·人工智能及识别技术· 文章编号：1000-3428(2017)08-0236-07 文献标志码：A 中图分类号：TP391

融合信任传播和奇异值分解的社会化推荐算法

李卫疆[1]，齐 静[1]，余正涛[1]，赵铁军[2]

（1. 昆明理工大学 信息工程与自动化学院，昆明 650500；
2. 哈尔滨工业大学 计算机学院，哈尔滨 150001）

摘 要： 针对用户信任矩阵中的数据稀疏问题，设计用户信任关系的传播规则，根据该规则计算用户之间的信任度，填充用户信任矩阵。在此基础上，结合用户信任传播算法和奇异值分解模型，提出一种社会化推荐算法，将用户评分矩阵与信任关系矩阵相结合，提高推荐系统的预测准确率。在 Epinions 和 Filmtrust 公开数据集上的实验结果表明，该算法相比传统推荐算法具有更高的推荐质量。

关键词： 推荐系统；社会化推荐；信任网络；信任传播；奇异值分解

中文引用格式： 李卫疆，齐 静，余正涛，等. 融合信任传播和奇异值分解的社会化推荐算法[J]. 计算机工程，2017，43(8)：236-242.

英文引用格式： Li Weijiang, Qi Jing, Yu Zhengtao, et al. Social Recommendation Algorithm Integrating Trust Propagation and Singular Value Decomposition[J]. Computer Engineering, 2017, 43(8): 236-242.

Social Recommendation Algorithm Integrating Trust Propagation and Singular Value Decomposition

LI Weijiang[1], QI Jing[1], YU Zhengtao[1], ZHAO Tiejun[2]

(1. Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China;
2. School of Computer, Harbin Institute of Technology, Harbin 150001, China)

【Abstract】 Aiming at the data sparsity of user trust matrix, this paper designs a propagation rule for trust relationships among users. It computes the trust degree of user according to the rule, and then uses the trust degree to fill the user trust matrix. It proposes a social recommendation algorithm based on users' trust propagation algorithm and Singular Value Decomposition (SVD) model, The user scoring matrix is combined with the trust relation matrix to improve The prediction accuracy of the recommended system. Experimental results on both Epinions and Filmtrust publicly available datasets show that compared with the traditional recommendation algorithm, the proposed algorithm has higher recommendation quality.

【Key words】 recommendation system; social recommendation; trust network; trust propagation; Singular Value Decomposition (SVD)

DOI: 10.3969/j.issn.1000-3428.2017.08.040

0 概述

随着 Web2.0 技术的飞速发展，现在人们正处于一个数据爆炸性增长的时代。互联网中充斥着各种飞速增长的数据，因此，如何从大量纷杂的互联网数据中高效准确地找出用户感兴趣的信息已成为推荐系统研究的焦点问题。目前推荐系统的主要任务是基于信息过滤技术向用户推荐一些用户感兴趣的项目，例如电影、音乐、新闻、朋友、网页等[1-3]。

传统推荐算法已经被广泛应用在电子商务产业，例如 Amazon，EBay 等[4-5]。然而，传统推荐系统中的推荐算法大多面临以下问题：1）大多数的传统推荐算法都假设用户之间是相互独立的。这种假设忽略了用户之间存在的社交关系和联系，这与现实世界中的人际关系不符[2,6]。根据社会选择理论和社会影响原理，人们很容易与他们相似的人们产生联系。除此之外，在社会网络中人们彼此之间很容易产生影响并且逐渐变得更加紧密。同时，随着社交网络（例如 Facebook，Twitter，Micro-blog 等）的日益流行，用户之间的关系逐渐成为提高推荐系统推荐质量的一个重

基金项目： 国家自然科学基金（61363045）；科技部科技创新人才基金（2014HE001）；云南省自然科学基金重点项目（2013FA130）。
作者简介： 李卫疆（1969—），男，副教授、博士，主研方向为自然语言处理、信息检索；齐 静，硕士研究生；余正涛、赵铁军，教授、博士。
收稿日期： 2016-06-23 **修回日期：** 2016-09-05 **E-mail：** hrbrichard@126.com

要因素[7-9]。2)现在越来越多的推荐算法为了考虑社交网络中用户之间的关联关系,把用户之间的社会信息融入到推荐算法中。然而,关于用户之间的信任信息非常少,这导致在大多数的数据集中只挖掘出极少数用户之间的关系数据。本文针对现有推荐算法所面临的问题和挑战,融合信任传播和奇异值分解(Singular Value Decomposition,SVD)模型,提出一种新的推荐算法。

1　相关工作

协同过滤技术[10-12]是传统推荐系统中较为成熟的一种推荐技术。协同过滤推荐技术得到了广泛应用的主要原因是其推荐过程独立于推荐对象的推荐内容,仅依赖用户的评分或者用户的喜好就能为用户推荐。同时,协同过滤技术也能帮助用户发现潜在兴趣。现在流行的协同过滤技术大多是基于低维因子模型。文献[13]针对许多现有的协同过滤方法在处理大数据集和稀疏数据集时推荐效果不明显的问题,提出一种概率矩阵分解(Probabilistic Matrix Decomposition,PMF)模型。该模型在推荐过程中仅通过用户-项目矩阵的现存评分数据预测用户对未知项目的评分。文献[14]提出了一种新的奇异值分解模型。该模型通过使用奇异值分解技术降低用户-项目矩阵的稀疏度。文献[15]提出了一种改进的 SVD 模型,即SVD ++模型。该模型在 SVD 的基础上引入隐式反馈,使用用户的历史浏览数据、用户历史评分数据、电影的历史浏览数据、电影的历史评分数据等作为新的参数。然而,以上方法都假设用户之间是相互独立的,该假设没有考虑到用户之间的社会关系和相互之间的联系,与现实社会中的社会关系不符。

近年来,随着社交软件的迅猛发展,社交网络已经日渐成熟,社交网络中用户之间的信任关系已经成为各大电商推荐平台向用户推荐商品的重要信息[16-17]。用户不仅只是作为简单的信息接受者,而且还能够主动地建立用户之间的联系。所以,大多数的推荐算法开始关注如何使用用户之间的社会关系提高推荐系统的推荐质量[18-20]。文献[9]提出一种混合方法,该方法融合了信任者和被信任者模型,即主动信任的用户和被信任的用户都将影响用户对位置项目的评分。文献[10]提出基于概率矩阵分解的推荐方法,通过一个共享的用户潜在特征空间,把社交网络结构和用户-项目评分矩阵连接起来。实验结果表明,在用户评分很少或没有评分的情况下该方法比当前流行的算法效率好,文献[11]从解决数据稀疏的角度出发,提出一种概率因素分析框架,该框架通过一个集合参数把用户的喜好和他们朋友的喜好进行融合。通过对复杂度进行分析得出该方法能够适用于较大的数据集,而且实验结果表明该方法比当前流行方法的效率更高。

上述方法都表明用户的社交网络信息对于提高推荐系统的效率具有重要作用。然而,上述方法都没有结合信任传播对推荐质量的影响提高推荐系统的推荐质量。在现实世界中,用户之间的信任关系是可以传递的。为此,本文设计了一种信任传播算法。该算法将具有直接关联关系的用户作为传输信任的纽带,推导出没有直接关联关系的用户之间的信任度,从而填充信任关系矩阵,解决了用户信任关系矩阵稀疏的难题,并且利用用户之间的信任关系进行推荐也大幅提高了推荐系统的推荐质量。同时,由于用户评分矩阵的稀疏性导致了推荐系统的推荐质量降低,因此奇异值分解模型可以将数据映射到低维空间,计算低维空间中检索对象之间的相似度,对用户未评分的检索对象进行评分预测,将预测评分高的检索对象推荐给用户,提高推荐系统的推荐质量。

2　信任传播规则

2.1　相关定义

为更直观地了解社交网络,本文给出一个社交网络实例如图 1 所示,其中,节点代表用户;边代表用户之间的社交依赖关系;? 表示用户之间没有直接关联关系,即需要通过信任传播规则计算用户之间的关联关系。在社交网络中,用户 u 的直接邻居为 N_u,用户与直接邻居之间的信任度存放在信任矩阵 $\boldsymbol{T}=[t_{u,v}]_{N\times N}$中。其中,$t_{u,v}$代表从用户 u 到用户 v 之间存在社交关联,其值可以取 0 或者 1,0 表示没有信赖关系,1 代表有信赖关系。由于有些社交网络(例如 Twitter,Micro-blog)中用户 u 信任用户 v 并不代表用户 v 也信任用户 u,因此信任矩阵 $\boldsymbol{T}$ 通常是一个非对称矩阵[18]。

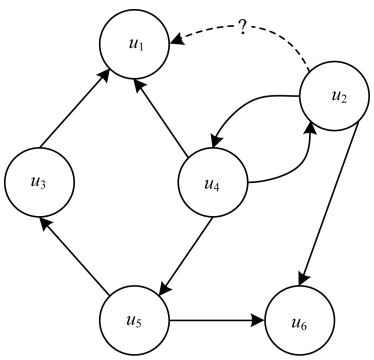


图 1　社交网络实例

信任网络中的信任传播问题是指某用户 u 在用户信任矩阵的基础上,将具有直接关联关系的用户作为传输信任的纽带,推导出没有直接关联关系的用户之间的信任度。换言之,由于现实数据集中关于用户之间的信任信息非常少,这导致了在大多数

数据集中只挖掘到了极少数用户之间的关系数据，用户信任矩阵非常稀疏，因此信任网络中的信任传播算法可以通过一定的信任传播规则推导出用户之间的间接信任关系，从而填充用户信任矩阵。如图 1 所示，用户 u_2 和用户 u_1 之间并不存在直接的信任关系，但是用户 u_2 和用户 u_1 之间的信任关系可以通过 $\{u_2 \rightarrow u_4, u_4 \rightarrow u_1\}$ 推导出来，该过程即为信任传播过程。

定义 1 信任传播的路径数 L：假设 $\boldsymbol{T} = \{u_1, u_2, \cdots, u_m\}$ 是由用户结点组成的信任网络，在信任网络 $\boldsymbol{T}$ 中任意选 2 个结点 u_i 和 u_j 分别作为起点结点和终点结点。从起点结点 u_i 到终点结点 u_j 所经历的中间路径数称为信任传播的路径数 L。

根据人际关系中的六度分隔理论（或小世界理论）和网络关系中的四度分离理论可知，任何 2 个陌生人之间所间隔的人不会超过 6 个，即最多通过5 个中间人就能够认识任何一个陌生人[21-22]。信任网络中的所有结点最多通过 5 个中间结点就能找到其他结点，因此，信任传播的跳数应设置为 $2 \leqslant L \leqslant 6$。由于信任传播路径超过路径数 L 的人不在用户的信任社区中，并且与用户之间没有信任关系，因此超过信任传播路径数 L 的结点将不会被加入到结点的关系中。

定义 2 用户 u 的朋友是一个用户集合，如果用户 u 直接信任用户 v，那么用户 v 是用户 u 的朋友；否则用户 v 不是用户 u 的朋友。

定义 3 用户 u 的社区是一个用户集合，如果在信任网络 T 中至少存在一条从用户 u 到达用户 v 的路径，那么用户 v 在用户 u 的社区中；否则用户 v 不在用户 u 的社区中。

定义 4 信任网络的信任度矩阵 $\boldsymbol{T} = (t_{uv})_{n \times n}$，其中，$t_{uv}$ 代表用户 u 信任用户 v 的程度，$t_{uv} \in [0,1]$，值越大表示用户 u 对用户 v 的信任度越高。

2.2 信任传播算法

信任关系的传递是信任网络的一个重要特征。在实际应用中，网络中的每一个用户都与其他用户有直接关联关系的情况是不可能的。但是，根据现实世界中的四度空间理论和现实世界的经验，用户之间的信任是可以传递的，可以通过具有直接关联关系的用户作为传输信任的纽带推导出没有直接关联关系的用户之间的信任度。信任传播算法（Trust Propagation Algorithm，TPA）首先按路径长度从 1 ~ L（L 为信任传播步长）搜索从源节点到目标节点的所有路径，按信任度随着路径长度衰减的规律计算每条路径传播的信任度，最后累加起来并进行归一化处理，得到源节点对目标节点的间接信任度值，从而计算出整个信任矩阵的信任度。

根据以上思路设计的信任传播算法具体描述如下：

算法 1 信任传播算法

输入 用户 u，用户信任矩阵 $\boldsymbol{TR}$，信任传播步长 L，信任衰减系数 p

输出 用户信任度向量 $\boldsymbol{T}_u$

1. 初始化：初始化信任度向量 T_u；Friends(u)表示用户的朋友集合，U_i 表示第 i 步搜索到的用户集合 $T_u^i = \{t'_{uv} \mid$经过i步之后从用户 u 传播到用户 v 的信任度$\}$
2. 对于所有的用户 v
3. if v ∈ Friends(u) then
4. 添加 $t_{uv} = 1$ 到 T_u
5. 添加 v 到 U_1
6. 从 TR 删除 tr_{uv}
7. else
8. $t'_{uv} = \text{Scale}(t_{uv})$，$\text{Scale}(x) = \frac{2}{\pi}\arctan(x)$
9. i 为 2 ~ L
10. 对于所有的 $v \in U_i$
11. 对于所有的 $w \in$ Friends(v)
12. if $w \notin$ Friends(v)
13. $t'_{uw} = t_{uw} + t^i_{uw} \cdot p$
14. 更新 T_u
15. 添加 w 到 U_{i+1}
16. 从 TR 中删除 tr_{vw}
17. 返回 T_u

复杂度分析：算法 1 的主要资源开销是计算一个用户及其社区里的其他用户之间的信任度的过程。该过程实际上是对用户信任关系图做一次宽度优先遍历，最坏情况下的时间复杂度为 $O(n^2)$，其中 n 为用户数量。但是实际中这种最坏情况是不可能存在的，因为本文提出的信任传播算法中引入了信任传播步长 L，根据六度空间理论，信任网络中的所有结点最多通过 5 个中间结点就能找到其他结点，所以信任传播的跳数应设置为 $2 \leqslant L \leqslant 6$。超过信任传播路径数 L 的结点将不会被加入到结点的关系中。由于信任矩阵稀疏性和信任传播步长 L 的限制，因此该信任度算法的复杂度远小于 $O(n^2)$。

3 SocialSVD 模型

3.1 基于 SVD 的推荐模型

SVD ++算法[15]是在 SVD(Singular Value Decomposition)的基础上的改进算法。该算法在 SVD 的基础上引入了隐式反馈，将用户历史浏览数据、用户历史评分数据、电影历史浏览数据、电影历史评分数据等作为新的参数。

本文假设有 m 个用户和 n 个 item，item 通常指被检索对象。评分矩阵 $\boldsymbol{R}$（$R = [r_{u,j}]_{m \times n}$）中存放用户对项目的评分，$r_{u,j}$ 表示用户 u 对项目 j 的评分，表示用户对项目的偏爱范围是 1 ~ 5。本文使用 I_u 表示用户 u 的评分集合，$\boldsymbol{p}_u$ 表示用户 u 的 d 维潜在特征向量，那么 $\boldsymbol{P} \in \boldsymbol{P}^{d \times m}$，$\boldsymbol{P}^{d \times m}$ 表示用户的潜在特征矩阵；$\boldsymbol{q}_j$ 表示项目 j 的潜在特征向量，那么 $\boldsymbol{Q} \in \mathbb{R}^{d \times n}$ 表示项目的潜在特征矩阵。因此，评分矩阵 $\boldsymbol{R} \approx \boldsymbol{P}^{\mathrm{T}}\boldsymbol{Q}$，

其中,$\boldsymbol{P}^{\mathrm{T}}$ 表示 $\boldsymbol{P}$ 的转置矩阵。因此,用户 u 对项目 j 的评分可以通过用户特征向量 $\boldsymbol{p}_u$ 和项目特征向量 $\boldsymbol{q}_j$ 的内积产生,即 $r_{u,j}=\boldsymbol{q}_j^{\mathrm{T}}\cdot\boldsymbol{p}_u$。推荐系统的主要任务是使预测值 $\hat{r}_{u,j}$尽可能接近真实值 $r_{u,j}$。形式上可以通过最小化以下差值函数学习用户特征矩阵和项目特征矩阵。

$$\boldsymbol{L}_r=\frac{1}{2}\sum_u\sum_{j\in I_u}(\boldsymbol{q}_j^{\mathrm{T}}\boldsymbol{p}_u-r_{u,j})^2 \tag{1}$$

SVD ++是利用隐式反馈信息的矩阵分解模型。在一般情况下,隐式反馈指的是能够有助于反映用户喜好的任何类型的用户历史信息。本文可以把 SVD ++模型形式化为如下方程[15]:

$$\hat{r}_{u,j}=\boldsymbol{b}_u+\boldsymbol{b}_j+\mu+\boldsymbol{q}_j^{\mathrm{T}}(\boldsymbol{p}_u+|I_u|^{-\frac{1}{2}}\sum_{i\in I_u}y_i) \tag{2}$$

其中,向量 $\boldsymbol{p}_j$ 表示电影 j 的评分相对于平均评分的偏差;向量 $\boldsymbol{p}_u$ 表示用户 u 做出的评分相对于平均评分的偏差,将平均评分记做 μ;y_i 表示用户 u 在过去对项目的评分在未来对未知项目的隐式影响。用户 u 可以进一步建模为 $\boldsymbol{p}_u+|I_u|^{-\frac{1}{2}}\sum_{i\in I_u}y_i$。

3.2　SocialSVD 模型描述

本文假设社交网络中直接信任网络图为 $G=(\nu,\varepsilon)$,其中,结点集 $\nu=\{v_i\}_{i=1}^m$表示社交网络中的所有的结点;边集 ε 表示用户之间的信任关系。$\boldsymbol{T}=[t_{u,v}]_{m\times m}$表示图 G 的 $m\times m$ 阶矩阵,也可以称为社交网络矩阵。其中,$t_{u,v}$表示用户 u 对用户 v 的信任程度。$\boldsymbol{p}_u$ 和 $\boldsymbol{w}_v$ 分别表示 d 维信任者特征向量和 d 维被信任者特征向量。因此,$\boldsymbol{P}^{d\times m}$和 $\boldsymbol{W}^{d\times m}$分别表示信任者特征矩阵和被信任者特征矩阵。为了能更好地连接用户信任矩阵与用户项目评分矩阵,在本文中使用相同的用户特征向量。与用户评分相同,本文使用 $\boldsymbol{T}\approx\boldsymbol{P}^{\mathrm{T}}\boldsymbol{W}$ 恢复社交网络矩阵,所以,推荐系统的主要任务是使预测值 $\hat{t}_{u,j}=\boldsymbol{w}_v\boldsymbol{p}_u$ 尽可能地接近真实的信任值 $t_{u,j}$。综上所述,本文通过最小化以下目标函数学习信任者特征矩阵和被信任者特征矩阵:

$$\boldsymbol{L}_t=\frac{1}{2}\sum_u\sum_{v\in\boldsymbol{T}_u}(w_v^t\boldsymbol{p}_u-t_{u,v})^2 \tag{3}$$

其中,$\boldsymbol{T}_u$ 表示用户 u 的信任度。

本文强调了信任关系的影响,因此,把信任关系的影响以与评分影响相同的方式加入到SVD ++模型中。所以,新的用户对项目的评分预测可以通过式(4)表示:

$$\hat{r}_{u,j}=\boldsymbol{b}_u+\boldsymbol{b}_j+\mu+\boldsymbol{q}_j^{\mathrm{T}}\left(\boldsymbol{p}_u+|I_u|^{-\frac{1}{2}}\sum_{i\in I_u}y_i+|\boldsymbol{T}_u|^{-\frac{1}{2}}\sum_{v\in\boldsymbol{T}_u}\boldsymbol{w}_v\right) \tag{4}$$

因此,与评分矩阵中的用户特征向量类似,本文将用户的信任特征融合到预测评分 $\hat{t}_{u,j}$中,即$|\boldsymbol{T}_u|^{-\frac{1}{2}}\sum\boldsymbol{w}_v$。所以,用户可以被进一步建模为 $\boldsymbol{p}_u+|I_u|^{-\frac{1}{2}}\sum y_i+|\boldsymbol{T}_u|^{-\frac{1}{2}}\sum\boldsymbol{w}_v$。在社会网络中,本文不仅考虑用户-项目评分矩阵,还要考虑用户与用户之间的信任矩阵[2,17]。总之,本文引入用户之间的信任关系以及 Frobenius 范数避免过拟合,经过整理之后获得的目标函数如下[16]:

$$\begin{aligned}\boldsymbol{L}=&\frac{1}{2}\sum_u\sum_{j\in I_u}(\hat{r}_{u,j}-r_{u,j})^2+\frac{\lambda_t}{2}\sum_u\sum_{v\in\boldsymbol{T}_u}(\hat{t}_{u,v}-t_{u,v})^2\\&+\frac{\lambda}{2}\sum_u|I_u|-\frac{1}{2}\boldsymbol{b}_u^2+\frac{\lambda}{2}\sum_j|U_j|-\frac{1}{2}\boldsymbol{b}_j^2\\&+\sum_u\left(\frac{\lambda}{2}|I_u|^{-\frac{1}{2}}+\frac{\lambda_t}{2}|\boldsymbol{T}_u|^{-\frac{1}{2}}\right)\|\boldsymbol{p}_u\|_{\mathrm{F}}^2\\&+\frac{\lambda}{2}\sum_j|U_j|-\frac{1}{2}\|\boldsymbol{q}_j\|_{\mathrm{F}}^2+\frac{\lambda}{2}\sum_i|U_i|-\frac{1}{2}\|y_i\|_{\mathrm{F}}^2\\&+\frac{\lambda}{2}|\boldsymbol{T}_v^+|^{-\frac{1}{2}}\|\boldsymbol{w}_v\|_{\mathrm{F}}^2\end{aligned} \tag{5}$$

其中,$\|\cdot\|_{\mathrm{F}}$ 表示 Frobenius 范数;λ 表示控制模型复杂度和防止过拟合的参数;U_j 和 U_i 分别是给项目 j 和项目 i 评分的用户集合;$\boldsymbol{T}_v^+$ 是指信任 v 的用户集合。

对于式(5)中的目标函数,可以通过在 $\boldsymbol{b}_u,\boldsymbol{b}_j,\boldsymbol{p}_u,\boldsymbol{q}_j,\boldsymbol{y}_i,\boldsymbol{w}_v$上使用梯度下降的方法进行求解,使其达到局部最小值:

$$\begin{aligned}\frac{\partial\boldsymbol{L}}{\partial\boldsymbol{b}_u}&=\sum_{j\in I_u}(\hat{r}_{u,j}-r_{u,j})+\lambda|I_u|^{-\frac{1}{2}}\boldsymbol{b}_u,\\\frac{\partial\boldsymbol{L}}{\partial\boldsymbol{b}_j}&=\sum_{u\in U_j}(\hat{r}_{u,j}-r_{u,j})+\lambda|U_j|^{-\frac{1}{2}}\boldsymbol{b}_j\\\frac{\partial\boldsymbol{L}}{\partial\boldsymbol{p}_u}&=\sum_{j\in I_u}(\hat{r}_{u,j}-r_{u,j})\boldsymbol{q}_j+\lambda_t\sum_{v\in\boldsymbol{T}_u}(\hat{r}_{u,j}-r_{u,j})\boldsymbol{w}_v\\&\quad+(\lambda|I_u|^{-\frac{1}{2}}+\lambda_t|\boldsymbol{T}_u|^{-\frac{1}{2}})\boldsymbol{p}_u\\\frac{\partial\boldsymbol{L}}{\partial\boldsymbol{q}_j}&=\sum_{u\in U_j}(\hat{r}_{u,j}-r_{u,j})(\boldsymbol{p}_u+|I_u|^{-\frac{1}{2}}\sum_{i\in I_u}y_i\\&\quad+|\boldsymbol{T}_u|^{-\frac{1}{2}}\sum_{v\in\boldsymbol{T}_u}\boldsymbol{w}_v)+\lambda|U_j|^{-\frac{1}{2}}\boldsymbol{q}_j\\\frac{\partial\boldsymbol{L}}{\partial\boldsymbol{y}_i}&=\sum_{j\in I_u}(\hat{r}_{u,j}-r_{u,j})\boldsymbol{q}_j|I_u|^{-\frac{1}{2}}+\lambda|U_i|^{-\frac{1}{2}}y_i\\\frac{\partial\boldsymbol{L}}{\partial y_i}&=\sum_{j\in I_u}(\hat{r}_{u,j}-r_{u,j})|\boldsymbol{T}_u|^{-\frac{1}{2}}\boldsymbol{q}_j\\&\quad+\lambda_t(\hat{r}_{u,j}-r_{u,j})\boldsymbol{p}_u+\lambda|\boldsymbol{T}_v|^{-\frac{1}{2}}\boldsymbol{w}_v\end{aligned} \tag{6}$$

3.3　SocialSVD 推荐算法

本文针对目前推荐算法所面临的问题和挑战提出一种推荐算法(SocialSVD)。该算法在基于已有的奇异值分解(SVD ++)模型的基础上加入用户的社交网络信息以提高推荐系统的效率,同时本文提

出信任传播算法(TPA)填充用户信任矩阵以解决社交网络矩阵的稀疏性问题。

SocialSVD 推荐算法的基本框架如图 2 所示,SocialSVD 推荐算法的主要输入是用户信任关系矩阵 $\boldsymbol{TR}$ 和用户评分矩阵 $\boldsymbol{R}$。该算法的主要步骤包括信任度计算、预测评分、产生推荐列表。

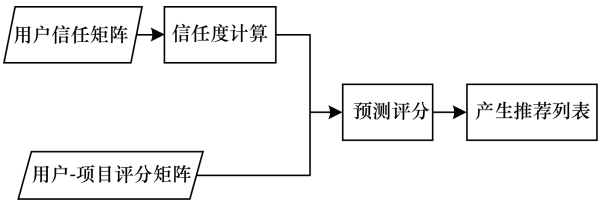


图 2 SocialSVD 推荐算法基本框架

SocialSVD 推荐算法的具体描述如下:

输入 当前用户 u,信任关系矩阵 $\boldsymbol{TR}$,用户-项目评分矩阵 $\boldsymbol{R}$,信任传播步长 $\boldsymbol{L}$,信任传播衰减率 $\boldsymbol{p}$

输出 用户推荐列表 N

1)使用信任传播算法计算得出信任度向量 $\boldsymbol{T}_u = TP(u,\boldsymbol{TR},\boldsymbol{L},\boldsymbol{p})$。

2)对于每个用户 u 未评分的项目集合中的每一个 i,预测评分 $\hat{r}_{u,j} = \boldsymbol{b}_u + \boldsymbol{b}_j + \mu + \boldsymbol{q}_j^{\mathrm{T}}(\boldsymbol{p}_u + |I_u|^{-\frac{1}{2}} \sum_{i \in I_u} y_i + |\boldsymbol{T}_u|^{-\frac{1}{2}} \sum_{v \in \boldsymbol{T}_u} \boldsymbol{w}_v)$。

3)更新用户推荐列表 N。

3.4 复杂度分析

SocialSVD 推荐算法的主要资源开销是计算目标函数 $\boldsymbol{L}$ 和变量的降阶过程,计算目标函数 $\boldsymbol{L}$ 的时间复杂度是 $O(d|\boldsymbol{R}| + d|\boldsymbol{TR}|)$,其中 d 表示 d 维特征向量。由于用户-项目矩阵 $\boldsymbol{R}$ 和用户信任矩阵 $\boldsymbol{TR}$ 的稀疏性,因此该时间复杂度的值要远低于矩阵基数。计算梯度下降的时间复杂度为 $O(d\boldsymbol{R}| + d|\boldsymbol{TR}|)$。此外,在 2.2 节算法 1 计算信任传播算法(TPA)的时间复杂度中已经给出由于信任矩阵稀疏性和信任传播步长 L 的限制,该信任度算法的复杂度远小于 $O(n^2)$,因此 SocialSVD 推荐算法的总时间复杂度与用户-项目矩阵和用户信任矩阵中的可观察实体数量成线性关系。

4 实验结果与分析

4.1 数据集描述

实验中使用的数据集是从 FilmTrust 网站爬取的 2011 年 7 月份的电影评价数据集,其中 FilmTrust 是一个电影共享和评分网站[12]。FilmTrust 数据集包括 1 508 个用户给 2 071 个电影的 35 497 条评分数据,而且 FilmTrust 数据集中包括 1 642 个用户的1 852 条信任关系数据,其中评分数据的稀疏度为1.136 607% 。同时本文使用的 Epinions 数据集是从 Epinions 网站上爬取的[16]。Epinions 网站是一个消费者发表意见和观点的网站,在该网站中用户能够通过数字评分或者写文本评论表达他们对于项目(例如车、电影、软件、书、音乐等)的观点,评分数值在1 ~5之间。同时,用户也可以通过判断其他用户的评论是否和自己的观点一致来决定该用户是否值得信任。Epinions 数据集包括 49 290 个用户对 139 738 个项目的 664 824 条评价,包括 487 181 条信任关系信息,评分数据的稀疏度为 0.011 845 84% 。

4.2 度量方法

本文使用 2 种典型的度量方法估计预测精度,即平均绝对误差(Mean Absolute Error,MAE)和均方根误差(Root Mean Square Error,RMSE),其中平均绝对误差和均方根误差的定义如下:

$$M_{\mathrm{MAE}} = \frac{\sum_{u,j \in N} |r_{u,j} - \hat{r}_{u,j}|}{N} \tag{7}$$

$$R_{\mathrm{RMSE}} = \sqrt{\frac{\sum_{u,j \in N} (r_{u,j} - \hat{r}_{u,j})^2}{N}} \tag{8}$$

其中,$\hat{r}_{u,j}$ 表示用户 u 对项目 j 的预测评分;$r_{u,j}$ 表示已有的用户评分;N 表示评分数量。MAE 值和 RMSE 值越小表示性能越好[22]。

4.3 信任衰减系数对实验结果的影响

本文在信任传播算法中提到的信任度大小随着信任传播长度增加而不断减小。因此,采用信任衰减系数(p)调整信任传播中信任的衰减率。实验中设置的信任衰减系数为 0.1 ~0.9,使用的测试数据集是 FilmTrust 数据集和 Epinions 数据集,分别进行 9 次信任传播实验将 MAE 值和 RMSE 值进行对比,结果如图 3 和图 4 所示。

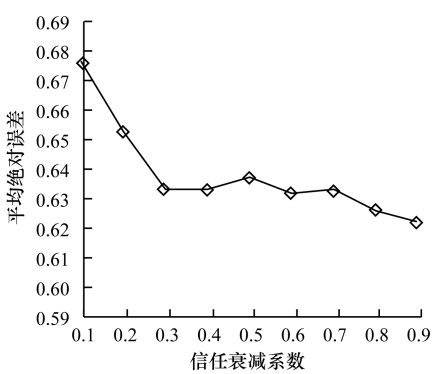


图 3 MAE 随信任衰减系数的变化情况

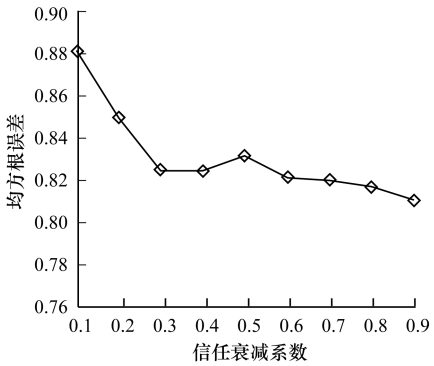


图 4　RMSE 随信任衰减系数的变化情况

从图 3、图 4 可以得出,当衰减系数 p 为 0.9 时,MAE 和 RMSE 值最小,因此,本文在实验中设置衰减系数 p 的值为 0.9。

4.4　结果分析

为了评估本文提出的 SocialSVD 算法的性能,将本文算法与 PMF[13],RSTE[10],SocialMF[18] 算法进行比较。

1)PMF:概率矩阵分解算法。该算法在推荐过程中只考虑了用户对项目的评分数据对推荐质量的影响,没有考虑用户之间的关联关系对推荐质量的影响。

2)RSTE:该算法的提出是为了解决数据稀疏问题。在该算法中给出了概率因素分析框架。该框架通过集合参数融合用户及其信任者的兴趣。

3)SocialMF:在 SoRec[6] 模型的基础上提出的一个新算法。该算法主要是为了凸显信任关系对提高算法推荐质量的影响作用。

在实验中,本文任意选择 80% 的 FilmTrust 数据集或者 80% 的 Epinions 数据集作为训练集,即任意选择 80% 的训练评分数据测试 20% 的评分数据。实验给出了五维和十维潜在特征向量的结果,如表 1 和表 2 所示。在实验中,最佳实验参数设置如下:1)PMF,$\lambda_u=\lambda_i=\lambda_b=0.01$;2)RSTE,$\lambda_u=\lambda_i=\lambda_b=0.001$,$\lambda_t=0.001$,在 Epinions 数据集中 $\alpha=0.4$,在 FilmTrust 数据集中 $\alpha=0.4$;3)SocialMF,$\lambda_u=\lambda_i=\lambda_b=0.001$,$\lambda_t=1$;4)SocialSVD,在 FilmTrust 数据集中 $\lambda_u=\lambda_i=\lambda_b=0.01$,$\lambda_t=0.01$,在 Epinions 数据集中 $\lambda_u=\lambda_i=\lambda_b=0.003$,$\lambda_t=0.002$。

表 1　五维特征向量实验结果对比

数据集	度量方法	PMF	RSTE	SocialMF	SocialSVD
FilmTrust	MAE	0.714	0.632	0.630*	0.622
	RMSE	0.949	0.814	0.811*	0.810
Epinions	MAE	0.979	0.905	0.843*	0.821
	RMSE	1.290	1.107	1.078*	1.059

表 2　十维特征向量实验结果对比

数据集	度量方法	PMF	RSTE	SocialMF	SocialSVD
FilmTrust	MAE	0.735	0.640	0.639*	0.622
	RMSE	0.968	0.833	0.827*	0.808
Epinions	MAE	0.928	0.857	0.849*	0.822
	RMSE	1.094	1.077*	1.130	1.060

从表 1 和表 2 可以得出,SocialMF 算法和 RSTE 算法由于加入了社交网络中用户之间的信任关系来提高推荐质量,因此在相同的维度条件下 SocialMF 算法和 RSTE 算法的 MAE 值和 RMSE 值都要比 PMF 算法小。这说明在推荐过程中用户之间的信任关系有利于改善推荐结果的质量。本文中提出使用信任传播和奇异值分解模型相结合的方法提高推荐系统的推荐质量,在 FilmTrust 和 Epinions 上的结果都比对比算法效果好,其中 * 标记了 3 种对比算法中性能较好的结果。这表明本文提出的融合用户信任关系和奇异值分解的推荐算法有利于提高推荐系统的推荐质量。

5　结束语

随着在线社交网络的兴起,推荐系统作为个性化服务领域中一个重要的研究分支越来越受到广大研究者的关注。针对用户信任矩阵中数据稀疏的问题,本文设计用户之间信任关系的传播规则,根据该信任传播规则计算用户之间的信任度,填充用户信任矩阵。在此基础上,提出一种融合用户之间信任传播算法和奇异值分解模型的推荐算法。该算法通过信任关系矩阵与用户评分矩阵的有效结合,提高推荐算法的推荐质量。从实验结果可以看出,信任传播在提高推荐系统的推荐质量方面具有重要作用。为凸显信任传播的重要性,本文仅考虑了信任信息对于信任建模和提高推荐系统的推荐质量的影响,而忽略了不信任信息对推荐质量的影响。因此,下一步将研究不信任信息对于推荐系统的影响,改进推荐质量。

参考文献

[1]　孟祥武,刘树栋,张玉洁,等. 社会化推荐系统研究[J]. 软件学报,2015,26(6):1356-1372.

[2]　郭　磊,马　军,陈竹敏,等. 一种结合推荐对象间关联关系的社会化推荐算法[J]. 计算机学报,2014,37(1):219-227.

[3]　Guo Guibing, Zhang Jie, Thalmann D. Merging Trust in Collaborative Filtering to Alleviate Data Sparsity and Cold Start[J]. Knowledge-Based Systems, 2014, 57: 57-68.

[4]　Bao Yang, Fang Hui, Zhang Jie. TopicMF: Simultaneously Exploiting Rating and Reviews for Recommendation[C]//

Proceedings of the 28th AAAI Conference on Artificial Intelligence. Québec City, Canada: AAAI Press, 2014: 1-5.
[5] Gupta P, Goel A, Lin J, et al. WTF: The Who to Follow Service at Twitter[C]//Proceedings of the 22nd International Conference on World Wide Web. New York, USA: ACM Press, 2013: 505-514.
[6] Ma Hao, Yang Haixuan, Lyu M R, et al. SoRec: Social Recommendation Using Probabilistic Matrix Factorization[C]//Proceedings of the 17th ACM Conference on Information and Knowledge Management. New York, USA: ACM Press, 2008: 931-940.
[7] Fang Hui, Bao Yang, Zhang Jie. Leveraging Decomposed Trust in Probabilistic Matrix Factorization for Effective Recommendation[C]//Proceedings of the 28th AAAI Conference on Artificial Intelligence. Québec City, Canada: AAAI Press, 2014: 30-36.
[8] Sun Zhu, Guo Guibing, Zhang Jie. Exploiting Implicit Item Relationships for Recommender Systems[C]//Proceedings of the 22nd International Conference on User Modeling, Adaptation and Personalization. New York, USA: ACM Press, 2015: 252-264.
[9] Yang Bo, Lei Yu, Liu Dayou, et al. Social Collaborative Filtering by Trust[C]//Proceedings of the 23rd International Jiont Conference on Artificial Intelligence. New York, USA: ACM Press, 2013: 2747-2753.
[10] Ma Hao, Zhou Dengyong, Liu Chao, et al. Recommender Systems with Social Regularization[C]//Proceedings of the 4th ACM International Conference on Web Search and Data Mining. New York, USA: ACM Press, 2011: 9-12.
[11] Ma Hao, King I, Lyu M R. Learning to Recommend with Social Trust Ensemble[C]//Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, USA: ACM Press, 2009: 203-210.
[12] Guo Guibing, Zhang Jie, Yorke-Smith N. A Novel Bayesian Similarity Measure for Recommender Systems[C]//Proceedings of the 23rd International Joint Conference on Artificial Intelligence. New York, USA: ACM Press, 2013: 2619-2625.
[13] Koren Y. Collaborative Filtering with Temporal Dynamics[C]//Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM Press, 2009: 447-456.
[14] Salakhutdinov R, Mnih A. Probabilistic Matrix Factorization[C]//Proceedings of the 21st Annual Conference on Neural Information Processing Systems. New York, USA: ACM Press, 2008: 1257-1264.
[15] Koren Y. Factorization Meets the Neighborhood: A Multifaceted Collaborative Filtering Model[C]//Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM Press, 2008: 426-434.
[16] Massa P, Avesani P. Trust-aware Bootstrapping of Recommender Systems[C]//Proceedings of ECAI'06. New York, USA: ACM Press, 2006: 29-33.
[17] Guo Guibing, Zhang Jie, Yorke S N. TrustSVD: Collaborative Filtering with Both the Explicit and Implicit Influence of User Trust and of Item Ratings[C]//Proceedings of the 29th AAAI Conference on Artificial Intelligence. New York, USA: ACM Press, 2015: 123-129.
[18] Jamali M, Ester M. A Matrix Factorization Technique with Trust Propagation for Recommendation in Social Networks[C]//Proceedings of the 4th ACM Conference on Recommender Systems. New York, USA: ACM Press, 2010: 135-142.
[19] 于洪涛,周 静,张付志. 融合信任传播和矩阵分解的协同推荐算法[J]. 燕山学报,2013,37(5):424-429.
[20] Wen Guanghui, Duan Zhisheng, Chen Guanrong. A Weighted Local-world Evolving Network Model with Aging Nodes[J]. Physica A: Statistical Mechanics & Its Applications, 2011, 390(21/22): 4012-4026.
[21] 沈益顺. 社交媒体快速发展背景下的网络暴力现象研究[J]. 新闻研究导刊,2015(24):62-62.
[22] 张 玲,王 磊,王姝媛. 基于聚类免疫网络的协同过滤推荐算法[J]. 计算机工程与应用,2008,44(27):141-144.

编辑 陆燕菲

(上接第 235 页)

[12] 杨胜文,史美林. 一种支持 QoS 约束的 Web 服务发现模型[J]. 计算机学报,2005,28(4):589-594.
[13] 何炎祥,吴 钊. 动态 Web 服务组合关键技术与性能分析[M]. 北京:清华大学出版社,2011.
[14] Wong K P, Yin W W. Combined Genetic Algorithm/Simulated Annealing/Fuzzy Set Approach to Short-term Generation Scheduling with Take-or-pay Fuel Contract[J]. IEEE Transactions on Power Systems, 1996, 11(1): 128-136.
[15] Al-Masri E, Mahmoud Q H. Investigating Web Services on the World Wide Web[C]//Proceedings of International Conference on World Wide Web. New York, USA: ACM Press, 2008: 795-804.
[16] 陈 亮,孙 敏. 基于免疫遗传算法的 Web 服务组合方法[J]. 计算机工程,2010,36(10):226-227.
[17] He Jun, Chen Liang, Wang Xiaolong, et al. Web Service Composition Optimization Based on Improved Artificial Bee Colony Algorithm[J]. Journal of Networks, 2013, 8(9): 2143-2149.

编辑 陆燕菲