

# 一种基于社会信任潜在因子模型的推荐方法

邢星<sup>1</sup> 张维石<sup>2</sup> 贾志淳<sup>1</sup>

(渤海大学信息科学与技术学院 锦州 121013)<sup>1</sup> (大连海事大学信息科学技术学院 大连 116026)<sup>2</sup>

**摘要** 随着社交网络的快速发展、社交网络用户规模的不断扩大,如何为用户推荐感兴趣的信息变得越发困难。传统的推荐方法利用用户兴趣的历史数据来预测用户未来感兴趣的项目,忽视了社交网络中的信任关系,导致推荐方法的推荐质量不高。针对上述问题,提出了基于社会信任潜在因子模型的推荐方法。该方法引入社会信任来度量社交网络中朋友之间的隐含信任关系,根据社会信任程度来选择用户信任的朋友,对用户信任的朋友与目标用户的共同兴趣进行潜在因子分析,构建基于社会信任的潜在因子模型,实现目标用户的前 $k$ 个项目推荐。真实数据集上的对比实验结果表明,基于社会信任潜在因子模型的推荐方法在推荐质量上优于现有的推荐方法。

**关键词** 社会信任计算,潜在因子分析,推荐系统,社会推荐,社交网络

**中图分类号** TP311 **文献标识码** A

## TBLFM: Trust Based Latent Factor Model for Social Recommendation

XING Xing<sup>1</sup> ZHANG Wei-shi<sup>2</sup> JIA Zhi-chun<sup>1</sup>

(School of Information Science and Technology, Bohai University, Jinzhou 121013, China)<sup>1</sup>

(School of Information Science and Technology, Dalian Maritime University, Dalian 116026, China)<sup>2</sup>

**Abstract** Recently online social networks have become the major platform with millions of registered users on the Web. The amount of information is increasing so quickly that users can't handle the information overload without the support of recommendation methods. Traditional recommendation methods have a limited performance in the context of social recommendation due to not considering the social network information, such as trust. Trust-based methods attempt to introduce a trust metric during the social recommendation. However, most of these methods are based on the explicit trust statements expressed by users, which are not available in the social networks such as Facebook, Twitter and Sina Weibo. This paper presented a trust metric to quantitatively measure the recommendation trust between pairs of users by aggregating the implicit trust and trust propagation values. We proposed a trust-based latent factor model, which incorporates the pairwise recommendation trust values into the probabilistic model for top-k item recommendation. The experiments on Sina Weibo demonstrate that our method outperforms the traditional recommendation methods and trust-based methods.

**Keywords** Social trust computation, Latent factor model, Recommender system, Social recommendation, Social network

## 1 引言

与传统的推荐方法<sup>[5,17]</sup>生成的推荐结果相比,人们更容易接受来自信任的人的推荐,而不是传统推荐方法生成的推荐<sup>[8,12]</sup>。在基于信任关系的推荐和传统推荐方法生成的推荐之间做出选择的时候,不管从推荐质量还是有效性上,用户更加倾向前者<sup>[9]</sup>。如何将信任机制引入到个性化推荐方法中,为用户提供基于信任关系的推荐,以便提高推荐方法的推荐质量是社交网络个性化推荐方法研究的热点问题和挑战<sup>[20]</sup>。

信任是一个非常复杂的概念,它关乎一个人的心理活动过程,与之相关的研究领域包括管理学、心理学、营销学、社会

学和计算机科学等诸多领域<sup>[1,4,15]</sup>。社会学家<sup>[18]</sup>将信任定义为两个部分:一部分是信念(Belief);另一部分是基于信念采取行动的意愿(Willingness)。社会学家对于信任的定义虽然不能直接应用在社交网络的实际个性化推荐场景中,但是为我们理解信任提供了一些启发。

近年来,随着社交网络的快速发展,社交网络中的信任(Social Trust)研究吸引了越来越多研究者的关注。Golbeck用信任关系来定量分析社交网络中的社会关系,提出了信任网络<sup>[6]</sup>,开发了个性化的电影评论网 FilmTrust,利用社交网络中的信任关系作为电影评分的权值,从而为用户提供更加准确可信的电影推荐<sup>[7,11]</sup>。

到稿日期:2013-03-27 返修日期:2013-06-14 本文受国家自然科学基金项目(61272172, 60973013),中央高校基本科研业务费项目(2011QN027)资助。

邢星(1982-),男,博士,副教授,主要研究方向为社交网络挖掘、社会计算、推荐系统等, E-mail: xing@cs.unm.edu; 张维石(1962-),男,博士,教授,博士生导师, CCF 会员,主要研究方向为软件工程、数据挖掘等; 贾志淳(1982-),女,博士,副教授,主要研究方向为 Web 服务、服务诊断等。

Massa 利用来自 Epinions 的用户之间事先给出的显示的信任数据建立一个信任预测模型来预测社会关系网络中传播的信任值,并利用这种信任值作为个性化推荐方法的推荐依据,取代基于用户兴趣相似度的个性化推荐<sup>[13]</sup>。

上述基于信任的个性化推荐方法只关注用户之间显式的信任关系,即用户之间的信任关系是通过用户事先给定的信任值来度量的,而很多有价值的隐式信任关系往往都被忽略了。更为重要的是,社交网络如 Facebook、Twitter、新浪微博等没有提供用户之间显示信任关系信息。因此,如何在没有提供显示信任信息的社交网络中,量化用户之间的信任关系是社交网络基于信任关系的个性化推荐方法研究亟待解决的问题。

针对上述问题,我们提出了基于社会信任潜在因子模型的推荐方法。该方法根据社交网络中之间的相似度来计算用户之间的直接信任,应用用户之间的直接信任值来度量用户之间信任关系的强弱;然后,定义两种社交网络中信任传递算子,计算社交网络用户之间的间接信任;定义推荐信任刻画社交网络用户之间的信任传播距离对用户之间的信任值的影响;最后,我们根据推荐信任在社交网络中用户匹配到目标用户信任的用户,将社交网络的显示朋友关系转换为信任网络的隐含信任关系。在此基础上,通过潜在因子分析目标用户和信任用户之间相同兴趣的成因,建立基于社会信任的潜在因子模型,实现对目标用户的个性化推荐。本文的主要贡献表现在以下几个方面:

(1)提出了一种新的社会信任计算方法,该方法结合社交网络中用户之间的兴趣相似度、信任传递路径和信任传播距离,计算社交网络用户之间隐含信任关系的信任程度。

(2)建立基于社会信任的潜在因子模型 TBLFM(Trust Based Latent Factor Model),将社交网络项目推荐问题转换为基于 TBLFM 的学习与推理问题,并实现了参数学习算法。

(3)在抽取到的真实数据集上对本文提出的推荐方法进行实验,实验结果显示基于 TBLFM 的推荐方法优于传统的推荐方法和基于信任的推荐方法。

## 2 社会信任计算

### 2.1 直接信任计算

社交网络相关研究发现具有相似个性的人有着相互联系的倾向,这就是社交网络的同质性(homophily)特点<sup>[2,14]</sup>。依据上述发现,我们使用社交网络用户朋友关系的相似度来度量朋友之间的信任程度。给定社交网络  $G$ 、用户  $u$  和  $u'$  之间的直接信任计算公式如下所示:

$$trust(u, u') = \frac{|F(u) \cap F(u')|}{|F(u) \cup F(u')|} \quad (1)$$

式中,  $F(u)$  表示用户  $u$  关注的朋友集合,  $F(u')$  表示用户  $u'$  关注的朋友集合,  $trust(u, u')$  表示目标用户  $u$  对用户  $u'$  的信任程度。

将式(1)归一化处理,得到社交网络朋友用户之间的直接信任计算公式:

$$trust(u, u') = \frac{trust(u, u')}{\sum_{u' \in F(u)} trust(u, u')} \quad (2)$$

根据式(2)计算朋友之间的直接信任,我们可以将无权的社交网络  $G$  转换为权重为直接信任值的信任网络  $G_t$ 。

### 2.2 间接信任计算

在实际应用中,社交网络中的每个用户结点可能只对一小部分其他用户结点得出这样的信任,而这一小部分用户结点同样会信任另一部分用户结点。如此传播下去,从社交网络  $G$  中的任意一个用户结点出发,这种信任关系都会沿着网络结构进行传播,在间接社会关系连接的用户结点之间建立信任关系,由此形成用户结点之间的信任网络  $G_t$ 。

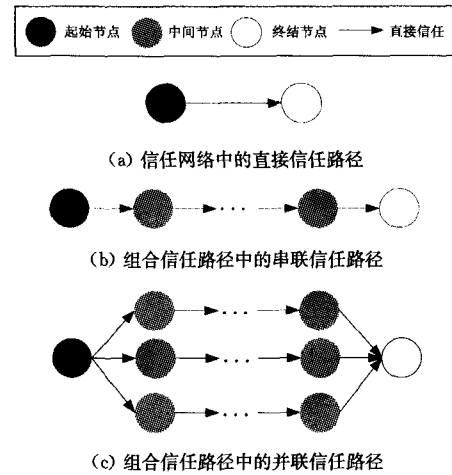


图1 信任网络  $G_t$  的组成元素

图1给出组成信任网络  $G_t$  中用户之间信任传递的主要元素,包括:

- 用户结点:社交网络中的用户。用户结点可以分为起始结点、中间结点和终结结点。
- 信任路径:连接结点之间的有向链路。依据信任路径中的结点数量分为直接信任路径(结点数量等于2)和组合信任路径(结点数量大于2)。直接信任路径是指不包含中间结点的信任路径,而组合信任路径是指由若干直接信任路径组成且包含多个中间结点的信任路径。
- 信任:表征用户之间信任程度的量化值,反映出用户之间的主观信任倾向。信任包括直接信任和间接信任。

通过信任网络  $G_t$ , 每个用户结点都可以对社交网络中未曾接触过的用户结点(非朋友关系结点)形成基本的认识并建立适当的信任关系。与此同时,用来度量用户结点之间的信任关系的信任值随着结点之间的距离增加而减少,信任的传递距离影响着用户之间信任关系的强弱。我们使用间接信任来计算在信任网络  $G_t$  中用户之间的信任传递,将组合信任路径分为串联信任路径和并联信任路径,通过定义下面两种信任传递算子计算社交网络中的间接信任。

连接算子:计算串联信任路径目标用户  $u$  和用户  $u'$  的间接信任,记作  $\otimes$ 。

$$\begin{aligned} trust^{p_1}(u, u') &= trust(u, u_1) \otimes trust(u_1, u_2) \otimes \dots \otimes trust \\ &\quad (u_{d-1}, u') \\ &= \min(trust(u, u_1), trust(u_1, u_2), \dots, trust \\ &\quad (u_{d-1}, u')) \end{aligned} \quad (3)$$

式中,  $p_1$  是连接目标用户  $u$  和用户  $u'$  的串联信任路径,目标用户  $u$  和用户  $u'$  的信任距离为  $d$ 。

并发算子:计算并联信任路径目标用户  $u$  和用户  $u'$  的间接信任,记作  $\oplus$ 。

$$trust(u, u') = trust^{p_1}(u, u') \oplus trust^{p_2}(u, u') \oplus \dots \oplus$$

$$\begin{aligned} & trust^{p_n}(u, u') \\ &= \max(trust^{p_1}(u, u'), trust^{p_2}(u, u'), \dots, \\ & \quad trust^{p_n}(u, u')) \end{aligned} \quad (4)$$

式中,  $p_1, p_2, \dots, p_n$  是  $n$  条连接目标用户  $u$  和用户  $u'$  的串联信任路径。

在计算用户之间的间接信任时, 首先通过连接算子计算用户在一个串联信任路径的间接信任, 然后通过并发算子计算用户在所有串联路径中的间接信任, 最后取所有间接信任的最大值作为用户之间的间接信任。

### 2.3 推荐信任计算

基于 TBLFM 推荐方法的另一个关键任务是决定用户之间信任传递的最大距离  $d_{max}$ 。一方面,  $d_{max}$  决定信任网络中与目标用户匹配的信任用户的最大搜索范围, 用户之间最大信任传递距离  $d_{max}$  越大, 在挖掘信任网络的过程中就越可以找到更多目标用户信任的用户; 另一方面, 在用户信任的传递过程中, 中间环节越多则传递效果会越差, 用户之间的信任会随着用户之间信任传递距离的增加而减少。

我们使用推荐信任将直接信任、间接信任和最大信任传递距离三者融合在一起, 为基于社会信任计算的社交网络个性化推荐提供依据。

$$recTrust(u, u') = trust(u, u') \times \frac{d_{max} - d_{u,u'} + 1}{d_{max}} \quad (5)$$

式中,  $d_{u,u'}$  是在信任网络  $G_t$  中目标用户  $u$  和用户  $u'$  的信任传递距离, 满足  $d_{u,u'} \leq d_{max}$ 。

### 2.4 基于社会信任的用户选择

社交网络中用户之间的信任决定了目标用户对社交网络中来自于其他用户的信息做出的相信或者不相信的判断。通常社交网络中的用户只会信任自己认为真正值得信赖的其他用户。由于信任是用户的主观信念和意愿, 用户对来自社交网络中其他用户的相信或者不相信的判断各不相同。

我们应用平均推荐信任来判断目标用户是否相信来自社交网络中其他用户的推荐。给定社交网络  $G$ 、目标用户  $u$  和最大信任传递距离  $d_{max}$ , 目标用户  $u$  的平均推荐信任计算如下:

$$\bar{t}_u = \frac{\sum recTrust(u, u')}{n} \quad (6)$$

通过目标用户  $u$  和用户  $u'$  之间的推荐信任与目标用户的平均推荐信任相比较, 得到社交网络目标用户  $u$  信任的用户集合, 计算公式如下:

$$TNeighbor(u) = \{u' | recTrust(u, u') \geq \bar{t}_u, d_{u,u'} \leq d_{max}\} \quad (7)$$

上述方法保证了在给定最大信任传递距离  $d_{max}$  的情况下, 高质量地选择目标用户在社交网络中所信任的用户集合, 同时, 也保证了基于社会信任的社交网络推荐方法的推荐质量。

为了方便计算, 将式(7)中用户之间的推荐信任进行归一化处理, 计算公式如下所示:

$$recTrust(u, u') = \frac{recTrust(u, u')}{\sum_{u' \in TNeighbor(u)} recTrust(u, u')} \quad (8)$$

## 3 基于社会信任的潜在因子模型

为了进一步研究基于社会信任的推荐方法, 我们对用户

之间的社会信任关系进行潜在因子分析<sup>[3,19]</sup>, 建立基于社会信任的潜在因子模型 TBLFM。

### 3.1 模型构建

首先, 我们定义社交网络项目推荐的目标函数  $\mathcal{F}$ , 用信任关系替代朋友关系进行个性化推荐。

$$\mathcal{F}: (\mathcal{I}^{rec}(u) | u, TNeighbor(u)) \rightarrow \mathbf{R} \quad (9)$$

式中,

•  $TNeighbor(u)$  表示目标用户  $u$  在社交网络  $G$  中的信任用户集合, 计算方法见式(7);

•  $\mathcal{A}(u) = \{i | (u, i) \in E^{click}, u \in \omega, i \in \mathcal{I}\}$ , 表示目标用户  $u$  所点击过的项目集合;

•  $\mathcal{I}^{rec}(u) = \sum_{u' \in TNeighbor(u)} \mathcal{A}(u') - \mathcal{A}(u)$ , 表示目标用户  $u$  推荐的项目集合。

为了进一步刻画目标函数  $\mathcal{F}$ , 定义下面的条件概率公式, 从概率的角度描述基于社交网络中的项目推荐问题。

$$\begin{aligned} p_u(\mathcal{A} | u, TNeighbor(u)) \\ = \prod_{u' \in TNeighbor(u)} \prod_{i \in \mathcal{I}} p_u(i | u, u')^{unlicked} \end{aligned} \quad (10)$$

式中,  $unlicked$  是指示变量, 满足:

$$unlicked = \begin{cases} 1, & i \in \mathcal{A}(u') - \mathcal{A}(u) \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

目标函数  $\mathcal{F}$  的求解过程表示为给定社交网络中目标用户和目标用户所信任的用户, 求解目标用户对信任用户所感兴趣项目的条件概率。

$$\begin{aligned} p_u(\mathcal{A} | u, TNeighbor(u)) \\ = \prod_{u' \in TNeighbor(u)} \prod_{i \in \mathcal{A}(u') - \mathcal{A}(u)} p_u(i | u, u') \\ = \prod_{u' \in TNeighbor(u)} \prod_{i \in \mathcal{A}(u') - \mathcal{A}(u)} p(i | t(u, u')) p_u(t | u, u') \end{aligned} \quad (12)$$

式中,  $p_u(t | u, u') = recTrust(u, u')$ , 根据式(8)计算。

引入潜在因子向量  $\mathcal{Z} = \{z_1, z_2, \dots, z_k\}$ , 对目标用户和信任用户的兴趣进行潜在因子分析。潜在因子向量  $\mathcal{Z}$  有  $k$  个潜在因子, 用来解释在观察到的数据元组  $\langle u, u', i \rangle$  中目标用户与信任用户共同兴趣的成因。

$$\begin{aligned} p_u(\mathcal{A} | u, TNeighbor(u)) \\ = \prod_{u' \in TNeighbor(u)} \prod_{i \in \mathcal{A}(u') - \mathcal{A}(u)} p(i | t(u, u')) p_u(t | u, u') \\ = \prod_{u' \in TNeighbor(u)} \prod_{i \in \mathcal{A}(u') - \mathcal{A}(u)} \sum_z p(i | z) p_u(z | t(u, u')) p_u(t | u, u') \end{aligned} \quad (13)$$

### 3.2 参数学习

为了学习模型中的潜在因子, 我们将式(13)表示为对数形式的似然函数:

$$\begin{aligned} \mathcal{L}(p_u; \theta) &\equiv \mathcal{L}(p_u(\mathcal{A} | u, TNeighbor(u); \theta)) \\ &= \sum_{u' \in TNeighbor(u)} \sum_{i \in \mathcal{A}(u') - \mathcal{A}(u)} [\sum_z [\log p(i | z) + \log p_u(z | t(u, u')) \\ & \quad + \log p_u(t | u, u')]] \\ &= \sum_{(u, u', i)} [\sum_z [\log p(i | z) + \log p_u(z | t(u, u'))]] + C \end{aligned} \quad (14)$$

式中,  $\theta = \{p(i | z), p_u(z | t(u, u'))\}$  是模型所要学习的参数,  $\langle u, u', i \rangle: u \times TNeighbor(u) \times \mathcal{I}^{rec}(u)$  是观察到的元组数据。

应用 EM 算法学习模型中的参数, 求解步骤如下所示:

E 步骤:

$$\mathcal{R}(z; t(u, u'), i, \theta^{old}) \equiv p_u(z | t(u, u'), i; \theta^{old})$$

$$= \frac{p(i | z)^{old} p_u(z | t(u, u'))^{old}}{\sum_{z'} p(i | z')^{old} p_u(z' | t(u, u'))^{old}} \quad (15)$$

M 步骤:

$$p(i|z)^{new} = \frac{\sum_{(u,u',i): i'=i} \mathcal{R}(z;t(u,u'),i,\theta^{old})}{\sum_{(u,u',i)} \mathcal{R}(z;t(u,u'),i,\theta^{old})} \quad (16)$$

$$p_u(z|t(u,u'))^{new} = \frac{\sum_{(u,u',i): u'=u} \mathcal{R}(z;t(u,u'),i,\theta^{old})}{|\{(u,u',i): u'=u\}|} \quad (17)$$

#### 算法 1 基于社会信任潜在因子模型参数学习算法

输入: 目标用户  $u$ , 信任集合  $TNeighbor(u)$ , 所有观察到的元组  $(u, u', i)$

输出: 参数  $\theta$

1. 初始化参数  $\theta$
2. repeat
3.  $\theta^{old} \leftarrow \theta$
4. 根据式(15)计算  $\mathcal{R}(z;t(u,u'),i,\theta^{old})$
5. 根据式(16)计算  $p(i|z)^{new}$
6. 根据式(17)计算  $p_u(z|t(u,u'))^{new}$
7.  $\theta^{new} \leftarrow p(i|z)^{new}, p_u(z|t(u,u'))^{new}$
8.  $\theta \leftarrow \theta^{new}$
9. until 式(14)满足收敛条件

### 3.3 基于社会信任潜在因子模型的推荐计算

根据算法 1 计算得到模型 TBLFM 的参数后,我们使用式(18)对目标用户  $u$  进行个性化推荐。

$$\begin{aligned} p(i|u) &= \sum_{u' \in TNeighbor(u)} p_u(i|u, u') \\ &= \sum_{u' \in TNeighbor(u)} p_u(t|u, u') p(i|t(u, u')), i \in \mathcal{P}^{rec}(u) \end{aligned} \quad (18)$$

将式(18)得到的计算结果降序排序,选择前  $k$  项最可能感兴趣的项目作为推荐结果返回给目标用户,从而实现基于社会信任潜在因子模型 TBLFM 的个性化推荐。

## 4 实验

### 4.1 数据来源

实验中的数据来源于新浪微博,应用新浪微博所提供的公开数据访问接口抽取新浪微博中的公开数据。我们随机选取 5 个用户,满足每个用户关注的用户数都大于等于 10,并且至少点击了 3 个项目。将这样的用户作为种子用户,应用广度优先的遍历方式,遍历种子用户所关注的用户,并将关注用户中至少点击 3 个项目的用户加入到用户集合中。根据六度分割原则,重复 6 次这样的搜索,并将收集到的用户信息、用户点击的项目信息和用户与用户之间关注关系信息全部存储到数据库中。

数据集包括 89 个推荐项目、3237 个用户、41894 个关注关系、20179 次项目点击。平均每个用户关注了 13 个用户,点击了 6 次项目。将抽取到的数据按照 80%~20% 的比例划分为训练集和测试集,我们随机进行 5 次这样的划分,应用交叉验证的方法检验推荐方法的推荐质量。

### 4.2 实验设置

我们采用下面 4 个常用的评价指标来评估不同推荐方法生成的前  $k$  个推荐项目的推荐质量<sup>[10,16]</sup>:

- ①前  $k$  个推荐项目的查准率 precision@ $k$ ;
- ②前  $k$  个推荐项目的查全率 recall@ $k$ ;
- ③前  $k$  个推荐项目的 F1 值 F1@ $k$ ;
- ④前  $k$  个推荐项目的平均准确率 AP@ $k$ 。

实验环境设置如下:32 核 CPU(Inter Xeon E5620, 2.40 GHz),32G 内存的 Dell 服务器,CentOS 6.2 操作系统,实验结果在 Matlab 中计算得到。

### 4.3 对比方法

为了评价基于社会信任潜在因子模型的推荐质量,我们将提出的方法与传统的个性化推荐方法和基于信任的个性化推荐方法相比较,对比方法具体描述如下:

- Baseline:基于项目(item-based)相似度计算的个性化推荐方法,使用 Jaccard 相似度计算项目之间的相似度<sup>[5]</sup>。该方法只利用用户项目点击矩阵中的数据,没有考虑社交网络中用户与用户之间的信任关系,是传统的个性化推荐方法。

- TidalTrust:由 Golbeck 提出的一种基于社交网络信任的个性化推荐方法<sup>[7]</sup>。该方法是建立在显示的用户信任网络基础上的,通过对信任网络的广度优先搜索,找到信任网络中与目标用户最近的用户作为信任评分的参考,使用加权平均信任网络中邻居节点的信任值的方法来计算用户之间的间接信任。TidalTrust 只考虑信任网络中邻居节点对目标用户节点的影响,所以会丢失很多信任路径中距离较大的信任用户。在实际推荐场景中,由于新浪微博没有提供显示的用户信任信息,因此将新浪微博中的关注关系看成是信任关系。换言之,在实现 TidalTrust 的时候,如果用户之间通过关注关系连接,那么将用户之间的直接信任值设置为 1,否则,设置为 0。使用文献[11]中的下述计算公式,计算 TidalTrust 中的信任传递:

$$t_{is} = \frac{\sum_{j \in adj(i)} t_{js}}{|adj(i)|} \quad (19)$$

式中, $t_{is}$ 表示用户  $i$  和用户  $s$  之间的信任传递, $adj(i)$ 表示信任网络中用户节点  $i$  的邻居节点。

- TidalTrust+:基于社交网络信任推荐 TidalTrust 的扩展方法。TidalTrust+与 TidalTrust 的不同之处在于 TidalTrust+应用 2.1 节中的直接信任计算将社交网络转换为信任网络,在显示信任网络上进行推荐;而 TidalTrust 将社交网络中的关注关系当作信任关系,在隐式的信任网络上进行推荐。我们应用式(1)和式(2)计算用户之间的直接信任,并将信任值返回给 TidalTrust+。

- TBLFM:基于社会信任潜在因子模型(Trust Based Latent Factor Model)的推荐方法。该方法中参数设置如下:最大信任传递距离  $d_{max}=3$ ,潜在因子个数设置为 10,算法 1 中的收敛条件设置为 0.001。

### 4.4 实验结果及分析

实验结果如表 1—表 4 所列,其中:表 1 显示参数  $k$  对前  $k$  个推荐项目平均准确率的影响,表 2 显示参数  $k$  对前  $k$  个推荐项目 F1 值的影响,表 3 显示参数  $k$  对前  $k$  个推荐项目查准率的影响,表 4 显示参数  $k$  对前  $k$  个推荐项查全率的影响。

表 1 参数  $k$  对前  $k$  个推荐项目的平均准确率的影响

k	Baseline	TidalTrust	TidalTrust+	TBLFM
5	0.0827	0.1141	0.1187	0.1541
6	0.1025	0.1258	0.1304	0.1813
7	0.1097	0.1442	0.1487	0.1899
8	0.1203	0.1507	0.1562	0.2090
9	0.1231	0.1625	0.1781	0.2221
10	0.1344	0.1782	0.1938	0.2355
11	0.1358	0.1800	0.1956	0.2392
12	0.1373	0.1919	0.1999	0.2481
13	0.1526	0.2060	0.2141	0.2637
14	0.1570	0.2149	0.2229	0.2686
15	0.1582	0.2169	0.2249	0.2721

表2 参数  $k$  对前  $k$  个推荐项目 F1 值的影响

k	Baseline	TidalTrust	TidalTrust+	TBLFM
5	0.1390	0.2006	0.2113	0.2440
6	0.1809	0.2142	0.2243	0.2799
7	0.1852	0.2326	0.2421	0.2770
8	0.2053	0.2380	0.2471	0.2959
9	0.2038	0.2585	0.2853	0.3032
10	0.2192	0.2849	0.3097	0.3227
11	0.2132	0.2750	0.2980	0.3161
12	0.2087	0.2892	0.2968	0.3199
13	0.2298	0.3059	0.3132	0.3349
14	0.2404	0.3137	0.3207	0.3336
15	0.2365	0.3065	0.3132	0.3321

表3 参数  $k$  对前  $k$  个推荐项目查准率的影响

k	Baseline	TidalTrust	TidalTrust+	TBLFM
5	0.2364	0.3091	0.3273	0.3818
6	0.2727	0.3182	0.3333	0.3939
7	0.2597	0.3117	0.3247	0.3636
8	0.2727	0.3068	0.3182	0.3636
9	0.2626	0.3131	0.3333	0.3535
10	0.2636	0.3273	0.3455	0.3545
11	0.2479	0.3058	0.3223	0.3388
12	0.2348	0.3030	0.3106	0.3258
13	0.2448	0.3077	0.3147	0.3287
14	0.2468	0.3117	0.3182	0.3247
15	0.2364	0.2970	0.3030	0.3152

表4 参数  $k$  对前  $k$  个推荐项目查全率的影响

k	Baseline	TidalTrust	TidalTrust+	TBLFM
5	0.1055	0.1783	0.1859	0.2099
6	0.1473	0.1983	0.2059	0.2650
7	0.1589	0.2336	0.2412	0.2775
8	0.1833	0.2480	0.2555	0.3161
9	0.1889	0.2929	0.3914	0.3520
10	0.2166	0.3341	0.4326	0.3927
11	0.2191	0.3365	0.4350	0.3995
12	0.2222	0.4397	0.4472	0.4936
13	0.2590	0.4765	0.4840	0.5304
14	0.2961	0.4955	0.5030	0.5403
15	0.3004	0.4998	0.5074	0.5522

均值 F1、前  $k$  个推荐项目查准率的平均值 Precision 和前  $k$  个推荐项目查全率的平均值 Recall。

表5 前  $k$  个推荐项目的推荐质量比较

推荐方法	AP	F1	Precision	Recall
Baseline	0.1285	0.2056	0.2526	0.2088
TidalTrust	0.1714	0.2654	0.3101	0.3394
TidalTrust+	0.1803	0.2783	0.3228	0.3717
TBLFM	0.2258	0.3054	0.3495	0.3936

图3显示了上述4个评价指标平均值的比较,数据参见表5。从图2和图3中可以看出,基于推荐信任的潜在因子模型 TBLFM 的推荐质量最高。

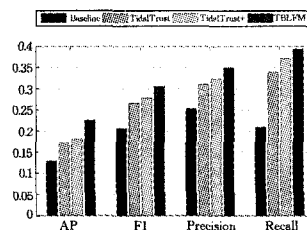


图3 不同评价指标平均值的比较

与基线方法 Baseline 相比较, TBLFM 的推荐质量在平均准确率 AP 上提高 43.09%, 在 F1 值上提高 32.68%; 与 TidalTrust 推荐方法相比较, TBLFM 的推荐质量在平均准确率 AP 上提高 24.92%, 在 F1 值上提高 13.98%; 与 TidalTrust+ 推荐方法相比较, TBLFM 的推荐质量在平均准确率 AP 上提高 20.15%, 在 F1 值上提高 8.87%。从推荐质量上看, 基于社会信任潜在因子模型 TBLFM 的推荐方法优于现有的基于项目的推荐方法和基于信任的推荐方法。一个原因是 TBLFM 准确地度量了社交网络用户之间的隐含信任关系, 并根据用户的主观信任程度对信任值偏低的用户进行个性化过滤, 进而选择来自社交网络目标用户的推荐信任程度高的用户群体。另一个原因是我们对选择到可信赖用户群体与目标用户的兴趣行为(用户-项目点击行为)进行潜在因子分析, 得到了目标用户与信任用户产生共同兴趣的潜在因素, 构建了基于推荐信任的潜在因子模型 TBLFM, 该模型根据已有的目标用户和信任用户群体的点击数据(共同兴趣)预测目标用户未来感兴趣的项目, 保证了对用户感兴趣项目预测的准确性, 从而进一步提高了推荐方法的推荐质量。

**结束语** 本文通过引入社会信任来度量社交网络用户和用户朋友之间的隐含信任关系, 根据用户之间的社会信任程度选择目标用户所信任的用户群体, 对选择出来的可信赖用户和目标用户的共同兴趣进行潜在因子分析, 构建基于社会信任的潜在因子模型 TBLFM, 实现基于 TBLFM 的推荐方法。将基于 TBLFM 的推荐方法应用在从新浪微博抽取到的真实数据中, 并与基于项目的推荐方法和基于信任的推荐方法相比较。实验结果显示, 本文提出的方法在前  $k$  个推荐结果中的平均准确率、F1 值、查全率和查准率优于所比较的推荐方法。

### 参考文献

[1] Andersen R, Borgs C, Chayes J, et al. Trust-based recommendation systems; an axiomatic approach[C]// Proceedings of the 17th international conference on World Wide Web, Beijing, China, 2008: 199-208

(下转第 191 页)

图2(a)-(d)分别显示了参数  $k=5, k=10$  和  $k=15$  对推荐算法的前  $k$  个推荐结果的平均准确率、F1 值、查准率和查全率的影响。

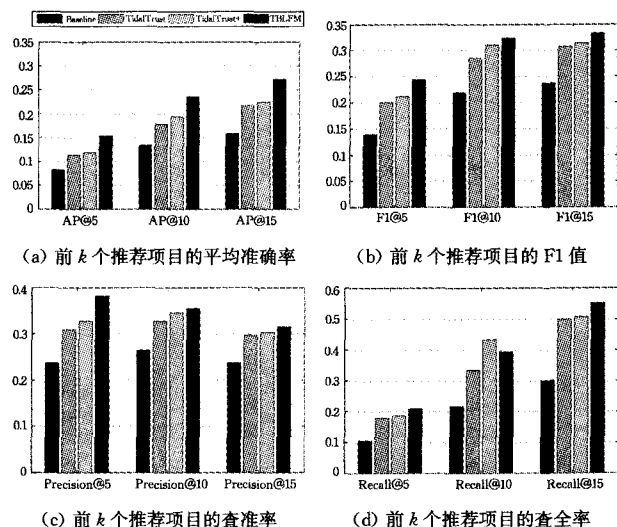


图2 参数  $k$  对推荐算法推荐质量的影响

表5显示了不同推荐方法前  $k$  个推荐项目的推荐质量, 分别计算不同推荐方法在  $k$  从 5 到 15 变化区间的前  $k$  个推荐项目平均准确率的平均值 AP、前  $k$  个推荐项目 F1 值的平

- [2] Davies S, Russel S. NP-completeness of searches for smallest possible feature sets [C]//Proceedings of the AAAI Fall 94 Symposiums on Relevance, Menlo Park, 1994; 37-39
- [3] Lakhina A, Crovella M, Diot C. Mining anomalies using traffic feature distributions [J]. ACM SIGCOMM Computer Communication Review, ACM, 2005, 35(4): 217-228
- [4] Berge A L, Pietra S D, Pietra V D. A maximum entropy approach to natural language processing[J]. Computational Linguistic, 1996, 22(1): 39-71
- [5] Li Yong-ming, Zhang Su-juan, Zeng Xiao-ping. Research of multi-population agent genetic algorithm for feature selection [J]. Expert Systems with Applications, 2009, 36(7): 11570-11581
- [6] Huang Cheng-long, Dun Jian-fan. A distributed PSO-SVM hybrid system with feature selection and parameter optimization [J]. Applied Soft Computing Journal, 2008, 8(4): 1381-1391
- [7] 范雪莉, 冯海泓, 原猛. 基于互信息的主成分分析特征选择算法 [J]. 控制与决策, 2012, 28(6): 915-919
- [8] Wang Yong, Li Lin, Ni Jun, et al. Feature selection using tabu search with long-term memories and probabilistic neural networks [J]. Pattern Recognition Letters, 2009, 30(7): 661-670
- [9] Marinake M, Marinakis Y, Doumpos M, et al. A comparison of several nearest neighbor classifier metrics using Tabu Search algorithm for the feature selection problem [J]. Optimization Letters, 2008, 2(3): 299-308
- [10] Tahir M A, Smith J. Improving nearest neighbor classifier using Tabu Search and ensemble distance metrics [C]//Proceedings of the Sixth International Conference on Data Mining (ICDM'06). Hong Kong, China; IEEE Computer Society, 2006; 1086-1090
- [11] Moore A W, Zuev D. Internet traffic classification using Bayesian analysis techniques [J]. ACM SIGMETRICS Performance Evaluation Review, ACM, 2005, 33(1): 50-60
- [12] 徐鹏, 林森. 基于 CA. 5 决策树的流量分类方法 [J]. 软件学报, 2009, 20(10): 2692-2704

(上接第 167 页)

- [2] Bisgin H, Agarwal N, Xu X. Investigating homophily in online social networks [C]//Proceedings of 2010 IEEE/WIC/ACM International Conference on Web Intelligence, WI, 2010; 533-536
- [3] Canny J. Collaborative filtering with privacy via factor analysis [C]//Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Tampere, Finland, 2002; 238-245
- [4] Caverlee J, Liu L, Webb S. Socialtrust: tamper-resilient trust establishment in online communities [C]//Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries. Pittsburgh PA, PA, USA, 2008; 104-114
- [5] Deshpande M, Karypis G. Item-based top-N recommendation algorithms [J]. ACM Transactions on Information Systems, 2004, 22(1): 143-177
- [6] Golbeck J. Computer science-Weaving a Web of trust [J]. Science, 2008, 321(5896): 1640-1641
- [7] Golbeck J. Computing And Applying Trust In Web-Based Social Networks [D]. University of Maryland, USA, 2005
- [8] Jamali M, Ester M. A matrix factorization technique with trust propagation for recommendation in social networks [C]//Proceedings of the Fourth ACM Conference on Recommender Systems. Barcelona, Spain, 2010; 135-142
- [9] 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. Paris, France, 2009; 397-406
- [10] Karypis G. Evaluation of Item-Based Top-N Recommendation Algorithms [C]//Proceedings of the tenth international conference on Information and knowledge management. Atlanta, Georgia, USA, 2001; 247-254
- [11] Kuter U, Golbeck J. SUNNY: a new algorithm for trust inference in social networks using probabilistic confidence models [C]//Proceedings of the 22nd national conference on Artificial intelligence. Columbia, Canada, 2007; 1377-1382
- [12] Ma H, 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. Boston, MA, USA, 2009; 203-210
- [13] Massa P, Avesani P. Trust-aware recommender systems [C]//Proceedings of the 2007 ACM conference on Recommender systems. Minneapolis, MN, USA, 2007; 17-24
- [14] McPherson M, Smith-Lovin L, Cook J M. Birds of a Feather; Homophily in Social Networks [J]. Annual Review of Sociology, 2001, 27(1): 415-444
- [15] Moghaddam S, Jamali M, Ester M, et al. FeedbackTrust: using feedback effects in trust-based recommendation systems [C]//Proceedings of the third ACM conference on Recommender systems. New York, New York, USA, 2009; 269-272
- [16] Niu S, Guo J, Lan Y, et al. Top-k learning to rank; labeling, ranking and evaluation [C]//Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval. Portland, Oregon, USA, 2012; 751-760
- [17] Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms [C]//Proceedings of the 10th International Conference on World Wide Web. Hong Kong, China, 2001; 285-295
- [18] Sztompka P. Trust: A Sociological Theory [M]. Cambridge: Cambridge Univ. Press, 1999
- [19] Xing X, Zhang W, Jia Z, et al. Learning to recommend top-k items in online social networks [C]//Proceedings of the 2012 World Congress on Information and Communication Technologies. WICT 2012, India, 2012; 1171-1176
- [20] 周涛. 个性化推荐的十大挑战 [J]. 计算机协会通讯, 2012, 8(7): 48-61