

复杂网络环境下基于信任传递的推荐模型研究

李慧^{1,2} 马小平² 施琚¹ 李存华¹ 仲兆满¹ 蔡虹¹

摘要 针对推荐系统中普遍存在的数据稀疏和冷启动等问题, 本文结合用户自身评分与用户的社会信任关系构建推荐模型, 提出了一种基于信任关系传递的社会网络推荐算法 (Trust transition recommendation model, TTRM). 该方法首先通过计算信任网络中节点的声望值与偏见值来发现信任网络中的不可信节点, 并通过对其评分权重进行弱化来减轻其对信任网络产生的负面影响. 其次, 算法又利用朋友的信任矩阵对用户自身的特征向量进行修正, 解决了用户特征向量的精准构建及信任传递问题. 同时为了实现修正误差的最小化, 算法利用推荐特性进行用户相似度计算并通过带有社会正则化约束的矩阵分解技术实现社会网络推荐. 实验结果表明, TTRM 算法较传统的社会网络推荐算法在性能上具有显著提高.

关键词 社会网络, 推荐, 信任度, 矩阵分解, 正则化

引用格式 李慧, 马小平, 施琚, 李存华, 仲兆满, 蔡虹. 复杂网络环境下基于信任传递的推荐模型研究. 自动化学报, 2018, 44(2): 363–376

DOI 10.16383/j.aas.2018.c160395

A Recommendation Model by Means of Trust Transition in Complex Network Environment

LI Hui^{1,2} MA Xiao-Ping² SHI Jun¹ LI Cun-Hua¹ ZHONG Zhao-Man¹ CAI Hong¹

Abstract To deal with the data sparsity and cool boot problem, a new method by means of trust relations called trust transition recommendation model (TTRM), as well as user rating and users' social trust network, is proposed. The first step of the method is to spot the untrustworthy nodes in the trust network through their reputation and deviation values and abate their negative effects on trust network by weakening their rating weights. Secondly, the method revises the users' feature vector from their friends' trust matrix to solve the problems like users' feature vector accuracy establishment and trust transmission. Meanwhile, in order to minimize the round-off error, it calculates the similarity of users based on the recommendation features and realizes social network recommendation through matrix factorization with social regularization constraints. The results of experiments of TTRM on public dataset reveal that the new recommendation performare has been greatly improved compared to the traditional collaborative recommendation.

Key words Social network, recommendation, trust, matrix factorization, regulation

Citation Li Hui, Ma Xiao-Ping, Shi Jun, Li Cun-Hua, Zhong Zhao-Man, Cai Hong. A recommendation model by means of trust transition in complex network environment. *Acta Automatica Sinica*, 2018, 44(2): 363–376

作为个性化服务的一项重要技术, 推荐系统在最初所建立的用户兴趣偏好模型, 都是根据用户的

历史行为进行相应的数据分析, 进而计算资源与用户偏好模型的相似度, 发现用户可能感兴趣的项目集, 将相似度高的项目推荐给用户. 然而伴随着不断扩展的推荐应用场景, 仅仅依据用户与项目之间存在的二元关系为用户产生相关的推荐结果, 通常无法获得令人满意的推荐效果. 与传统的协同推荐算法不同, 社会网络推荐将丰富的信息资源和用户行为交互信息引入到推荐系统中, 该方法可以对现实社会关系进行很好的模拟, 从而真实反映了用户在现实世界中的社会活动.

到目前为止针对社会网络推荐系统的研究, 广大专家学者已做了许多的研究工作, 并且获得了丰富的研究成果. 许多研究是建立在对社会化网络的拓扑结构、用户间的信任关系以及项目在社会网络中的流行度分析的基础上, 从而提出各种社会网络推荐框架模型^[1–4]. Granovetter 引入链接强度来表示社会网

收稿日期 2016-05-13 录用日期 2016-10-09
Manuscript received May 13, 2016; accepted October 9, 2016
国家自然科学基金 (61403156, 61403155), 连云港市科技计划项目 (SH1507, CXY1530, CG1315, CG1413), 江苏高校品牌专业建设工程资助项目 (PPZY2015A038), 江苏省青蓝工程, 淮海工学院科研基金资助项目 (Z2017012, Z2015012) 资助
Supported by National Natural Science Foundation of China (61403156, 61403155), Science and Technology Planning Project of Lianyungang (SH1507, CXY1530, CG1315, CG1413), Top-notch Academic Programs Project of Jiangsu Higher Education Institutions (PPZY2015A038), Jiangsu Blue Project, Science Foundation of Huaihai Institute of Technology (Z2017012, Z2015012)

本文责任编辑 赵铁军
Recommended by Associate Editor ZHAO Tie-Jun
1. 淮海工学院计算机工程学院 连云港 222002 2. 中国矿业大学信息学院 徐州 221008
1. Department of Computer Science, Huaihai Institute of Technology, Lianyungang 222002 2. School of Information & Electrical Engineering, China University of Mining & Technology, Xuzhou 221008

络中用户之间的信任关系,用户之间的信任关系用连接关系的强弱来度量^[5]. Gilbert 等将连接强度的概念运用到社交网络的场景中,对于 Facebook 的数据集,作者将用户之间的信任关系映射为二元模式,即用户之间只存在信任或不信任的关系^[6].但在现实生活中,用户之间的信任程度是不同的,不能用简单的二元模式来描述. Xiang 等提出了一种无监督学习的方法来确定社会网络中信任关系的强度大小^[7]. 张燕平等通过对历史用户声誉构建声誉推荐系统,结合隐语言模型提出了基于用户声誉的鲁棒协同推荐算法^[8]. 但目前已有的方法在信任计算时,均将推荐信息不加区分的统一对待,从而导致一些恶意和虚假的推荐信息被加入到信任计算中,使信任推荐的精准度受到了影响.

针对推荐系统中普遍存在的数据稀疏和冷启动等问题,本文结合用户自身评分信息与社会信任关系两项要素,构建一种全新的以信任关系传递为基础的社会网络推荐方法 (Trust transition recommendation model, TTRM). 该方法充分考虑到不可信节点对推荐系统产生的不利影响,在对信任网络中不可信节点进行合理消除的基础上,进一步提升推荐结果的准确性. 本文的主要贡献体现在以下几个方面:

1) 计算信任网络中节点的偏见值与声望值,发现网络中的不可信节点,将该节点对其他节点的评价权重进行弱化,从而降低其对推荐系统整体的影响力,有效消除了不实评价的负面影响,进而提升推荐系统的准确率.

2) 鉴于用户的喜好会受其朋友的影响,算法又利用朋友的信任矩阵对用户自身的特征向量进行修正,解决了用户特征向量的精准构建及信任传递问题.

3) 为了实现修正误差的最小化,算法对已有的基于社会正则化约束的矩阵分解技术进行了修正,提高了社会网络推荐系统的准确度.

本文第 1 节介绍社会化推荐算法的相关工作. 第 2 节提出信任网络中不可信节点的发现算法. 第 3 节具体介绍了基于信任关系的社会网络推荐算法. 第 4 节是本文的实验部分,通过与传统基于信任的推荐方法的对比,分析我们提出算法的有效性. 第 5 节是本文的总结.

1 相关工作

协同过滤技术 (Collaborative filtering, CF) 被广泛用于构建推荐系统,并且现今大多数社会网络推荐系统也都是在协同过滤技术的基础上构建起来的. 社会网络推荐系统一般有两个输入,即用户评分信息和社会关系信息. 大多数现有的社会网络推荐系统选择协同过滤模型作为基本的模型来构建系统,

并提出基于社会网络分析结果来提炼所需信息的技术. 因此,一个基于协同过滤技术的社会网络推荐系统一般包含两部分: 一个基本的协同过滤模型和一个社会化信息模型. 这可以形式化的描述为

一个社会网络推荐的 CF 模型 =

一个基本的 CF 模型 + 一个社会化信息模型 (1)

根据 CF 推荐系统的类型,我们把社会网络推荐系统分为两类: 基于内存的和基于模型的社会网络推荐系统.

1.1 基于内存的社会网络推荐系统

在基于内存的协同过滤推荐算法的基础上,基于内存的社会网络推荐算法逐渐发展起来. 对于某一用户,缺失评分项可通过汇总与他相关用户的评分来获得,记为 N^+ . 给定某一用户,传统的面向用户的推荐系统采用相似用户的评分数据,而基于内存的社会网络推荐系统使用的是从用户评分信息和社会关系信息中获得的相关用户 N^+ 的数据. 基于内存的社会网络推荐系统通常遵循两个步骤. 第一步,对于特定的用户 $u(i)$,获取与其相关的用户集合 N^+ ; 第二步,与基于内存的协同过滤方法的最后一步相同,即汇总第一步获取的相关用户的评分,从而进行用户缺失评分项的预测. 在此类型的社会网络推荐系统中,不同的推荐算法使用不同的方法来获取第一步中的相关用户集合. 接下来,我们将详细介绍一些有代表性的方法.

TidalTrust 算法^[9]: 在这个算法中所有用户通过信任关系互相连接. TidalTrust 算法基于以下两个结论来评估用户间的信任值: 较短的传播路径产生更准确的信任估计值; 具有较高信任值的路径会产生更好的推荐结果.

MoleTrust 算法^[10]: 一种新的信任度量标准,该算法认为在信任网络中用户不可能获得全部用户都认可的信任值,因此提出了局部信任度的概念. MoleTrust 为了获得信任值,需要进行大量的信任传播机制. 因此,在信任网络建立前去除闭合关系可以确保网络中同一个结点仅被访问一次,从而可以大大提高算法的执行效率. 随着该操作的进行,原有的信任网络被转化为一个有向无环图. TrustWalker 算法^[11]: 本算法的提出源自两个重要的发现. 首先,当一个用户的社会网络信息与一个和他类似用户的社会网络信息存在小部分的重叠时,该重叠部分将提供了一个独立的社会信息来源. 其次,对同一目标项目,强信任好友的评分比弱信任好友更为可靠. 第一个现象表明以信任为基础的方法的重要性,第二个现象表明基于项目的推荐方法的有效性.

1.2 基于模型的社会网络推荐系统

以基于模型的协同过滤模型为基础,基于模型

的社会网络推荐系统得以发展完善. 矩阵分解技术被广泛应用于基于模型的协同过滤方法中. 矩阵分解技术存在如下优势^[12-13]: 1) 多种优化途径(如基于梯度的方法)可用于寻找局部最优解; 2) 矩阵分解对高斯噪声有很好的概率解释; 3) 矩阵分解方法是非常灵活的. 现有的大多数基于模型的社会网络推荐系统都是基于矩阵分解的. 这些方法背后的共同理论基础就是用户与其朋友的偏好是相似的或易受其在社会网络中相关联用户的影响. 然而, 社会关系形成的低成本优势会导致同质性社会关系的形成. 由于具有强关系的用户比弱关系的用户更可能具有相似的品味, 如果对等看待用户的各种社交关系势必会导致推荐性能的退化. 因此, 对于每一个社会关系, 基于模型的社会推荐方法将其与一个强度相关联, 这个强度通常是在现有的社交推荐系统中通过评分相似性计算而获得的.

与传统的基于矩阵分解的推荐系统相比, 基于模型的社会网络推荐系统的不同之处在于它可以充分利用各种类型的社会信息, 根据所利用的信息类型不同, 可以进一步将基于模型的社会网络推荐系统分为三类: 协同分解方法, 基于集合的方法, 正则化方法. 下面将分别介绍每种类型的代表系统. 协同分解方法: 此类推荐系统的假设前提是: 在评分空间(评分信息)和社交空间(社会信息)中用户应该共享相同的用户偏好向量. 采用此种类型的社会网络推荐系统能够以共享相同用户偏好潜在因子的方式来对用户-项目矩阵和用户-用户社会关系矩阵进行协同分解. SoRec^[14]和LOCABAL^[15]是此类推荐系统的代表系统. 基于集成的方法: 集成方法的基本思想是用户和他们的社会网络应该对项目具有相似的评分, 并且对于特定的用户, 缺失的评分项是利用社会评分网络中其他相关用户的评分信息的线性组合而进行预测的. STE^[16]和mTrust^[17]是此类推荐系统的代表系统.

正则化方法: 正则化方法关注用户的偏好, 并假设用户的偏好应该与他人在社会网络中的用户偏好类似. 对于特定的用户 u_i , 正则化方法要求他的偏好应该与他社会网络用户的偏好相似. SocialMF^[18]和Social regularization^[19]是此类方法中的两个代表系统.

2 不可信节点的发现算法

本文考虑了两种类型的信任网络. 第一种类型的信任网络中, 用户可以通过对其他用户给出正/负信任值来表达对其信任/不信任的程度. 此时, 当正值评分越大时, 表明具有越高的可信度, 当评分为0时, 表明持中立态度, 当评分为负值时表明具有较低的可信度. 这种类型的信任网络有Slashdot (www.slashdot.org)和Epinions

(www.epinions.com). 第二种类型的信任网络是规定用户只能给出非负的信任值. 在这种情况下, 评分越高表明信任度越高, 评分越低表明信任度越低. 这种类型的信任网络有IMDB (www.imdb.com)和Youtube (www.youtube.com).

2.1 相关定义

用有向带权图 $G = (V, E, W)$ 作为社会网络的抽象表述, 其中 V 代表网络中的所有节点集合, E 代表边的集合, 而用户之间的具体关系则用每节点之间的连边表示. 在网络图形中, 权重 w_{ij} 表示节点 i 指向节点 j 的信任值. 给定信任网络 G , 我们的目标是基于节点间的信任度计算各节点的偏见值和声望值. 节点的偏见值反映了该节点所给出的评价与其他节点的差异程度(用 b_i 来表示). 如果某个节点的偏见值为0, 则表示其给出的评价是没有偏差的, 可以完全被接受. 反之, 如果节点的偏见值较高, 说明其给出的评价与其他节点的差异度较大, 故不能被完全接受. 节点的声望值反映了其被其他节点所信任的程度(用 r_i 来表示). 有关节点偏见值与声望值的具体定义如下:

定义 1. 偏见值(偏离度): 在社会网络中, 某节点与其他节点之间的信任与不信任关系具体可以通过偏见值进行量化. 因此, 可以借助该节点对其他节点的评分与其应得评分之间的差异度来衡量该节点的偏见值. 若该差值越大, 则说明这个节点就越不可信. 节点 i 的偏见值 b_i 定义如下:

$$b_i = \frac{\theta}{|O_i|} \sum_{j \in O_i} (w_{ij} - r_j) \quad (2)$$

其中, $|O_i|$ 表示节点 i 相应的出度集合, 也就是以用户 i 为初始位置指向的所有节点集合. 参数 $\theta \in [0, 1]$, 在第4节给出了其取值的验证实验. 为了确保其取值范围在 $[-1, 1]$ 之间, 通常对偏见值进行标准化处理. 如果网络中的节点是真正可信赖的, 则该节点的偏见值为0.

定义 2. 不可信节点: 若某节点的偏见值表现为较高的正值, 且存在一条权值较大(权值符号为正)的出边, 或者网络中的节点偏见值表现为较高的负值, 并且存在一条权值较大(权值符号为负)的出边, 则将该节点定义为不可信节点. 若节点的偏见值为正值, 表明该节点表现出正面评价特征, 也就是网络图内该节点的某条出边权值为正, 反之也是如此. 如果一个节点对其他节点给出了其不应得的正值评分, 则该节点的偏见值将会增大. 如果网络中的节点具有较高的偏见值, 则其给出的评分权重应当被弱化, 从而减少其评分的重要程度. 为了达到这个目的, 通常采用减少该节点出边权值来降低影响. 若某一节点的偏见值与出边权值符号不相同, 即为异号时, 则无需对其进行任何修改. 直观地可以这样解释, 即当

一个人被认为通常会给出一个负面反馈时, 实际上他却给出了一个正面反馈, 则此时他的观点应该被重视. 因此, 当一个节点具有正 (负) 偏见值和一条负 (正) 值权重的出边时, 我们对这条边的权重不需要做任何一点的变化.

为了实现这一目标, 我们通过引进一个辅助参数 X_{kj} , 用以量化节点 k 的偏见值对节点 j 的每条出边权值的影响作用.

$$X_{kj} = \max\{0, b_k \times \text{sign}(w_{kj})\} \quad (3)$$

其中, $\text{sign}(w_{kj})$ 表示从节点 k 到节点 j 的信任值符号. 从上式可以得到, 若某一节点偏见值与其出边权值符号不相同, X_{kj} 的值为 0, 即并不会影响该节点的偏见值结果. 反之, 辅助参数 X_{kj} 的值为该偏见值的绝对值. 引入了辅助变量之后可以为网络中的节点重新计算新的信任权重 w_{kj}^* , 计算公式如下:

$$w_{kj}^* = w_{kj}(1 - X_{kj}) \quad (4)$$

由式 (4) 可知, 若节点出边的权值符号和该节点偏见值的符号为异号时, 则新的权重不会受到影响. 反之, 经式 (4) 调整后的权重将被弱化. 信任网络中节点应该得到的真实可信度可以用该节点的声望值来度量. 声望值可以通过偏见值来计算.

定义 3. 声望值: 在社会网络中, 某节点的声望值表示来自网络中其他可信节点对其入边的权重期望. 声望值的大小取决于入边的质量. 在对入边进行分析时, 可以采用调整权值的方式来减少偏见值的影响, 在去除了不可信节点之后计算全部入边的平均值. 现将 i 的声望值 r_i 定义如下:

$$r_i = \frac{1}{|I_i|} \sum_{j \in I_i} w_{ji}(1 - X_{ji}) \quad (5)$$

其中, $|I_i|$ 表示节点 i 的入度集合, 即以 i 为终点, 所有指向 i 的节点集合. 由式 (5) 可知, 声望值的取值范围在 $[-1, 1]$. 通过有关节点偏见值与声望值的定义, 使得构建的模型可以能够处理评价值为负值的信任网络, 并且可以将网络中权重为 0 (中立观点) 与无连接的情况 (图中无边相连) 加以区分. 如图 1 给出了一个包含 3 个节点的社会网络结构图, 边上的权值代表节点之间的信任度. 与现实情况相符, 图 1 中不

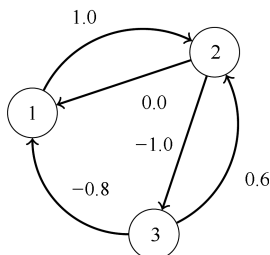


图 1 带有信任度的社会网络示例

Fig. 1 A example of social network with trust weights

是所有节点都对其他节点有评价, 因此社会网络关系通常是一个稀疏矩阵. 以该图所示的社会关系图为例, 下一节中将会给出计算该图中 3 个节点的偏见值与声望值的具体计算过程.

2.2 偏见值与声望值的计算

网络中节点的声望值和偏见值的具体计算方法将在本节给出. 由式 (2) 和式 (5) 可知, 声望值与偏见值的计算过程是交替迭代过程. 相邻节点的声望值决定该节点的偏见值, 而该节点的声望值则又由相邻节点的偏见值决定. 因此, 声望值与偏见值的具体求解, 可以通过不动点迭代的方法完成. 具体过程如下: b_i^t 和 r_i^t 分别表示节点 i 在第 t 次迭代得到的偏见值及声望值结果. 并以此为基础, 对该节点 $t+1$ 次迭代进行计算, 算法过程如下:

步骤 1. 首先以节点的偏见值与声望值为初值, 分别计算其他相邻节点在下一迭代后的声望值;

步骤 2. 根据上一步求得的声望值再重新计算该节点的偏见值. 从算法的执行过程可知 r_i^{t+1} 的结果取决于 b_i^t , 其值依次利用 r_i^t 可以得到求解. 因此, 式 (2) 和式 (5) 以如下形式重新定义:

$$b_i^{t+1} = \frac{\theta}{|O_i|} \sum_{j \in O_i} (w_{ij} - r_j^{t+1}) \quad (6)$$

$$r_j^{t+1} = \frac{1}{|I_j|} \sum_{k \in I_j} w_{kj}(1 - X_{kj}^t) \quad (7)$$

表 1 给出了由图 1 所示的网络示例图中三个节点的最终偏见值与声望值. 表 2 给出了图 1 所示的示例中每次迭代时各个节点所对应的偏见值与声望值的计算结果. 由于初始用户之间还没有产生信任交互, 用户之间的信任度最低, 因此将节点的偏见值与声望值初始时置为 -1 (表 2 中的第一行).

表 1 图 1 中各节点最终的偏见值与声望值

Table 1 Final bias and prestige values for the nodes in Fig. 1

	节点 1	节点 2	节点 3
偏见值 (Bias)	0.13	0.08	-0.14
声望值 (Prestige)	-0.33	0.73	-1.00

2.3 偏见值与声望值计算时间分析

对社会网络中任意节点在每一次迭代时计算其偏见值与声望值的计算复杂度可以用 $O(|d^0||d^i|)$ 来表示, 其中 $|d^0|$ 和 $|d^i|$ 表示所有节点的平均出度和入度. 虽然这样的计算复杂度看起来代价非常高, 但分摊后的计算代价可以逐渐降低到 $O(m)$, 其中 m 为整个网络中的边数总和. 每次新的迭代 t 开始时,

表 2 图 1 中每次迭代各节点的偏见值与声望值
Table 2 Bias and prestige values after each iteration in Fig. 1

迭代次数	节点 1		节点 2		节点 3	
	偏见值	声望值	偏见值	声望值	偏见值	声望值
0	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
1	0.10	0.00	-0.25	0.8	-0.25	0.00
2	0.12	-0.30	0.01	0.75	-0.16	-0.75
3	0.13	-0.33	0.08	0.73	-0.15	-1.00
4	0.13	-0.33	0.08	0.73	-0.14	-1.00
5	0.13	-0.33	0.08	0.73	-0.14	-1.00

首先, 利用式 (6) 计算所有节点的 r_i^t . 如果 m 表示网络中的所有边数, 则为所有节点计算声望值的时间代价为 $O(m)$. 基于这些声望值, 再利用式 (7) 可以求得所有结点的偏见值 b_i^t , 这一计算过程的分摊时间代价仍然为 $O(m)$. 因此, 对于 k 次迭代, 总的的时间代价为 $O(km)$. 只有当图结构非常紧密时, 计算时间代价可以增大到 $O(km^2)$. 然而, 对于多数的信任网络如 Slashdot、Epinions、IMDB、Youtube 等网站, 其网络中的结点数与边数均为线性关系, 因此整个网络计算所有节点的偏见值与声望值的总体时间代价为 $O(km)$. 通过以上计算可以求得网络中任意节点的偏见值与声望值, 节点偏见值大小与其可信度之间的关联呈反向变动关系, 即偏见值越小, 表明节点可信度越高. 基于以上判断, 可以借助式 (4) 实现对节点的权重进行弱化, 即减少了不可信节点对信任网络产生的负面影响. 在此基础上再进行推荐, 可以使推荐系统的推荐精度与用户对推荐结果的满意度得到提高.

2.4 迭代算法的收敛性证明

不可信节点的发现算法中节点的偏见值与声望值的计算是一个反复迭代的过程, 因此本节给出算法收敛性的证明. 由于节点的声望值可以通过其他节点的偏见值来求得, 这说明节点的声望值与偏见值具有相同的收敛性, 因此下面仅给出节点声望值的收敛性验证过程.

定理 1. 对于信任网络的任意节点 i , 均有 $|r_i^{t+1} - r_i^t| \leq \theta^t \|r^1 - r^0\|_\infty$ 成立

证明. 当迭代次数 $t = 1$ 时, 则

$$\begin{aligned}
 |r_i^2 - r_i^1| &= \left| \frac{1}{|I_i|} \sum_{j \in I_i} w_{ji} (X_{ji}^0 - X_{ji}^1) \right| \leq \\
 &\frac{1}{|I_i|} \sum_{j \in I_i} w_{ji} |X_{ji}^0 - X_{ji}^1| \leq \\
 &\frac{\theta}{|I_i|} \sum_{j \in I_i} w_{ji} \|r^1 - r^0\|_\infty \leq \\
 &\theta \|r^1 - r^0\|_\infty \quad (8)
 \end{aligned}$$

设 $t = k$ 成立, 则 $t = k + 1$ 时

$$\begin{aligned}
 |r_i^{k+2} - r_i^{k+1}| &= \left| \frac{1}{|I_i|} \sum_{j \in I_i} w_{ji} (X_{ji}^k - X_{ji}^{k+1}) \right| \leq \\
 &\frac{1}{|I_i|} \sum_{j \in I_i} w_{ji} |X_{ji}^k - X_{ji}^{k+1}| \leq \\
 &\frac{\theta}{|I_i|} \sum_{j \in I_i} w_{ji} \|r^{k+1} - r^k\|_\infty \leq \\
 &\theta \|r^{k+1} - r^k\|_\infty \leq \\
 &\theta^{k+1} \|r^1 - r^0\|_\infty \quad (9)
 \end{aligned}$$

□

3 基于信任关系的社会网络推荐模型

协同过滤等传统推荐算法仅仅考虑了用户-项目评分矩阵所提供的信息, 但忽略了社会信任关系的分析. 事实上与卖家推荐相比, 来自信任朋友的推荐常常更容易让用户所接受. 因此, 本节综合考虑用户信任关系与用户自身的评分信息, 创建一种利用矩阵分解与社会正则化技术实现基于可信度分析的社会化推荐算法, 以提高社会化推荐系统的准确性.

3.1 问题定义

设 $\{u_1, \dots, u_N\}$ 代表用户集合, $\{v_1, \dots, v_M\}$ 代表项目集合, $[R_{u,v}]_{N \times M}$ 代表用户对项目的评分矩阵, $R_{u,v}$ 代表用户 u 对项目 v 的实际评分. 由于通常情况下, 用户进行评分时一般采用 1~5 分的评分标准来说明自身的喜好情况, 因此本文也将用户对项目的评分结果映射到区间 $[0, 1]$. 由前文论述可知, 在信任网络中, 对于具有高偏见值的节点其给出的评价重要性应当被降低. 因此, 可以利用式 (4) 求得的修正信任度 (记为 w_{kv}^*) 来进行信任度矩阵 B 的构建. 本文具体通过概率矩阵分解技术^[20-21] 对用户-项目评分矩阵 $R_{u,v}$ 与信任关系矩阵 B 进行联合分解, 从而获得用户 u 对项目 v 评分的预期值. 矩阵分解技术通过矩阵分解获得用户、项目两项因素的低维潜在特征矩阵. 将一个稀疏的评分矩阵 R 分解为两个矩阵 U 和 V 就是矩阵分解的目标, 其中 U 和 V 分别表示用户特征矩阵和项目特征矩阵, 这些特征是描述用户和推荐对象的关键因素. 分别取

上述矩阵的一行与一列向量进行内积处理, 就能够获得用户对项目评分的预测值^[22].

$$R \approx U^T V \quad (10)$$

在现实中, 由于每个用户对项目的评分较少, 因而导致用户-项目评分矩阵成为一个稀疏矩阵. 传统地使用奇异值分解 (Singular value decomposition, SVD) 方法使评分预测误差最小, 即

$$\frac{1}{2} \|R - U^T V\|_F^2 \quad (11)$$

其中, $\|\cdot\|$ 表示 Frobenius 范数. 但是由于评分矩阵包含大量的缺失值, 我们只需要对矩阵中的评分值进行分解, 因此可以将上式变形为

$$\min_{U, V} \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n I_{ij} (R_{ij} - U_i^T V_j)^2 \quad (12)$$

其中, I_{ij} 是一个指示函数, 如果用户 u_i 对项目 v_j 有评分, 则其值为 1; 否则, 为 0. 为了避免过拟合, 在上式中引入两个正则化项, 可得:

$$\min_{U, V} \left(\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n I_{ij} (R_{ij} - U_i^T V_j)^2 + \frac{\lambda_1}{2} \|U\|_F^2 + \frac{\lambda_2}{2} \|V\|_F^2 \right) \quad (13)$$

其中, $\lambda_1 > \lambda_2 > 0$. 式 (13) 的最优解问题相当于最小化带二次正则项的误差平方和函数. 运用梯度下降法^[23] 可以找到局部最优解, 并且具有很好的概率解释, 因此成为协同过滤研究中的主流算法.

3.2 用户-项目评分矩阵分解

本文借助因式分解方法完成用户-项目评分矩阵 R 的分解与降维, 在此基础上分析用户的潜在特征. 将用户-项目评分矩阵分解成低维的用户特征矩阵 U 和项目特征矩阵 V 是矩阵分解的主要思想. 假设 m 个用户对 n 个项目完成的评分构成用户的评分矩阵 R , R_{ij} 代表用户 u_i 对项目 v_j 的实际评分, $U \in R^{l \times m}$ 及 $V \in R^{l \times n}$ 分别表示分解得到的用户与项目特征矩阵, l 代表用户特征数. 列向量 U_i 、 V_j 表示用户 U_i 、项目 V_j 的潜在特征向量. 为了让预测评分 $U^T V$ 与实际评分 R 的误差达到最小, 定义评分矩阵 R 的条件概率分布满足:

$$p(R|U, V, \sigma_R^2) = \prod_{i=1}^m \prod_{j=1}^n [N(r_{ij}|g(U_i^T V_j), \sigma_R^2)]^{I_{ij}^R} \quad (14)$$

其中, $N(x|\mu, \sigma^2)$ 表示参数 x 服从期望为 μ 、标准差为 σ^2 的高斯分布. I_{ij}^R 表示指示函数, 其取值 0 或 1 分别表示用户 u_i 对项目 v_j 没有或已经做出评分.

函数 $g(x) = 1/(1 + \exp(-x))$ 的作用是将预测评分限定在 $[0, 1]$ 区间. 为了防止过拟合, 用户与项目特征矩阵仍然假设服从均值为 0 的球形高斯先验, 如式 (15) 和式 (16) 所示:

$$p(U|\sigma_U^2) = \prod_{i=1}^m N(U_i|0, \sigma_U^2 I) \quad (15)$$

$$p(V|\sigma_V^2) = \prod_{j=1}^n N(V_j|0, \sigma_V^2 I) \quad (16)$$

由贝叶斯定理可知, 在已知部分用户对商品的评分信息的前提下, 求解其分解得到的用户和商品特征矩阵 U 和 V 的后验概率满足式 (17), 当该式取最大值时便可以得到最佳的 U 和 V , 使得预测评分与实际评分的误差最小:

$$\begin{aligned} p(U, V|R, \sigma_R^2, \sigma_U^2, \sigma_V^2) &\propto \\ &p(R|U, V, \sigma_R^2) p(U|\sigma_U^2) p(V|\sigma_V^2) = \\ &\prod_{i=1}^m \prod_{j=1}^n [N(r_{ij}|g(U_i^T V_j), \sigma_R^2)]^{I_{ij}^R} \times \\ &\prod_{i=1}^m N(U_i|0, \sigma_U^2 I) \times \prod_{j=1}^n N(V_j|0, \sigma_V^2 I) \end{aligned} \quad (17)$$

式 (17) 中以用户已有评价信息为基础, 对其潜在特征进行评估预测, 但在社会网络中, 用户会经常采纳其朋友 (所信任的用户) 的推荐. 这是由于一般来说, 用户与其朋友会有共同的喜好, 当其对朋友的信任值越高, 其推荐的成功率就越大. 因此在下一节中介绍如何利用信任用户信息来提高推荐的准确率.

3.3 结合信任关系与评分矩阵的社会化推荐

社会关系网络中, 邻居节点 $\Gamma(u)$ 常常会对用户 u 的行为与喜好产生影响, 因此来自用户所信任好友的推荐则更容易被用户所接受. 也就是说, 相邻节点 $v \in \Gamma(u)$ 会对用户 u 的潜在特征向量产生特定影响. 该影响作用可以用下式进行表示:

$$\hat{U}_u = \sum_{v \in \Gamma(u)} w_{uv}^* U_v \quad (18)$$

其中, \hat{U}_u 表示对任意用户 u , 已知其邻居节点 $v \in \Gamma(u)$ 特征向量的前提下, 对用户 u 特征向量的估计值. w_{uv}^* 是利用式 (4) 求得的修改后的用户信任权值. 从式 (18) 可以看出, 某用户潜在特征向量的估计值是其直接相邻用户的特征向量的权重平均值. 因此上式可以表示为

$$\begin{pmatrix} \hat{U}_{u,1} \\ \hat{U}_{u,2} \\ \vdots \\ \hat{U}_{u,m} \end{pmatrix} = \begin{pmatrix} U_{u,1} & U_{2,1} & \cdots & U_{n,1} \\ U_{u,2} & U_{2,2} & \cdots & U_{n,2} \\ \vdots & \vdots & \vdots & \vdots \\ U_{1,K} & U_{2,K} & \cdots & U_{n,m} \end{pmatrix} \begin{pmatrix} w_{u,1}^* \\ w_{u,2}^* \\ \vdots \\ w_{u,m}^* \end{pmatrix}$$

(19)

这里虽然考虑了社会网络中朋友的喜好可以影响用户特征向量的构建, 但这并不会改变用户-项目评分矩阵 R 的条件概率分布. 并且用户 u 的特征向量可以反映其所信任朋友特征向量的条件概率分布, 即

$$p(U|B, \sigma_U^2, \sigma_B^2) \propto P(U|\sigma_U^2)P(U|B, \sigma_B^2) = \prod_{i=1}^m N(U_i|0, \sigma_U^2 I) \times \prod_{i=1}^m N(U_i | \sum_{k \in \Gamma(i)} w_{ik}^* U_k, \sigma_C^2 I) \quad (20)$$

上式可以看作是两个正态分布的乘积因此结果仍然符合正态分布, 这样即可以保证用户特征向量不会太大并且与其邻居节点的特征向量非常相近. 同样为了防止过拟合, 对于项目特征向量也假设服从均值为 0 的高斯分布, 即

$$p(V|\sigma_V^2) = \prod_{j=1}^n N(V_j|0, \sigma_V^2 I) \quad (21)$$

运用贝叶斯推理可知:

$$p(U, V|R, B, \sigma_B^2, \sigma_U^2, \sigma_V^2) \propto p(R|B, U, V, \sigma_B^2) p(U|B, \sigma_U^2) p(V|B, \sigma_V^2) \quad (22)$$

仍然假设 B 与低维矩阵 U 和 V 是相互独立的, 则式 (22) 可表示为

$$p(U, V|R, B, \sigma_B^2, \sigma_U^2, \sigma_V^2) \propto p(R|U, V, \sigma_B^2) p(U|B, \sigma_U^2) p(V|\sigma_V^2) = \prod_{i=1}^m \prod_{j=1}^n [N(R_{ij}|g(U_i^T V_j), \sigma_B^2)]^{I_{ij}^R} \times \prod_{i=1}^m N(U_i | \sum_{k \in \Gamma(i)} w_{ik}^* U_k, \sigma_U^2 I) \times \prod_{i=1}^m N(U_i|0, \sigma_U^2 I) \times \prod_{j=1}^n N(V_j|0, \sigma_V^2 I) \quad (23)$$

其中, $p(U|\sigma_U^2)$ 与 $p(V|\sigma_V^2)$ 分别表示用户与项目特征矩阵仍然假设服从均值为 0 的球形高斯先验. 式 (18) 可以理解为仅依靠用户所信任朋友的品味进行推荐, 其对应的概率图模型可以用图 2 表示. 通过该模型, 可以很好地解决初始信任不足的情况. 因为在社交推荐中, 有不少用户的信任用户信息不充分, 因此给基于信任度的推荐工作带来了信任值缺失等问题, 称为信任冷启动问题. 通过该模型用户间的信任计算可以通过相邻用户传递下去, 因此可以很有效解决信任冷启动问题.

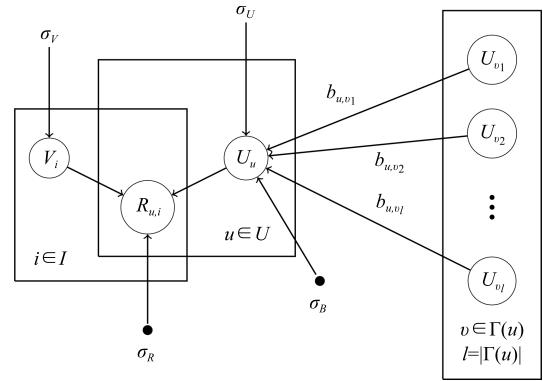


图 2 基于用户信任关系推荐的概率图模型

Fig. 2 Graphic model for recommendation based on trust relationship

给定上述推荐模型后, 其对数联合后验概率分布可以表示为

$$\begin{aligned} \ln p(U, V|R, B, \sigma_B^2, \sigma_U^2, \sigma_V^2) = & -\frac{1}{2\sigma_B^2} \sum_{i=1}^m \sum_{j=1}^n I_{ij}^R (R_{ij} - g(\alpha U_i^T V_j))^2 - \\ & \frac{1}{2\sigma_U^2} \sum_{i=1}^m U_i^T U_i - \frac{1}{2\sigma_V^2} \sum_{j=1}^n V_j^T V_j - \\ & \frac{1}{2\sigma_B^2} \sum_{i=1}^m \left(\left(U_i - \sum_{k \in \Gamma(i)} w_{ik}^* U_k \right)^T \left(U_i - \sum_{k \in \Gamma(i)} w_{ik}^* U_k \right) \right) - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n I_{ij}^R \ln \sigma_B^2 - \\ & \frac{1}{2} (ml \ln \sigma_U^2 + nl \ln \sigma_V^2 + ml \ln \sigma_B^2) + \epsilon \end{aligned} \quad (24)$$

其中, ϵ 为与参数无关的常数. 在参数 U, V 固定时, 求解式 (24) 对数后验概率的最大值问题可以转换为对以下带二次正则项的误差平方和函数求最小值问题:

$$\begin{aligned} E(U, V, R, B) = & \frac{1}{2} \prod_{i=1}^m \prod_{j=1}^n I_{ij}^R (R_{ij} - g(U_i^T V_j))^2 + \\ & \frac{\lambda_U}{2} \sum_{i=1}^m U_i^T U_i + \frac{\lambda_V}{2} \sum_{j=1}^n V_j^T V_j + \\ & \frac{\lambda_B}{2} \sum_{i=1}^m \left(\left(U_i - \sum_{k \in \Gamma(i)} w_{ik}^* U_k \right)^T \times \right. \\ & \left. \left(U_i - \sum_{k \in \Gamma(i)} w_{ik}^* U_k \right) \right) \end{aligned} \quad (25)$$

其中, $\lambda_U = \sigma_R^2/\sigma_U^2$, $\lambda_V = \sigma_R^2/\sigma_V^2$, $\lambda_B = \sigma_R^2/\sigma_B^2$, $\|\cdot\|_F^2$ 表示 Frobenius 范数. 上式中, 仅考虑了用户所信任的朋友会对用户产生影响, 但在现实中用户所信任的朋友具有的兴趣爱好未必与用户相同. 特别是在社会网络环境中, 用户可能有成百上千个朋友, 在众多朋友中, 有些朋友的喜好与用户相似, 但另外一些朋友的喜好可能和目标用户完全不同. 如果用户与其朋友的喜好差异过大, 反而会导致求解用户特征向量的准确度下降. 为解决这样的问题, 我们通过引入正则化项 (如式 (26) 所示) 对用户与朋友的偏好差异水平进行控制. 推荐的目标就是发现那些用户所信任并与其具有相似偏好的用户集合, 这些人的推荐才会更有可能被用户所接受. 因此, 为了防止过拟合, 提高泛化能力, 我们在上述推荐模型中引入下式的正则化项:

$$\frac{\alpha}{2} \sum_{i=1}^m \sum_{k \in \Gamma(i)} sim(i, k) \|U_i - U_k\|_F^2 \quad (26)$$

将该正则化项引入到原始目标函数 (式 (25)) 对误差函数进行约束, 从而达到优化误差函数的目的.

$$\begin{aligned} \min_{U, V} E^*(U, V, R, B) = & \\ & \frac{1}{2} \prod_{i=1}^m \prod_{j=1}^n I_{ij}^R (R_{ij} - g(U_i^T V_j))^2 + \\ & \frac{\lambda_U}{2} \sum_{i=1}^m U_i^T U_i + \frac{\lambda_V}{2} \sum_{j=1}^n V_j^T V_j + \\ & \frac{\alpha}{2} \sum_{i=1}^m \sum_{k \in \Gamma(i)} sim(i, k) \|U_i - U_k\|_F^2 + \\ & \frac{\lambda_C}{2} \sum_{i=1}^m ((U_i - \sum_{k \in \Gamma(i)} w_{ik}^* U_k)^T \times \\ & (U_i - \sum_{k \in \Gamma(i)} w_{ik}^* U_k)) \end{aligned} \quad (27)$$

其中, $sim(i, k) \in [0, 1]$ 表示相似度函数, 用来度量用户 u_i 和用户 u_k 之间的相似度大小. 在该推荐模型中利用推荐特性进行用户相似度的计算. 我们主要考虑了两类推荐特性: 产品的频率与声望. 令 fre 和 rep 分别代表产品的购买频率与声誉, 为了计算用户 u_i 和 u_k 之间的相似度, 我们应用下式计算:

$$sim(i, k) = \frac{\sum_{j \in I_i \cap I(k)} \sqrt{\frac{1}{rep_j^2} + \frac{1}{fre_j^2}} (R_{ij} - \bar{R}_i)(R_{kj} - \bar{R}_k)}{\sigma_i \sigma_k} \quad (28)$$

其中, fre_j 和 rep_j 分别表示产品 j 的购买频率与声望值 (产品的购买频率在购物网站中是很容易获取

的, 产品的声望值已经在第 2.2 节中求得). \bar{R}_i 表示用户 i 的平均评分. $I(i)$ 表示用户 i 的评分项目集合. $sim(i, k)$ 的值越大表明用户之间的喜好越相似. σ_i 与 σ_k 分别表示用户 i 与用户 k 的标准方差, 其计算公式如下所示:

$$\begin{aligned} \sigma_i &= \sqrt{\sum_{j \in I_i \cap I(k)} (R_{ij} - \bar{R}_i)^2} \\ \sigma_k &= \sqrt{\sum_{j \in I_i \cap I(k)} (R_{kj} - \bar{R}_k)^2} \end{aligned} \quad (29)$$

对于目标函数 E^* (式 (27) 所示) 求解局部最小值, 可以通过对所有用户和项目在 U_i 和 V_j 上使用梯度下降法获得局部极小值:

$$\begin{aligned} \frac{\partial E}{\partial U_i} &= \sum_{j=1}^n I_{ij}^R g'(U_i^T V_j) (g(U_i^T V_j) - R_{ij}) V_j + \\ & \lambda_U U_i + \alpha \sum_{k \in \Gamma(i)} sim(i, k) (U_i - U_k) + \\ & \alpha \sum_{g \in B_i} sim(i, g) (U_i - U_g) + \\ & \lambda_B (U_i - \sum_{k \in \Gamma(i)} w_{ik}^* U_k) - \\ & \lambda_B \sum_{\{k | i \in \Gamma(k)\}} w_{ik}^* (U_k - \sum_{g \in \Gamma_k} w_{gk}^* U_g) \end{aligned} \quad (30)$$

其中, $g'(x)$ 为函数 $g(x)$ 的导数, 即 $g'(x) = \exp(x)/(1 + \exp(x))^2$, $B(x)$ 表示所有信任用户 u_i 的用户集合. 为了减少算法的复杂度, 令 $\lambda_U = \lambda_V$. 参数 α 的作用是控制推荐模型中融入朋友喜好的权重, 此值越大, 说明在推荐中以朋友的喜好为主, 反之, 此值越小, 表明推荐中以用户自身的喜好为主. 当然, 为了获得最佳的推荐效果, 应当兼顾用户自身与朋友两方面的信息. 在第 4 节我们给出了参数 α 取值的验证实验.

3.4 算法分析

算法主要的计算代价在于目标函数 E^* 与用户和项目特征空间向量的梯度学习. 令每个用户的平均评分次数用 \bar{r} 表示, 每个用户的平均邻居个数用 \bar{k} 表示, 则计算目标函数的复杂度为 $O(m\bar{r}l + m\bar{k}l)$. 由于用户-项目评分矩阵 R 和信任度矩阵 B 都是稀疏矩阵, 因此目标函数的计算还是非常快的, 与社会评分用户数量成线性关系. 计算梯度的复杂度为 $O(m\bar{r}l + m\bar{k}^2l)$, 也是与社会评分用户数量成线性关系. 文献 [16] 中提出的 STE 方法中计算复杂度 $O(m\bar{r}\bar{k}^2l)$ 相比, 本文使用的算法复杂度更低, 性能更好.

4 实验与结果分析

为了验证本文提出的 TTRM 算法的有效性, 本节将其与现有的推荐方法在现实公开的包含信任关系的社交网络上进行对比实验, 实验还对模型中所涉及的参数进行验证实验。

4.1 数据集与评价指标

本文验证环节具体以 Epinions、Slashdot 这两项现实数据集为基础。上述数据集均可以从 SNAP (<http://snap.stanford.edu>) 获取。Epinions 是专业的产品评论网站, 拥有大规模的用户群体。该网站提供了产品评分与评价功能, 并在所有用户对某一产品的全部评分基础上, 整理并挑选出最有权权威性的评论。数据集中收录了从 1999 年至 2004 年的所有评论数据, 总共包含 841 000 条边与 119 217 个节点信息。实验挑选了其中评分信息较为完整的 51 670 个用户和其对 83 509 个项目的具体数据。在此基础上构建而成的用户-项目评分矩阵的密度不高于 0.015%。Slashdot 则是技术类新闻站点, 基于其独创的 Slashdot Zoo 技术, 所有注册用户可以评估自身与他人关系并进行标记。从网站上获取了从 2009 年开始的包含 549 202 条边和 82 144 个用户的数据后, 我们发现其中 70 284 个节点至少存在一条带符号的边, 而且其中 77.4% 是正数。实验使用平均绝对误差 (Mean absolute error, MAE)、均方根误差 (Root mean square error, RMSE) 及用户均方根误差 (User root mean square error, URMSE) 作为算法性能的评价指标。MAE 表示所有单个观测值与算术平均值的偏差的绝对值, 该值越小, 表明推荐算法的性能越好, 其定义如下:

$$\text{MAE} = \frac{\sum_{i,j} |r_{i,j} - \bar{r}_{i,j}|}{N} \quad (31)$$

$r_{i,j}$ 的含义为用户 i 向项目 j 做出的具体评分, $\bar{r}_{i,j}$ 的含义为基于应用推荐模型, 用户对项目评分的预期值, N 为评分样本个数。RMSE 表示观测值与真值偏差的平方与观测次数 N 比值的平方根, 该值越小, 表明推荐算法的性能越好, 其定义如下:

$$\text{RMSE} = \sqrt{\frac{\sum_{i,j} (r_{i,j} - \bar{r}_{i,j})^2}{N}} \quad (32)$$

URMSE 用来评估推荐算法针对不同用户所表现出来的推荐效果。MAE 和 RMAE 均只能衡量算法对所有用户的推荐性能, 但 URMSE 可以用来衡量推荐算法对每个用户的推荐效果, 其定义如下:

$$\text{URMSE} = \frac{\sum_{u_k \in U} \sqrt{\frac{\sum_{i=1}^{n_k} (r_{i,j} - \bar{r}_{i,j})^2}{n_k}}}{|U|} \quad (33)$$

其中, n_k 表示第 K 个用户需要预测的评价数量, $|U|$ 表示用户总数。

4.2 不可信节点的验证实验

在计算信任网络中节点偏见值的式 (2) 中参数 $\theta \in [1, 0)$, 下面进行参数 θ 取值的验证实验。首先, 我们使用节点 i 的信任方差来度量其偏见值求解的准确度。有关节点方差的定义如下:

$$\text{var}(i) = \frac{\theta}{|O_i|} \sum_{j \in O_i} (w_{ij} - \bar{r}_j)^2 \quad (34)$$

其中, $\bar{r}_j = \frac{1}{|I_j|} \sum_{i \in I_j} w_{ij}$, 将网络中各结点按其方差值进行排名, 并将该排名作为“基准排名”。如果网络中节点的方差值越大说明该结点的偏见值越高。然后在参数 θ 取不同值下应用 TTRM 算法求解各个节点的偏见值, 最后对求得的偏见值排名与“基准排名”计算 Kendall Tau 距离^[24]。Kendall Tau 距离常被用来评估应用某算法求得的排名与基准排名之间的相关性, 该相关性越大说明算法性能越好, 图 3 给出了参数验证结果。由实验结果可知, 当参数 $\theta = 1/2$ 时, 算法性能最好。

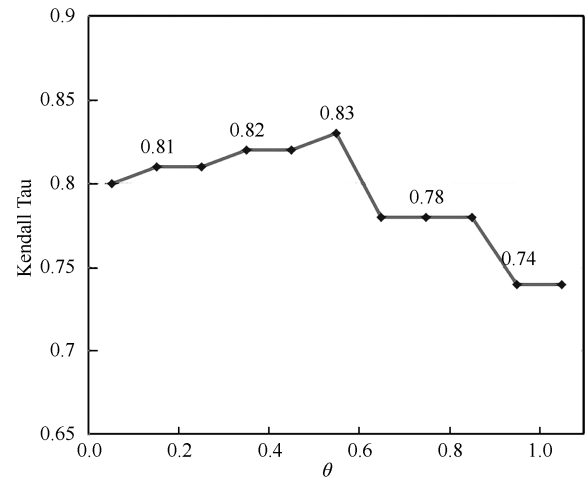
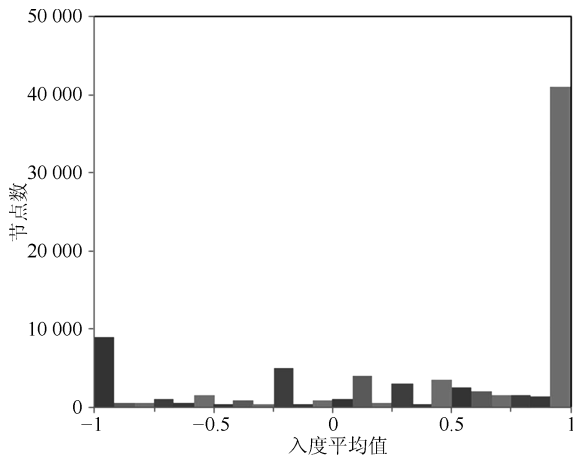


图3 参数 θ 的取值验证实验

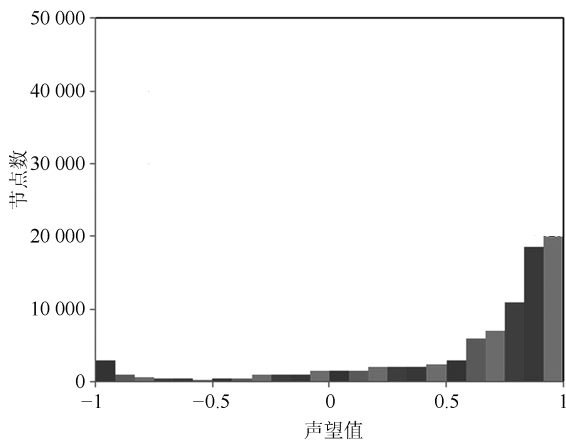
Fig. 3 Verification experiment of parameters θ

TTRM 算法的主要工作是基于社会网络中的信任关系产生推荐, 因此算法的第一步是通过识别并弱化网络中的不可信节点, 对社会网络的信任矩阵进行优化, 从而提高推荐系统的性能。首先对不可信节点识别的有效性进行验证。具体通过节点入度平均值 (所有入边权重的期望值) 与声望值分布状况之间的影响关系进行评估分析。图 4 给出了在 Slashdot 数据集上的入度平均值与声望值的分布结果。从图 4 可以看出, 在 Slashdot 数据集上, 存在大量入度平均值 = 1 的节点, 导致这一现象的原因在于 Slashdot 数据集内有接近 80% 的边上的权重都为正值。由图 4 中节点声望值的分布可知, 因为消除了偏见值的不良影响, 从而有效提升了节点声望值

的平滑度. 这一实验结果充分表明了本方法在提高信任网络中不可信节点发现算法上的有效性. 在消除了不可信节点的信任网络中进行推荐, 必定可以提升推荐系统的精准度.



(a) 入度平均值
(a) In-degree mean



(b) 声望值
(b) Deserve

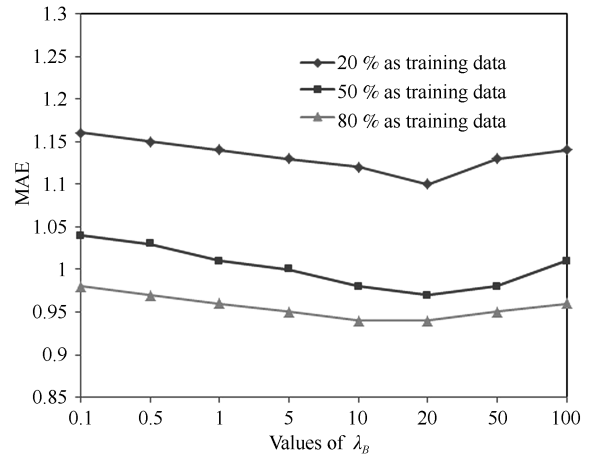
图 4 Slashdot 数据集的入度平均值与声望值

Fig. 4 In-degree mean and deserve for Slashdot datasets

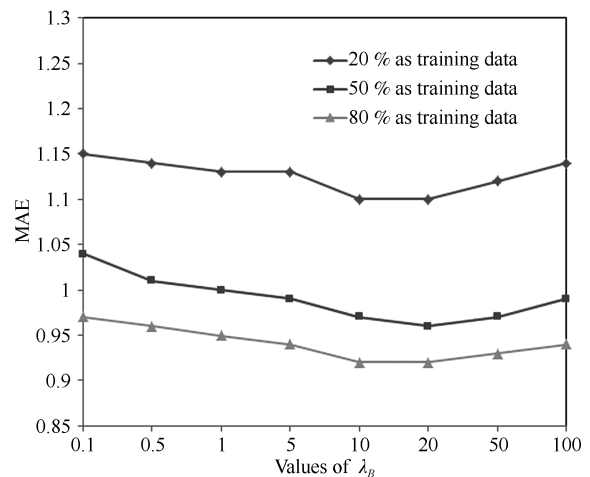
4.3 参数的影响实验

基于信任度的社会网络推荐算法的最大优势在于在推荐模型中融入了社会网络信息, 而这些社会网络信息将有助于优化用户兴趣模型的构建. 在 TTRM 推荐模型中, 参数 λ_B 是用于平衡来自用户-项目评分矩阵和用户信任关系的信息比例. 如果 λ_B 取值为 0, 表示仅依靠用户评分矩阵信息进行推荐; 反之, 当 λ_B 取无穷大时, 表示推荐模型在推荐时仅依赖于用户的信任关系矩阵进行用户喜好的预测. 显然, 只有联合用户自身的评分信息与信任关系信息才能使推荐效果达到最优. 下面通过实验验证参数 λ_B 的不同取值对 TTRM 推荐性能的产生影响. 实验中对用户特征矩阵的维度 L 分别取

5 和 10 两种取值, 测试集比重分别为 20%, 50%, 80% 三种比例, 使用 MAE 衡量 λ_B 对推荐效果的影响. 图 5 给出了在 Epinions 数据集上参数 λ_B 在用户特征矩阵维度分别为 5 和 10, 及其在不同比例测试集下 MAE 的结果.



(a) 用户特征维度 $L = 5$
(a) User feature dimension $L = 5$

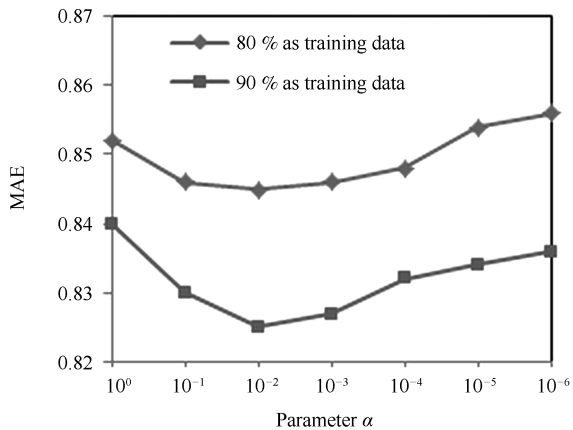


(b) 用户特征维度 $L = 10$
(b) User feature dimension $L = 10$

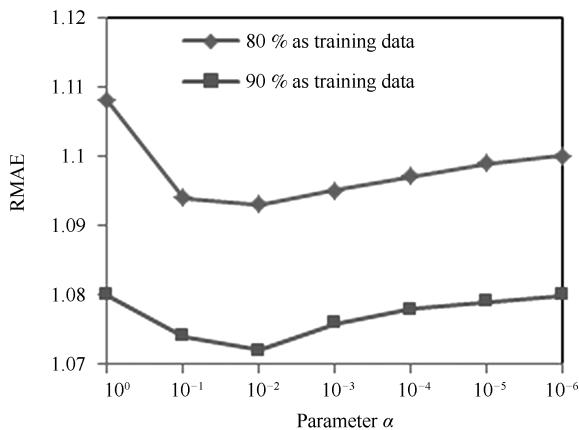
图 5 不同维度与不同训练集比例下参数 λ_B 的 MAE 结果
Fig. 5 Effect experiment of parameter λ_B under different training percent

由图 5 可知, 参数 λ_B 对算法的推荐性能有着重要的影响, 也验证了在推荐时综合考虑用户评分矩阵与信任关系可以大大提高推荐性能. 随着 λ_B 的增大, 推荐的准确度逐渐提高, 但当 λ_B 值继续增大时, 推荐性能反而降低, 由此可知, 参数 λ_B 的取值为 10 的时候, MAE 取值最低, 算法推荐性能最好. 并且对比用户特征矩阵维度分别为 5 和 10 下的 MAE 结果可知, 维度为 10 时的推荐效果更好, 这是由于用户特征矩阵越大则对应的用户特征分类越细, 从而算法的推荐性能越好. 因此我们最终将参数 λ_B 的取值设为 10, 用户特征矩阵的维度 L 设为 10.

在 Epinions 数据集上参数 α 的不同取值下 MAE 和 RMSE 的结果。



(a) MAE 结果
(a) MAE result



(b) RMAE 结果
(b) RMAE result

图6 参数 α 的影响实验 (维度 = 10)
Fig. 6 Impact experiment of parameter
(dimensionality = 10)

为了解决用户与朋友之间的兴趣差异会对用户特征向量产生负面影响的问题, TTRM 算法通过引入正则化项 (如式 (27) 所示) 来对用户与朋友的喜好差异进行约束. 其中参数 α 决定了算法优化过程中引入社会网络信息量的大小. 如果其值过小, 表明在推荐过程中仅考虑用户自身的喜好, 反之, 如果取值过大, 表明在推荐过程中只考虑其朋友的喜好. 显然, α 的取值不宜取极大或极小值, 因为 α 取极值均会降低推荐性能. 图 6 给出了在 Epinions 数据集上参数 α 的不同取值下 MAE 和 RMSE 的结果.

由图 6 可知, 随着参数 α 的增大, TTRM 算法性能首先是提高的, 这也验证了对用户与朋友喜好差异进行约束的有效性. 但是当参数 α 的取值小于 0.01 以后, MAE 和 RMSE 的取值反而增大, 说明推荐性能在降低. 因此我们最终将参数 α 的取值设为 0.01.

4.4 算法性能对比实验

为了验证 TTRM 算法的有效性, 本文利用衡量推荐算法性能所常用的几种算法进行仿真测试. 实验与以下几种算法进行对比, 以验证本文所提模型的正确性: UserMean: 该方法是利用每个用户对项目评分的均值进行用户未评分项的预测. Item-Mean: 该方法是利用每个项目评分的平均值对评分数据进行预测. PMF (Probabilistic matrix factorization)^[25]: 该研究分析工具来自于 Salakhutdinov 等的研究成果. PMF 具体以概率分布的矩阵因式分解为基础, 对用户-项目评分矩阵开展协同推荐计算. NMF (Non-negative matrix factorization)^[26]: 该方法是由 Lee 等提出的. NMF 是也是一种仅利用评分矩阵信息进行推荐的方法. Trust^[27]: 该协同推荐算法具体以信任网络为基础. 但与本文提出的推荐算法的最大区别在于, 该算法并未考虑信任网络中不可信节点对推荐结果的干扰与影响作用, 未能消除干扰因素的负面影响. STE^[16]: 该方法是由 Jamali 等提出的社会信任模型, 是一种典型的基于信任关系的推荐算法. SocialMF^[18]: 该算法具体以用户偏好传播为基础进行模型构建, 并完成偏好推荐. 这种由 Jamali 等提出的模型也为本文的研究基础, 但该算法没有考虑到不可信节点对信任建模的影响, 并且对用户与朋友的喜好差异也没有进行约束. 实验选取的测试集比重分别为 20%, 50%, 80%, 90%. 算法中用到的参数设置如下: $\lambda_B = 10$, 用户特征矩阵的维度 $L = 10$. $\lambda_U = \lambda_V = 0.1$, $\alpha = 0.01$. 数据稀疏性问题是影响推荐系统有效性的最大因素, 历史评分数据的缺乏, 则直接影响和限制了推荐系统的性能与效率. 因此, 为了确定本文所提出的 TTRM 算法的推荐性能, 需要进行充分的对比实验. 由于 TTRM 算法是一种基本信任关系的推荐, 因此首先将其与经典的信任推荐模型 Trust、STE 和 SocialMF 这三种算法相对比. 具体采取以下步骤完成检验: 以用户做出评分的数量为划分依据, 将训练集用户划分为 7 组: 这 7 组的评价数量分别为 “= 0”, “1 ~ 10”, “11 ~ 20”, “21 ~ 40”, “41 ~ 80”, “81 ~ 160” 以及 “> 160”, 取 TTRM 与其他推荐算法对不同评价数量用户的推荐效果以 URMSE 作为评测标准进行对比, 结果如表 3 所示. 由实验结果可知, 测试集在 90% 时, 当用户评价数量为 0 时, TTRM 在 URMSE 性能上与其他算法的差值最大, 说明 TTRM 对冷启动用户的推荐效果最好. 这是由于 TTRM 算法中在信任度计算时考虑了信任传递问题, 因此可以很好地实现冷启动用户的推荐问题. 当用户的评价数量继续增大时, 与其他算法在 URMSE 的取值差异慢慢减小, 但总体来看, TTRM 的推荐效果由于考虑了不可信节点的影响, 对其评价权重进行了弱化, 因此降低了不可信节点

给信任网络带来的负面影响. 同时, 该算法在推荐过程中引入了社会正则化项对用户与朋友的喜好差异进行了约束, 因此推荐质量较其他推荐算法都有所提升.

表 4 给出了在不同测试集比例下 TTRM 算法与其他各类推荐算法在 MAE、RMSE 与 URMSE

上的对比结果. 由表 4 可知本文提出的 TTRM 算法在平均绝对误差 (MAE)、均方根误差 (RMSE) 与用户均方根误差 (URMSE) 等评估指标结果的准确性均显著高于 Trust、NMF、PMF、SocialMF 等传统推荐算法, 尤其在用户评分数据缺乏的条件下, TTRM 算法的优势尤为突出.

表 3 不同用户评价数量下各推荐算法对比结果

Table 3 Comparative results of different algorithms under different user evaluation number

评价数量	0	1~5	6~10	11~20	21~40	4~80	81~160	161~320	321~640	> 640	
90 %	Trust	1.79	1.25	1.08	1.02	1.00	0.98	0.95	0.92	0.90	0.88
	STE	1.69	1.18	1.03	1.00	0.99	0.97	0.95	0.92	0.90	0.88
	SocialMF	1.32	1.12	1.01	0.98	0.93	0.91	0.90	0.88	0.86	0.85
	TTRM	1.07	1.02	0.95	0.90	0.88	0.86	0.85	0.84	0.83	0.82
80 %	Trust	1.85	1.28	1.09	1.04	1.02	0.99	0.97	0.95	0.92	0.90
	STE	1.68	1.15	1.05	1.03	1.00	0.96	0.95	0.92	0.91	0.89
	SocialMF	1.23	1.10	1.02	0.98	0.97	0.94	0.93	0.91	0.90	0.88
	TTRM	1.09	1.05	0.95	0.94	0.93	0.92	0.91	0.90	0.88	0.86
50 %	Trust	1.89	1.32	1.13	1.09	1.06	1.03	1.01	0.99	0.94	0.92
	STE	1.75	1.25	1.10	1.07	1.03	0.99	0.97	0.95	0.88	0.86
	SocialMF	1.35	1.15	1.08	1.02	1.00	0.98	0.95	0.93	0.85	0.85
	TTRM	1.12	1.10	0.98	0.97	0.94	0.93	0.92	0.91	0.89	0.85
20 %	Trust	1.98	1.30	1.13	1.12	1.11	1.11	1.09	1.04	1.04	1.03
	STE	1.78	1.23	1.10	1.08	1.07	1.06	1.05	1.02	1.02	1.01
	SocialMF	1.42	1.21	1.09	1.05	1.01	1.02	0.98	0.94	0.89	0.89
	TTRM	1.13	1.10	1.00	0.98	0.97	0.96	0.95	0.94	0.88	0.88

表 4 各推荐算法的性能对比结果

Table 4 Performance comparison results of different recommendation algorithm

	UserMean	ItemMean	PMF	NMF	Trust	SocialMF	TTRM	
90 %	MAE	0.913	0.877	0.865	0.871	0.832	0.802	0.789
	RMSE	1.169	1.238	1.154	1.162	1.101	1.051	1.021
	URMSE	1.740	1.652	1.156	1.142	1.132	0.937	0.902
80 %	MAE	0.929	0.891	0.889	0.895	0.854	0.813	0.801
	RMSE	1.182	1.259	1.177	1.183	1.126	1.053	1.028
	URMSE	1.802	1.756	1.215	1.192	1.168	0.988	0.875
50 %	MAE	0.932	0.955	0.923	0.921	0.912	0.875	0.832
	RMSE	1.192	1.263	1.185	1.193	1.236	1.089	1.092
	URMSE	1.820	1.752	1.237	1.214	1.192	1.017	1.002
20 %	MAE	0.946	0.957	0.932	0.929	0.918	0.904	0.885
	RMSE	1.205	1.206	1.185	1.123	1.056	1.094	1.085
	URMSE	1.840	1.756	1.255	1.210	1.177	1.036	1.011

5 结论

社会网络分析的诸多应用中, 对网络节点的可信度进行评估是其中的一项关键技术, 尤其是基于可信网络的应用. 从朋友推荐这一非常典型的社交网络推荐关系出发, 提出一种将关系信任度与用户-项目评分矩阵进行联合矩阵分解的社会化推荐方法 TTRM, 通过朋友的喜好对用户特征向量的构建进行修正, 实现了信任传递从而解决了新用户无信任值的冷启动问题. 推荐模型中通过加入社会化正则项对目标函数进行约束, 从而最小化了用户与朋友之间的喜好差异, 提高了基于信任度推荐的准确度. 实验结果表明 TTRM 方法由于考虑了不可信节点对信任网络的负面影响与信任传递等问题, 较其他推荐方法有大幅度的提高, 特别是对稀疏数据和伪信任网络效果非常明显.

References

- 1 Saravanan M, Buveneswari S, Divya S, Ramya V. Bayesian filters for mobile recommender systems. In: Proceedings of the 2011 International Conference on Advances in Social Networks Analysis and Mining. Kaohsiung City, Taiwan, China: IEEE, 2011. 715–721
- 2 Yeung K F, Yang Y Y, Ndzi D. A proactive personalised mobile recommendation system using analytic hierarchy process and Bayesian network. *Journal of Internet Services and Applications*, 2012, **3**(2): 195–214
- 3 Yin Gui-Sheng, Zhang Ya-Nan, Dong Yu-Xin, Han Qi-Long. A constrained trust recommendation using probabilistic matrix factorization. *Acta Electronica Sinica*, 2014, **42**(5): 904–911
(印桂生, 张亚楠, 董宇欣, 韩启龙. 基于受限信任关系和概率分解矩阵的推荐. 电子学报, 2014, **42**(5): 904–911)
- 4 Guo Lei, Ma Jun, Chen Zhu-Min. Trust strength aware social recommendation method. *Journal of Computer Research and Development*, 2013, **50**(9): 1805–1813
(郭磊, 马军, 陈竹敏. 一种信任关系强度敏感的社会化推荐算法. 计算机研究与发展, 2013, **50**(9): 1805–1813)
- 5 Granovetter M S. The strength of weak ties. *American Journal of Sociology*, 1973, **78**(6): 1360–1380
- 6 Gilbert E, Karahalios K. Predicting tie strength with social media. In: Proceedings of the 2009 SIGCHI Conference on Human Factors in Computing Systems. New York, NY, USA: ACM, 2009. 211–220
- 7 Xiang R J, Neville J, Rogati M. Modeling relationship strength in online social networks. In: Proceedings of the 19th International Conference on World Wide Web. New York, NY, USA: ACM, 2010. 981–990
- 8 Zhang Yan-Ping, Zhang Shun, Qian Fu-Lan, Zhang Yi-Wen. Robust collaborative recommendation algorithm based on user's reputation. *Acta Automatica Sinica*, 2015, **41**(5): 1004–1012
(张燕平, 张顺, 钱付兰, 张以文. 基于用户声誉的鲁棒协同推荐算法. 自动化学报, 2015, **41**(5): 1004–1012)
- 9 Golbeck J. Computing and Applying Trust in Web-based Social Networks [Ph.D. dissertation], University of Maryland, USA, 2005
- 10 Massa P, Avesani P. Trust metrics on controversial users: balancing between tyranny of the majority. *International Journal on Semantic Web and Information Systems*, 2007, **3**(1): 39–64
- 11 Jamali M, Ester M. *TrustWalker*: a random walk model for combining trust-based and item-based recommendation. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM, 2009. 397–406
- 12 Li B Y S, Yeung L F, Yang G K. Pathogen host interaction prediction via matrix factorization. In: Proceedings of the 2014 IEEE International Conference on Bioinformatics and Biomedicine. Belfast, United Kingdom: IEEE, 2014. 357–362
- 13 Menon A K, Elkan C. Link prediction via matrix factorization. *Machine Learning and Knowledge Discovery in Databases*. Berlin Heidelberg: Springer, 2011. 437–452
- 14 Ma H, Yang H X, Lyu M R, King I. SoRec: social recommendation using probabilistic matrix factorization. In: Proceedings of the 17th ACM Conference on Information and Knowledge Management. New York, NY, USA: ACM, 2008. 931–940
- 15 Tang J L, Hu X, Gao H J, Liu H. Exploiting local and global social context for recommendation. In: Proceedings of the 23rd International Joint Conference on Artificial Intelligence. New York, NY, USA: ACM 2013. 2712–2718
- 16 Ma H, King I, Lyu M R. Learning to recommend with social trust ensemble. In: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, NY, USA: ACM, 2009. 203–210
- 17 Tang J L, Gao H J, Liu H. mTrust: discerning multi-faceted trust in a connected world. In: Proceedings of the 5th ACM International Conference on Web Search and Data Mining. New York, NY, USA: ACM, 2012. 93–102
- 18 Jamali M, Ester M. A matrix factorization technique with trust propagation for recommendation in social networks. In: Proceedings of the 4th ACM Conference on Recommender Systems. New York, NY, USA: ACM, 2010. 135–142
- 19 Ma H, Zhou D Y, Liu C, Lyu M R, King I. Recommender systems with social regularization. In: Proceedings of the 4th ACM International Conference on Web Search and Data Mining. New York, NY, USA: ACM, 2011. 287–296
- 20 Xiao R L, Li Y N, Chen H T, Ni Y C, Du X. SRSP-PMF: a novel probabilistic matrix factorization recommendation algorithm using social reliable similarity propagation. *Intelligent Computing Theories and Methodologies: Lecture Notes in Computer Science*. Switzerland: Springer, 2015. 80–91
- 21 Ji K, Sun R Y, Li X, Shu W H. Improving matrix approximation for recommendation via a clustering-based reconstructive method. *Neurocomputing*, 2016, **173**: 912–920

- 22 Yu Z, Wang C, Bu J J, Wang X, Wu Y, Chen C. Friend recommendation with content spread enhancement in social networks. *Information Sciences*, 2015, **309**: 1102–1118
- 23 Kayacan E, Khanesar M A. Gradient descent methods for type-2 fuzzy neural networks. *Fuzzy Neural Networks for Real Time Control Applications: Concepts, Modeling and Algorithms for Fast Learning*. Amsterdam: Elsevier, 2016. 45–70
- 24 Kendall M G. A new measure of rank correlation. *Biometrika*, 1938, **30**(1–2): 81–93
- 25 Salakhutdinov R, Mnih A. Probabilistic matrix factorization. In: *Advances in Neural Information Processing Systems 20*. Cambridge, MA, USA: MIT Press, 2008. 1257–1264
- 26 Lee D D, Seung H S. Learning the parts of objects by non-negative matrix factorization. *Nature*, 1999, **401**(6755): 788–791
- 27 Yahyaoui H, Al-Mutairi A. A feature-based trust sequence classification algorithm. *Information Sciences*, 2016, **328**: 455–484



李 慧 博士, 淮海工学院计算机工程学院副教授. 主要研究方向为个性化推荐, 社会网络分析.
E-mail: shufanzs@126.com
(**LI Hui** Ph.D., associate professor in the Department of Computer Science, Huaihai Institute of Technology. Her research interest covers personality recommendation and socila network analysis.)



马小平 博士, 中国矿业大学信电学院教授. 主要研究方向为智能计算. 本文通信作者.
E-mail: xpma@cumt.edu.cn
(**MA Xiao-Pin** Ph.D., professor at the School of Information & Electrical Engineering, China University of Mining & Technology. His main research

interests is intelligent computing. Corresponding author of this paper.)



施 珺 淮海工学院计算机工程学院教授. 主要研究方向为智能信息处理.
E-mail: sj_lfg@hotmail.com
(**SHI Jun** Professor in the Department of Computer Science, Huaihai Institute of Technology. Her main research interest is information processing.)



李存华 博士, 淮海工学院计算机工程学院教授. 主要研究方向为数据挖掘.
E-mail: cli2000@126.com
(**LI Cun-Hua** Ph.D., professor is the Department of Computer Science, Huaihai Institute of Technology. His main research interest is data mining.)



仲兆满 博士, 淮海工学院计算机工程学院副教授. 主要研究方向为中文信息处理.
E-mail: zhongzhaoman@163.com
(**ZHONG Zhao-Man** Ph.D., associate professor in the Department of Computer Science, Huaihai Institute of Technology. His main research interest is Chinese information processing.)



蔡 虹 淮海工学院计算机工程学院讲师. 主要研究方向为智能信息处理.
E-mail: caihong@263.net
(**CAI Hong** Lecturer in the Department of Computer Science, Huaihai Institute of Technology. Her main research interest is information processing.)