络中的信任能量 in_i 有多少,任意节点的信任值都可以在有限次循环中收敛。

为了保证能量在网络中流通并伴随着能量的衰减,文中规定了衰减因子 d ,即能量流入后继节点的占比。图 6 显示了不同衰减因子取值情况下信任度的分布情况。实验证明了参数 d 的值越大,节点的信任度分配得就越均匀,既不会过分信任距离源点近的节点,也不会过分不信任离源点远的节点。当 $d=1$ 时,无法在节点上累积能量,算法就会失效。于是,文中使用 $d=0.8$ 作为衰减因子,既保证能够获取网络中各节点的信任值,又使信任值分布均匀。

¹⁾ Epinions 数据集获取网站: <http://www.epinions.com>

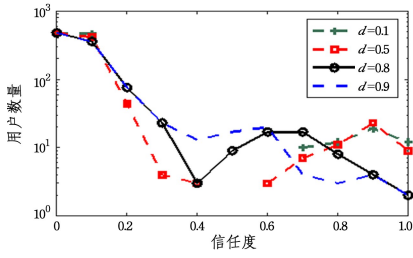


图6 衰减因子对用户信任度分布的影响

2) sim 相似度量效果验证

计算用户间的偏好相似度是以评分矩阵为输入,综合用户间的兴趣相似度 $f_s(x,y)$ 和服务体验相似度 $e_s(x,y)$ 得到的。表1对比了基于用户的协同过滤算法在不同相似度量方法下的推荐效果。

表1 相似度量性能的对比如

相似度量	MAE	RMSE
Pearson	0.826	1.009
余弦相似度	0.801	1.003
对数似然值	0.810	1.010
曼哈顿距离	0.827	1.073
Jaccard系	0.785	0.968
sim 相似度	0.772	0.958

可以得出,使用本文提出的 sim 相似度量来进行协同过滤推荐能得到更好的推荐效果。

3) 推荐算法参数估计实验

本文定义的服务推荐权重衡量标准 $TP(x,y)$ 如下:

$$TP(x,y) = (1-r)sim(x,y) + r * trust(x,y)$$

其中, $r = \frac{\alpha}{2 + e^{-\rho(s)/t(s)}}$, 作为信任度对整体推荐度的贡献。图7是在不同 α 的情况下抽样 11 个子网络后所得到的推荐结果的 DCG(搜索引擎算法指标)比较。可以看出,当 $\alpha=1$ 时推荐结果较好。

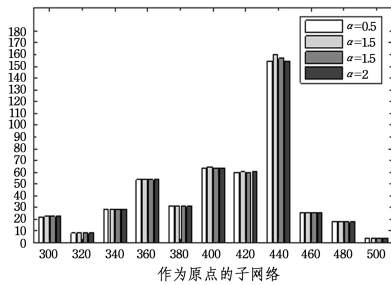
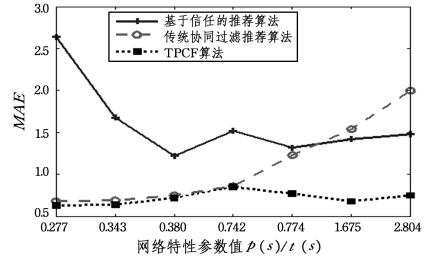


图7 不同 α 值对整体推荐效率的影响

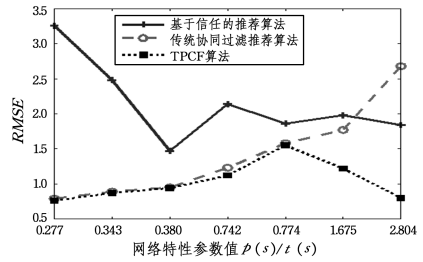
除了参数 α 之外, r 还受到直接信任关系数量 $p(s)$ 与评分数据数量 $t(s)$ 比值的影响。这两个参数的比值与选取的用户关系集合和评价集合相关。在用户社交关系稀疏而评分数据充足时, $p(s)/t(s)$ 较大, r 较大, 信任关系对整体推荐权重的贡献大。在社交关系密集而评分数据稀疏时, $p(s)/t(s)$ 较小, r 较小, 信任关系对整体推荐权重的贡献小。

图8中采样了 Epinions 数据集中代表 7 种具备不同评分稀疏度和信任稀疏度的源点形成的子网络,随着比值的增加,信任关系 $p(s)$ 逐渐增加,而评价数量 $t(s)$ 逐渐减少;在每个子网络上使用经典协同过滤算法、信任推荐算法以及 TPCF 算法进行预测。比较 7 个组的推荐质量可知,推荐结果与 $p(s)/t(s)$ 比值的大小存在一定联系。该参数比值增大时,基于传统协同过滤算法的误差率增大,而基于信任推荐的算法的误差率呈减小趋势。出现该情况的原因在于当评价数据减少、

网络信任关系增多时,基于协同过滤推荐算法的准确率会因为数据稀疏而减小,而基于信任的算法会因为具备相对更多的信任关系而获得更准确的推荐。TPCF 算法的整体准确率波动不大,这是由于 TPCF 算法结合了信任推荐算法以及传统协同过滤推荐算法,使用参数 $r = \frac{\alpha}{2 + e^{-\rho(s)/t(s)}}$ 动态地控制两部分对整体 TPCF 推荐算法的贡献占比,从而一定程度地消除了网络中信任值稀疏或评分值稀疏时造成的误差。



(a) MAE 的对比

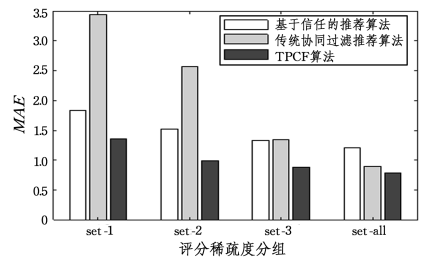


(b) RMSE 的对比

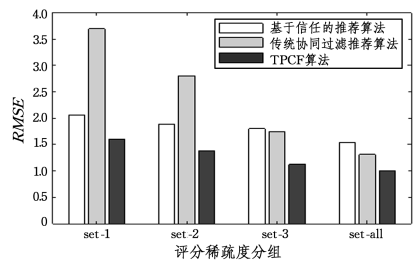
图8 网络特性参数对整体推荐效率的影响

4) 3 种推荐算法效果比较

TPCF 算法是通过动态结合信任度和偏好相似度两部分指标得到推荐权重,并进行协同推荐的算法。图9对比了 TPCF 算法与单纯的信任度协同推荐和传统协同过滤算法推荐效果。其中, set-1 表示采样评分数据数量小于 3 的源点集合, set-2 是采样评分数量位于 3 到 5 之间的源点集合, set-3 是采样评分数据数量位于 5 到 10 之间的源点集合,而 set-all 表示整体 Epinions 集合中随机采样源点集合。从图9中可以看到,相比于其他两种方法,TPCF 方法的 MAE 值及 RMSE 值较小,在预测的准确率上优于其他两种算法。



(a) MAE 的对比



(b) RMSE 的对比

图9 3 种推荐算法性能的比较

在评分数据稀疏时,计算用户间的偏好相似度的误差较

大,因而经典 CF 算法的推荐效果不佳。但基于信任权重的推荐的准确性取决于信任数据是否稀疏,与评分数据的数量无关。TPCF 算法由于动态地结合了经典 CF 算法以及基于信任权重的推荐,受到评分数据稀疏的影响较小,因此在评分数据稀疏时依然能够获取较高的推荐准确率。

当 $r=0$ 时,TPCF 算法退化成为以 sim 为相似度量标准的协同过滤算法,整个 TP 综合衡量标准只考虑用户偏好相似度而不考虑信任关系,存在冷启动数据稀疏等问题;当 $r=1$ 时,TPCF 算法成为单纯基于信任权重的协同过滤方法,虽然能够解决用户冷启动问题,但由于用户信任关系较为稀疏,只与网络中少数用户存在联系,因此单纯使用信任模型的推荐存在较大误差。TPCF 综合了信任与偏好关系,使得用户间存在更多的可衡量关系,结合了两种推荐算法的优点,扩充了推荐算法的覆盖率。如图 10 所示,采样了编号 0—5000 的全体用户,并将其平均分成 5 组,每一组为连续的 1000 个用户及其评价值。计算其各自不同的平均覆盖率。由图 10 可知,单纯基于信任的推荐方法由于数据集中信任数据稀疏的问题而平均覆盖率较低;传统协同过滤算法在用户评分数据稀疏时的平均覆盖率也较低;TPCF 算法使用了社交网络数据,对只根据用户评分计算整体信任权重的协同过滤方法进行了补充。尽管单一使用社交网络数据形成信任度并进行推荐算法的平均覆盖率并不高,但结合两者优势的 TPCF 算法在整体覆盖率上有较大提升。

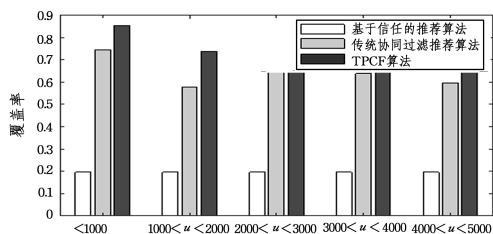


图 10 3 种推荐算法的覆盖率比较

5) 算法性能对比

将本文提出的 TPCF 算法的平均性能与经典推荐算法以及其他结合信任网络的推荐算法进行对比,结果如表 2 所列。经典的 CF 算法在评分数据稀疏时,覆盖率非常低,无法满足用户需求。其余几种基于信任网络的推荐算法在覆盖率方面有了显著提升,但在预测准确度方面还有待提高。TPCF 算法综合了用户间信任度和偏好相似度,有效解决了在评分数据稀疏时无法根据偏好相似度进行高效推荐的问题。从表 2 可以看到,在 Epinions 数据集上,TPCF 方法在 MAE 和 RMSE 上都有一定程度的降低,评分预测更加准确;而 TPCF 算法综合了信任度和偏好相似度两个指标,使得任意用户都可以得到推荐权值,极大地增大了预测评分成功的数量,提升了覆盖率。

表 2 推荐算法的性能对比

推荐方法	MAE	RMSE	COV/%
CF	0.844	1.272	49.52
PMF	0.835	1.264	78.99
SocialMF	0.832	1.262	79.12
TidalTrust	0.830	1.261	80.03
TrustWalker	0.820	1.145	79.86
TPCF	0.798	1.001	80.87

结束语 本文首先介绍了当前推荐系统领域的发展状况及目前流行的推荐算法;其次介绍了信任的相关理论知识,包

括信任的定义、传递、聚合等基本问题;然后介绍了信任传播激活模型的建立,并通过信任传播方式获得整个网络的信任度;另一方面,设计相似度量标准 sim ,将用户评分数据作为依据,综合考虑用户间的共同评分产品数量和评分值的差异,确定用户间兴趣相似度,提升了推荐准确率;接着提出了本文核心算法 TPCF,将用户间偏好关系与社交关系动态结合,得出了新的衡量标准来进行协同过滤;根据网络性质动态调节了两部分标准的占比,不仅扩大了用户之间存在的联系,还提升了在评分数据稀疏或者信任数据稀疏时算法的可用性;最后通过实验证明,TPCF 算法有效地提高了推荐系统的预测准确度和覆盖率,一定程度上解决了传统方法对于冷启动以及数据稀疏时推荐覆盖率低的问题,提升了推荐的整体准确率。然而,利用信任网络虽然能够有效解决用户的冷启动问题,但是解决不了项目的冷启动问题。在某项目没有被任何用户评分的情况下,还是无法通过本文算法得到较好的推荐。此时,利用项目固有属性计算项目相似度能够解决项目的“冷启动”问题。因此,未来可以通过在推荐系统中加入项目间相似度的相关内容来进一步提高 TPCF 的覆盖率。

参 考 文 献

- [1] 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.
- [2] LIU G, WANG Y, ORGUN M A, et al. A Heuristic Algorithm for Trust-Oriented Service Provider Selection in Complex Social Networks[C] // IEEE International Conference on Services Computing. IEEE Computer Society, 2010: 130-137.
- [3] GARTRELL M, XING X, LV Q, et al. Enhancing group recommendation by incorporating social relationship interactions[C] // International ACM Siggroun Conference on Supporting Group Work. 2010: 97-106.
- [4] TANG M, XU Y, LIU J, et al. Combining Global and Local Trust for Service Recommendation[C] // IEEE International Conference on Web Services. IEEE Computer Society, 2014: 305-312.
- [5] ZHAO T, MCAULEY J, KING I. Leveraging Social Connections to Improve Personalized Ranking for Collaborative Filtering[C] // ACM. 2014: 261-270.
- [6] HWANG C S, CHEN Y P. Using Trust in Collaborative Filtering Recommendation[M] // New Trends in Applied Artificial Intelligence. Berlin: Springer, 2007: 1052-1060.
- [7] HANG C W. Probabilistic Trust Models for Social and Service Networks[D]. North Carolina State University, 2013.
- [8] JAMALI M, ESTER M. TrustWalker: a random walk model for combining trust-based and item-based recommendation[C] // Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2009: 397-406.
- [9] ZIEGLER C N, LAUSEN G. Propagation Models for Trust and Distrust in Social Networks[M]. Kluwer Academic Publishers, 2005.
- [10] TANG J, GAO H, LIU H, et al. eTrust: understanding trust evolution in an online world[C] // ACM SIGKDD International

Conference on Knowledge Discovery and Data Mining. ACM, 2012;253-261.

- [11] DENG S, HUANG L, XU G, et al. On Deep Learning for Trust-Aware Recommendations in Social Networks[J]. IEEE Transactions on Neural Networks & Learning Systems, 2016, 28(5): 1164-1177.
- [12] YANG B, LEI Y, LIU J, et al. Social Collaborative Filtering by Trust[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(8): 1633-1647.
- [13] XIONG L, WANG L, HUANG Y. An Approach for Top-k Recommendation Based on Trust Information[C]// Conference on Service-Oriented Computing and Applications. IEEE Computer

Society, 2017; 198-205.

- [14] GUO L, ZHANG C, FANG Y. A Trust-Based Privacy-Preserving Friend Recommendation Scheme for Online Social Networks[J]. IEEE Transactions on Dependable & Secure Computing, 2015, 12(4): 413-427.
- [15] 宗刚, 孙玮, 任蓉. 基于信任机制的复杂网络知识传播模型的研究[J]. 价值工程, 2009, 28(12): 94-97.
- [16] 牛常勇, 刘国枢. 基于局部全局相似度的 SVD 的协同过滤算法[J]. 计算机工程与设计, 2016, 37(9): 2497-2501.
- [17] REAL R, VARGAS J M. The Probabilistic Basis of Jaccard's Index of Similarity[J]. Systematic Biology, 1996, 45(3): 380-385.

(上接第 376 页)



图 11 16×16 阈值矩阵加网提取效果图

对防伪信息的提取清晰程度进行评价,结果如表 2 所列。

表 2 提取信息清晰度的客观评价数据

阈值矩阵	FSIM					
	S-C	C-S	S-R	R-S	C-R	R-C
8×8	0.9092	0.9088	0.9104	0.9088	0.9080	0.9085
16×16	0.9679	0.9682	0.9698	0.9784	0.9674	0.9684

分析表 2 可知,同一阈值矩阵、不同网点组合下的 FSIM 相近,防伪信息提取效果均较好。其中,16×16 阈值矩阵的 FSIM 值更接近 1,验证了阈值矩阵越大,图像防伪效果越佳的主观评价。综合考虑,选取圆形-菱形组合网点,能获得较好的隐蔽性,且提取的防伪信息的显著性也较好。

结束语 结合最小阈值矩阵法和图像模板匹配原理,设计了基于不同网点组合加网半色调信息隐藏方案,从而保证了图像的优质再现及防伪性能的有效提升,为半色调信息组合防伪技术的网点选择提供了有效的参考依据。不同形状网点生成的防伪图像之间存在差异,其中圆形-菱形形成的防伪图像的图像还原效果及防伪性能相对较优,是最佳的组合选项,为通过调节网点形状改善防伪性能的相关研究提供了有价值的理论依据。

参考文献

- [1] 吕辉辉. 基于半色调信息隐藏的同步信息嵌入与提取算法研究[D]. 北京:北京印刷学院, 2016.
- [2] WUX T, SUN W. Visual data hiding in dot diffusion images [C] // International Conference on Computer Sciences and Convergence Information Technology. IEEE, 2011: 588-593.
- [3] 衣旭梅. 基于半色调加网防伪印刷的信息隐藏技术的研究[D].

- 北京:北京印刷学院, 2011.
- [4] 孟凡俊. 基于伪随机信号调制的半色调加网防伪技术研究[D]. 北京:北京印刷学院, 2011.
- [5] 郝发义, 邓开发. 基于网点形状的半色调全息防伪算法研究[J]. 包装工程, 2014(9): 122-126.
- [6] 徐锦林, 陈素明. 调幅加网阈值矩阵设计方法与加网原理[J]. 中国印刷与包装研究, 2009, 1(3): 38-42.
- [7] XU J L. A Rational Tangent Halftoning Algorithm with a Single Threshold Matrix[C] // Proceedings of 2008 International Conference on Computer Science and Software Engineering. Washington, USA: IEEE Computer Society, 2008: 181-184.
- [8] 徐锦林, 徐咏驰. 调幅加网因子研究[J]. 中国印刷与包装研究, 2009, 1(6): 20-24.
- [9] MEAE M, VAIDYANATHAN P P. Recent advances in digital halftoning and inverse halftoning methods[J]. IEEE Transactions on Circuits & Systems I Fundamental Theory & Applications, 2002, 49(6): 790-805.
- [10] LEWIS J P. Fast Template Matching[J]. Vision Interface, 1995, 32(4): 351-361.
- [11] ZHANG S, ZHOU Y. Template matching using grey wolf optimizer with lateral inhibition[J]. Optik-International Journal for Light and Electron Optics, 2017, 130: 1229-1243.
- [12] GIR. What's wrong with mean-squared error [M] // Digital Images and Human Vision. Cambridge: MIT Press, 1993: 207-220.
- [13] 王宇庆, 刘维亚, 王勇. 一种基于局部方差和结构相似度的图像质量评价方法[J]. 光电子·激光, 2008, 19(11): 1546-1553.
- [14] ESKICIOGLU A M, FISHER P S. Image quality measures and their performance[J]. IEEE Transactions on Communications, 1995, 43(12): 2959-2965.
- [15] KITE T D, EVANS B L, BOVIK A C. Modeling and quality assessment of halftoning by error diffusion[J]. IEEE Transactions on Image Processing, 2000, 9(5): 909-922.
- [16] 曾台英, 邵雪, 汪祖辉. 一种基于特征相似性的印刷图像质量评价方法[J]. 包装工程, 2016, 37(11): 153-157.
- [17] ZHANG L, ZHANG L, MOU X, et al. FSIM: A Feature Similarity Index for Image Quality Assessment[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2011, 20(8): 2378.
- [18] NIZAMI I F, MAJID M, AFZAL H, et al. Impact of Feature Selection Algorithms on Blind Image Quality Assessment[J]. Arabian Journal for Science & Engineering, 2017, 43(4): 1-14.