

# 融合社会标签与信任关系的社会网络推荐方法

胡云<sup>1</sup> 张舒<sup>2</sup> 李慧<sup>3,4</sup> 施珺<sup>3</sup> 仲兆满<sup>3</sup>

(1. 南京中医药大学信息技术学院, 南京, 210023; 2. 淮海工学院商学院, 连云港, 222005; 3. 淮海工学院计算机工程学院, 连云港, 222005; 4. 江苏省海洋资源开发研究院, 连云港, 222005)

**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, 本文将标签与基于信任的社交推荐方法相结合, 提出了一种融合社会标签和信任关系的社会网络推荐方法。该方法利用概率因式分解技术实现了社会信任关系、项目标记信息和用户项目评分矩阵的集成。从不同维度出发, 实现了用户和项目潜在特性空间的互连。在此基础上, 通过概率矩阵因式分解技术实现降维, 从而实现了有效的社会化推荐。在 Epinions 和 Movielens 数据集上的实验结果表明本文所提出的方法优于传统的社会化推荐和社会标签推荐算法, 特别是当用户评分数据较少时该算法的优越性体现得更好。

**关键词:** 社会网络; 推荐; 信任度; 矩阵分解; 标签

**中图分类号:** TP391      **文献标志码:** A

## Social Networks Recommendation Based on Social Tag and Trust Relation

Hu Yun<sup>1</sup>, Zhang Shu<sup>2</sup>, Li Hui<sup>3,4</sup>, Shi Jun<sup>3</sup>, Zhong Zhaoman<sup>3</sup>

(1. College of Information Technology, Nanjing University of Chinese Medicine, Nanjing, 210023, China; 2. School of Business, Huaihai Institute of Technology, Lianyungang, 222005, China; 3. Department of Computer Science, Huaihai Institute of Technology, Lianyungang, 222005, China; 4. Marine Resource Development Institute of Jiangsu, Lianyungang, 222005, China)

**Abstract:** Aiming at the universal problem of data sparsity and cold start in recommendation system, a social network recommendation method based on the combination of social label and trust relation is proposed in the paper. The proposed method uses the probabilistic factorization technique to integrate the social trust, item marking information and user item score matrix. The interconnection of users and potential item feature space is realized from different dimensions. On this basis, the realization of dimension reduction by the factorization of probability matrix can achieve effective social recommendation. Experimental results on Epinions and Movielens data sets show that the proposed method is superior to traditional social recommendation and social label recommendation algorithms, especially in the case of less user score data.

**Key words:** social network; recommendation; trust; matrix factorization; tag

**基金项目:** 国家自然科学基金(61403156, 61403155)资助项目; 江苏高校“青蓝工程”资助项目; 连云港市“521工程”科研资助项目; 连云港市科技计划(JC1608)资助项目; 淮海工学院自然科学基金(Z2017012, Z2015012)资助项目; 淮海工学院教学改革项目(XJG2017-2-5)资助项目; 教育部协同育人项目(201701028110, 201701028011, 201702134005)资助项目。

**收稿日期:** 2016-09-25; **修订日期:** 2016-11-29

## 引言

伴随着社交网络的盛行,在线用户生成的信息呈指数型增长,社会语义信息的分析对于许多 Web 应用程序变得越来越重要,然而用户冷启动问题导致的用户-项目矩阵数据稀疏性等问题严重影响了推荐质量。现有的推荐方法均未能很好地处理具有极少评分项的用户推荐。目前推荐系统与各种社会背景信息相关联,包括用户的社会信任网络、用户标签<sup>[1-3]</sup>或与项目相关的标签信息等<sup>[4-5]</sup>。

社交标签可以通过从新的方向反映用户对网络资源的偏好来描述用户选择的类型,即社会化标签反映了用户感兴趣的一些主题,可以利用一些主题偏好比如标签的使用来推断用户的兴趣方向。目前大量关于社会标签系统的研究调查<sup>[6-10]</sup>已经证实了标签可以很精准地代表用户对网页内容的判断及喜好,也可以很好地对资源进行描述。Heymann 等<sup>[6]</sup>对社会化标签预测进行了研究,发现基于标签的关联规则可以产生非常精确的标签预测;Ramage 等<sup>[11]</sup>采用标签作为一种辅助数据源应用于页面文本和锚文本,用来提高网页自动聚类;Wu 等<sup>[12]</sup>在带有权重的标签社会网中应用了基于扩散的推荐方法来提高推荐系统的精度,然而对是否可以利用社会标签信息来帮助提高推荐质量的研究甚少。因与以往研究工作不同,本文的研究将社会标签信息与信任度相融合来提高社会化推荐的质量,从而有效地解决了推荐系统中的冷启动问题。

信任网络是指由社会网络内部所有存在信任关系的用户所共同构成的一种网络结构。现在,进行社会网络推荐的算法有很多,它们主要以信任关系为基础,但大多数还是偏向于如何进行准确的信任度计算和信任模型的推理。如对用户进行推荐预测时可以使用多种相似度度量方法的线性组合<sup>[13]</sup>;基于用户信任关系矩阵的概率分解和传统评分矩阵的概率分解提出了一种新的推荐方法,该方法基于受限关系和概率分解矩阵<sup>[14]</sup>;当对用户的兴趣进行建模时,进一步通过辨别与目标用户有相同爱好的朋友来解决恢复信任度的过程<sup>[15]</sup>;通过合并多个目标函数,在矩阵分解的框架下解决问题<sup>[16]</sup>。

为了有效解决社会网络推荐中的冷启动问题,让社会网络推荐变得更加高效,本文提出了一种基于社会标签与信任关系的社交网络推荐方法(Tag-based and trust-based recommendation, TTR),该方法结合社交标签和信任度的社会化推荐方法,在基于概率因式分析的基础上集成了社会信任关系、项目标签信息以及用户项目评分矩阵。将这些维度不同的数据资源通过公共的用户潜在特征空间(或项目潜在特征空间)进行相连接,利用概率矩阵因式分解技术实现降维,从而在获得低维度的用户及项目潜在特征空间的基础上实现高效的社会化推荐。在 Epinions 和 Movielens 数据集上的实验结果表明,与现有的基于信任度的社会化推荐及社会标签推荐算法相比,TTR 算法的推荐性能更加优越,特别是当用户的评分数据量非常少时,其优势就变得很明显。另外,通过算法的复杂度分析,显示该算法适用于非常大的数据集,因为算法的复杂性和观测数据的数量之间存在线性关系。

## 1 信任矩阵分解模型

信任矩阵分解的基本思想就是对社会网络的信任关系进行分析,从而获得两个分别表示用户和信任关系的低维潜在特征矩阵。通过以上处理,能够完成评分矩阵  $\mathbf{R}$  向用户特征矩阵  $\mathbf{U}$  和项目特征矩阵  $\mathbf{V}$  的转换,这些特征是描述用户和推荐对象的重要因素。为了进一步发掘出潜藏在信任关系之后的用户特征,通过信任关系来提升推荐的准确度,使用信任关系矩阵对目标函数进行限制,将社会关系矩阵和用户评分矩阵在共享用户特征空间上进行联合概率分解,挖掘出评分相近又具有相同兴趣嗜好的信任用户成为邻居用户产生推荐,与之对应的概率图模型如图 1 所示。用户信任关系矩阵  $\mathbf{B}$  可用用户特征矩阵  $\mathbf{U}$  和信任特征矩阵  $\mathbf{Q}$  的内积表示;用户-项目评分矩阵  $\mathbf{R}$  可用用户特征矩阵  $\mathbf{U}$  和项目特征矩阵  $\mathbf{V}$  的内积表示。

假设在  $\mathbf{R}$  中包含  $m$  个用户和  $n$  个项目, $R_{ij}$  则表示用户  $u_i$  对项目  $v_j$  的评分,分解得到的用户特征矩

阵用  $\mathbf{U} \in \mathbf{R}^{l \times m}$  表示, 用  $\mathbf{Q} \in \mathbf{R}^{l \times m}$  表示分解得到的信任关系特征矩阵, 用  $l$  表示用户特征个数, 列向量  $\mathbf{U}_i, \mathbf{Q}_k$  分别表示用户  $u_i$  和信任矩阵  $\mathbf{B}$  所对应的潜在特征向量。为了使预测评分  $\mathbf{U}^T \mathbf{Q}$  与用户信任关系矩阵  $\mathbf{B}$  的误差最小, 定义信任关系矩阵  $\mathbf{B}$  的条件概率分布为

$$p(\mathbf{B} | \mathbf{U}, \mathbf{Q}, \sigma_B^2) = \prod_{i=1}^m \prod_{k=1}^m [N(b_{ij} | g(\mathbf{U}_i^T \mathbf{Q}_k), \sigma_B^2)]^{I_{ij}} \quad (1)$$

式中:  $N(x | \mu, \sigma^2)$  表示随机变量  $x$  服从数学期望为  $\mu$ 、标准方差为  $\sigma^2$  的高斯分布;  $I_{ij}$  表示指示函数, 其取值为 0 表示用户  $u_i$  信任  $u_k$ , 取值为 1 则表示用户  $u_i$  不信任  $u_k$ ; 函数  $g(x) = 1/(1 + \exp(-x))$ , 其作用是限定信任预测值在  $[0, 1]$  之间。为了防止这种情况的发生, 假设用户和可靠性矩阵仍然是平均为 0 的球形高斯先验, 即

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

$$p(\mathbf{Q} | \sigma_Q^2) = \prod_{k=1}^m N(\mathbf{Q}_k | 0, \sigma_Q^2 \mathbf{I}) \quad (3)$$

通过贝叶斯推荐可得

$$p(\mathbf{U}, \mathbf{Q} | \mathbf{B}, \sigma_B^2, \sigma_U^2, \sigma_Q^2) \propto p(\mathbf{B} | \mathbf{U}, \mathbf{Q}, \sigma_B^2) p(\mathbf{U} | \sigma_U^2) p(\mathbf{Q} | \sigma_Q^2) = \prod_{i=1}^m \prod_{k=1}^m [N(b_{ij} | g(\mathbf{U}_i^T \mathbf{Q}_k), \sigma_B^2)]^{I_{ij}} \times \prod_{i=1}^m N(\mathbf{U}_i | 0, \sigma_U^2 \mathbf{I}) \times \prod_{k=1}^m N(\mathbf{Q}_k | 0, \sigma_Q^2 \mathbf{I}) \quad (4)$$

将信任矩阵  $\mathbf{B}$  代入式(4), 可得

$$p(\mathbf{U}, \mathbf{V}, \mathbf{Q} | \mathbf{B}, \mathbf{R}, \sigma_R^2, \sigma_B^2, \sigma_V^2, \sigma_U^2, \sigma_Q^2) \propto p(\mathbf{R} | \mathbf{U}, \mathbf{V}, \sigma_R^2) p(\mathbf{B} | \mathbf{U}, \mathbf{Q}, \sigma_B^2) p(\mathbf{U} | \sigma_U^2) p(\mathbf{V} | \sigma_V^2) p(\mathbf{Q} | \sigma_Q^2) \quad (5)$$

式(5)利用用户自己评论的信息与信任矩阵对用户和信任的潜在特征进行估算, 具体的求解过程与标签模型相似, 在此就不再展开。

## 2 融合标签与信任关系的推荐模型

基于如何利用用户的社会信任关系进行推荐的总体框架, 可以很容易地扩展到融合用户项目评分矩阵与社会标签推荐模型上来。与用户评分相似, 具有相似标签的项目信息同样也可以作为推荐的重要来源, 因此本节利用相同的因式分解技术对用户项目评分与项目标签进行分解。通过共享的项目特征将用户的矩阵与项目标签信息进行联合, 得到如图 2 所示的基于标签推荐的概率图模型。其中  $\mathbf{T}$  表示每个标签的潜在特征,  $F_{jk}$  表示标签, 其值为项目  $v_j$  被标签  $t_k$  标记的次数。用户潜在特征空间从评分与标签角度反映了用户的喜好, 而项目潜在特征空间也可以由项目所收到的评分与标签信息来体现。

与信任矩阵分解模型相似, 可以得到

$$p(\mathbf{U}, \mathbf{V}, \mathbf{T} | \mathbf{F}, \mathbf{R}, \sigma_R^2, \sigma_F^2, \sigma_V^2, \sigma_U^2, \sigma_T^2) \propto p(\mathbf{R} | \mathbf{U}, \mathbf{V}, \sigma_R^2) p(\mathbf{F} | \mathbf{T}, \mathbf{V}, \sigma_F^2) p(\mathbf{U} | \sigma_U^2) p(\mathbf{V} | \sigma_V^2) p(\mathbf{T} | \sigma_T^2) \quad (6)$$

联合用户评分矩阵与项目标签矩阵分解的后验概率值满足

$$\ln p(\mathbf{U}, \mathbf{V}, \mathbf{T} | \mathbf{R}, \mathbf{F}, \sigma_R^2, \sigma_F^2, \sigma_V^2, \sigma_U^2, \sigma_T^2) = -\frac{1}{2\sigma_R^2} \sum_{i=1}^m \sum_{j=1}^n I_{ij}^R (r_{ij} - g(\mathbf{U}_i^T \mathbf{V}_j))^2 -$$

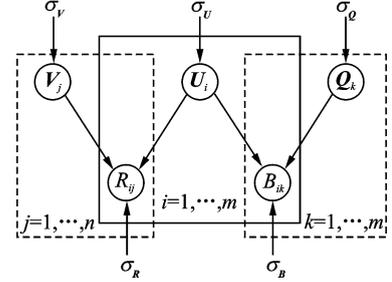


图 1 基于信任推荐的概率图模型  
Fig. 1 Probabilistic graphical model based on trust recommendation

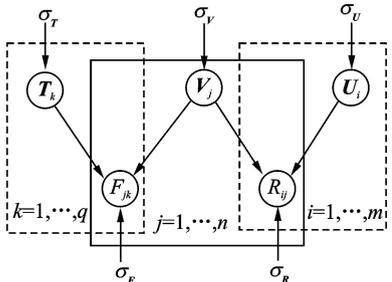


图 2 基于标签推荐的概率图模型  
Fig. 2 Probabilistic graphical model based on tag recommendation

$$\begin{aligned} & \frac{1}{2\sigma_F^2} \sum_{j=1}^n \sum_{k=1}^q I_{ij}^F (f_{jk} - g(\mathbf{V}_j^T \mathbf{T}_k))^2 - \\ & \frac{1}{2\sigma_U^2} \sum_{i=1}^m \mathbf{U}_i^T \mathbf{U}_i - \frac{1}{2\sigma_V^2} \sum_{j=1}^n \mathbf{V}_j^T \mathbf{V}_j - \frac{1}{2\sigma_T^2} \sum_{k=1}^q \mathbf{T}_k^T \mathbf{T}_k - \frac{1}{2} \left( \left( \sum_{i=1}^m \sum_{j=1}^n I_{ij}^R \right) \ln \sigma_R^2 + \left( \sum_{j=1}^n \sum_{k=1}^q I_{jk}^F \right) \ln \sigma_F^2 \right) - \\ & \frac{1}{2} (ml \ln \sigma_U^2 + nl \ln \sigma_V^2 + ql \ln \sigma_T^2) + \epsilon \end{aligned} \quad (7)$$

式中  $\epsilon$  为与参数无关的常数。当式(7)取最大值时,可知项目评分矩阵与标签矩阵最小偏差值可分解相对应的用户特征矩阵、项目特征矩阵和标签特征矩阵。将式(7)简化成为式(8),可将解决最大值的问题转换为解决相对应的式(8)的最小值问题。

$$\begin{aligned} L(\mathbf{R}, \mathbf{F}, \mathbf{U}, \mathbf{V}, \mathbf{T}) = & \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n I_{ij}^R (r_{ij} - g(\mathbf{U}_i^T \mathbf{V}_j))^2 + \frac{\lambda_F}{2} \sum_{j=1}^n \sum_{k=1}^q I_{jk}^F (f_{jk} - g(\mathbf{V}_j^T \mathbf{T}_k))^2 + \\ & \frac{\lambda_U}{2} \|\mathbf{U}\|_F^2 + \frac{\lambda_V}{2} \|\mathbf{V}\|_F^2 + \frac{\lambda_T}{2} \|\mathbf{T}\|_F^2 \end{aligned} \quad (8)$$

式中:  $\lambda_U = \sigma_R^2 / \sigma_U^2$ ,  $\lambda_V = \sigma_R^2 / \sigma_V^2$ ,  $\lambda_F = \sigma_R^2 / \sigma_F^2$ ,  $\lambda_T = \sigma_R^2 / \sigma_T^2$ ,  $\|\cdot\|_F^2$  表示 Frobenius 范数。若求得如式(8)所示的目标函数  $L$  的部分最小值,可在  $\mathbf{U}_i$ ,  $\mathbf{V}_j$  和  $\mathbf{T}_k$  上通过对所有用户、所有项目和标签来使用梯度下降法来获得,即

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{U}_i} = & \sum_{j=1}^n I_{ij}^R g'(\mathbf{U}_i^T \mathbf{V}_j) (g(\mathbf{U}_i^T \mathbf{V}_j) - r_{ij}) \mathbf{V}_j + \lambda_U \mathbf{U}_i \\ \frac{\partial L}{\partial \mathbf{V}_j} = & \sum_{i=1}^m I_{ij}^R g'(\mathbf{U}_i^T \mathbf{V}_j) (g(\mathbf{U}_i^T \mathbf{V}_j) - r_{ij}) \mathbf{U}_i + \lambda_V \mathbf{V}_j + \lambda_F \sum_{k=1}^q I_{jk}^F g'(\mathbf{V}_j^T \mathbf{T}_k) (g(\mathbf{V}_j^T \mathbf{T}_k) - f_{jk}) \mathbf{T}_k \\ \frac{\partial L}{\partial \mathbf{T}_k} = & \lambda_T \sum_{j=1}^n I_{jk}^F g'(\mathbf{V}_j^T \mathbf{T}_k) (g(\mathbf{V}_j^T \mathbf{T}_k) - f_{jk}) \mathbf{V}_j + \lambda_T \mathbf{T}_k \end{aligned} \quad (9)$$

式中:  $g'(x)$  为函数  $g(x)$  的导数,即  $g'(x) = \exp(x) / (1 + \exp(x))^2$ ; 为了减少算法的复杂度,令  $\lambda_U = \lambda_V = \lambda_T$ ; 参数  $\lambda_T$  用于控制将多少项目的标签信息融入到推荐模型中。算法对于  $\mathbf{U}$ ,  $\mathbf{V}$  和  $\mathbf{T}$  的学习方法如下: 首先随机初始化  $\mathbf{U}$ ,  $\mathbf{V}$  和  $\mathbf{T}$  的取值,然后根据梯度值迭代更新矩阵  $\mathbf{U}$ ,  $\mathbf{V}$  和  $\mathbf{T}$ ,直至目标函数达到收敛。

算法主要的计算代价在于目标函数  $L$  和在矩阵  $\mathbf{U}$ ,  $\mathbf{V}$  和  $\mathbf{T}$  的梯度学习上。由于用户-项目评分矩阵  $\mathbf{R}$  和标签矩阵  $\mathbf{F}$  都是稀疏矩阵,假设  $\mathbf{R}$  和  $\mathbf{F}$  中的非零元素数目分别为  $\rho_R$  和  $\rho_F$ ,则目标函数  $L$  的计算复杂度为  $O(\rho_R l + \rho_F l)$ ; 式(9)中梯度  $\partial L / \partial \mathbf{U}_i$ ,  $\partial L / \partial \mathbf{V}_j$  和  $\partial L / \partial \mathbf{T}_k$  的计算复杂度分别为  $O(\rho_R l + \rho_F l)$ ,  $O(\rho_R l)$  和  $O(\rho_F l)$ 。因此算法运行一次总的计算复杂度为  $O(\rho_R l + \rho_F l)$ ,这也说明了算法的复杂度与观测数据成线性关系。

### 3 实验与结果分析

#### 3.1 数据集与评价指标

本文研究的社会信任关系类型是有向的,所以选择两个公开的产品评论网站中的数据作为此次实验的数据集: Epinions 和 MovieLens 数据集。在该网站上的用户不仅可以对不同的产品进行评分,还可以对产品进行评论。另外,用户还可以对其他用户对该产品写下的评论进行评价,并将可信用户添加到“信任列表”里。MovieLens 数据集拥有众多的社会标签信息,包含了 10 000 054 条评分数据和 95 580 条标记数据,是由 71 567 名用户对 10 681 部电影标注产生的。本节中使用其中的 10M/100M 数据集进行测试。两个数据集的统计数据如表 1 所示。

MovieLens 10M/100M 数据相聚随机选择 80% 作为训练集,20% 作为测试集。每次实验验证都是独立的运行几次,最终将其结果取平均值。实验采用平均绝对误差 (Mean absolute error, MAE) 和均方根误差 (Root mean square error, RMSE) 作为系统评估指标,对各个推荐系统的性能进行对比。

表 1 Epinions 和 MovieLens 数据集的统计数据

Tab. 1 Statistical data of Epinions and MovieLens data sets

数据集	Epinions	MovieLens
用户数	22 166	7 375
项目数	296 277	106 797
评分数量	922 267	284 086
社会关系	300 548	111 781
每个用户平均评分数量	41.61	38.52
每个项目平均评分数量	3.11	2.66
信任网络密度	0.001 2	0.004 1

MAE 为全部单个观测值和算术平均值的偏差的绝对值,其大小与推荐性能成反比,即其值越小,则说明其推荐性能越好,定义为

$$\text{MAE} = \frac{\sum_{i,j} |r_{i,j} - \bar{r}_{i,j}|}{N} \quad (10)$$

式中:  $r_{i,j}$  表示用户  $i$  为项目  $j$  所打的评分;  $\bar{r}_{i,j}$  表示使用推荐算法计算得到的预测评分;  $N$  表示要观测评分的个数。

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,其大小也与推荐性能成反比,即其值越小,其推荐性能越好,定义为

$$\text{RMSE} = \sqrt{\frac{\sum_{i,j} (r_{i,j} - \bar{r}_{i,j})^2}{N}} \quad (11)$$

### 3.2 参数的影响实验

在 TTR 推荐模型中,融入用户-项目评分矩阵和标签信息之间的比例大小由参数  $\lambda_r$  来控制。如果  $\lambda_r$  取值为 0,表示它的推荐只对用户评分矩阵信息依靠来进行;如果  $\lambda_r$  取无穷大时,表示推荐模型只对项目所被标注的标签信息依赖进行用户喜好的预测。只有联合订阅者自身的评价信息和项目标签信息,推荐效果才最有利。

实验环境的参数为:用户特征矩阵的维度  $L=10$ ,  $\lambda_u = \lambda_v = 0.1$ 。训练集比例分别选取 10%, 30%, 50% 和 80%,图 3 给出了当用户特征矩阵维度  $L=10$  时对 MAE 与 RMSE 在不同比例训练集下的实验结果。由图 3 可知,参数  $\lambda_r$  对算法的推荐性能会产生重要影响,也验证了用户评分矩阵与社会标签的融合将有助于提升推荐性能。随着  $\lambda_r$  的增大,推荐精准度逐渐提高,但当  $\lambda_r$  值继续增大时,推荐性能反而随之降低。由图可知,当参数  $\lambda_r$  的取值为 20 时,MAE 的值最低,算法推荐性能最好,因此将参数  $\lambda_r$  的取值设为 20。

### 3.3 算法性能对比实验

通过与以下几种算法的对比实验来对本文提出的 TTR 算法进行验证。

(1) UserMean: 使用每个用户为每个项目所打分的平均值来预测用户没有得分项。

(2) ItemMean: 使用每个用户为每个项目所打分的平均值来预测用户的评分数据。

(3) 概率矩阵分解 (Probabilistic matrix factorization, PMF)<sup>[17]</sup>: 是由 Salakhutdinov 等提出的一种基于概率的矩阵分解技术。

(4) 非负矩阵分解 (Non-negative matrix factorization, NMF)<sup>[18]</sup>: 是由 Lee 等提出的一种只使用评价矩阵信息的推荐方法。

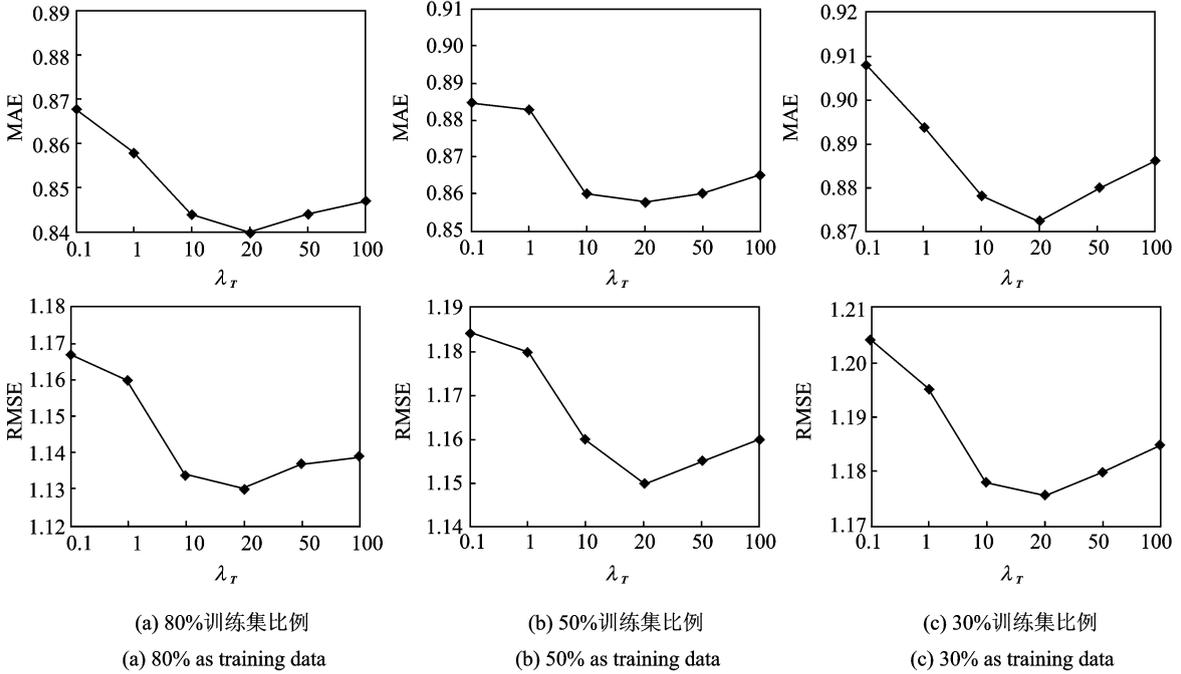


图3 不同训练集比例下参数  $\lambda_r$  对算法性能的影响实验 ( $L=10$ )

Fig. 3 Influence of parameter  $\lambda_r$  to algorithm performance under different training set ratios

(5) 奇异值分解 (Singular value decomposition, SVD)<sup>[19]</sup>: 通过有效处理评分矩阵中的缺失项来提高推荐系统的性能。

(6) 社会网络矩阵分解 (Social matrix factorization, SocialMF)<sup>[20]</sup>: 利用用户兴趣爱好传播进行建模的一种推荐方法, 也是本文的研究基础之一。可是该算法没有考虑不可信节点影响信任建模的情况, 对用户和好友之间的偏好差异没有限制。

实验中使用的参数设置如下:  $\lambda_r = 20$ ; 用户特征矩阵的维度  $L = 10$ ;  $\lambda_u = \lambda_v = 0.1$ 。本文提出的 TTR 算法最大的贡献是在传统社会推荐算法的基础上融入了社会标签信息, 优化了项目属性的构建。为了更进一步验证项目标签比例对算法性能的影响, 首先将训练集中按照项目所被标注的标签数量划分为 5 组, 分别是项目标签数量为“0”, “1-5”, “6-10”, “11-20”, “> 21”, 取 TTR 算法在不同训练集上对不同项目标签数量下, 以 MAE 和 RMSE 作为评测标准进行推荐效果的对比, 实验结果如图 4 所示。

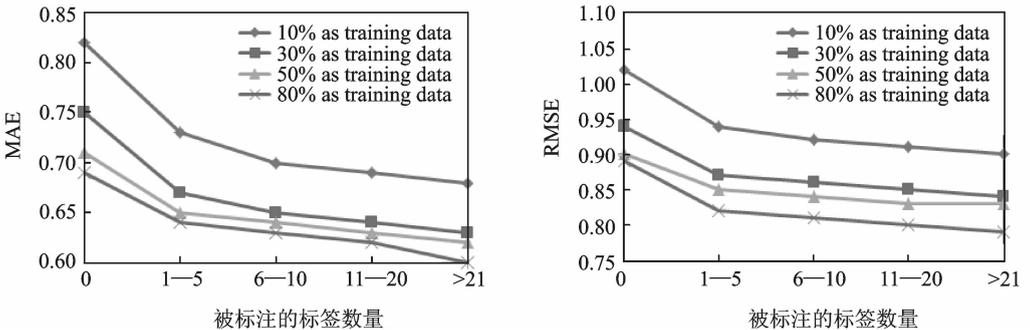


图4 推荐算法在不同训练集比例下的性能比较

Fig. 4 Performance comparison of the recommendation algorithms under different training set ratios

由图 4 可知,本文提出的 TTR 算法由于考虑了项目标签的信息,因此在推荐精度上有显著提高。随着项目被标注的标签信息增大,推荐系统的质量首先显著提升,但随着标签信息量继续增加至 20 以上,系统性能趋于稳定。这个结果也是合理的,因为随着标签信息的增加,项目属性被描述得越准确;但当项目标签过多时,将会导致项目描述信息的冗余,最终实验选择项目标签数量为 20。

表 2 给出了所有算法在 Epinions 和 MovieLens 数据集上的对比实验结果。实验分别在潜在特征向量维度为 5, 10 和 20 三种取值下进行验证。从对比结果可以看出,本文提出的 TTR 算法在两个数据集上的所有维度取值下性能均优于其他算法。

表 2 各种推荐算法在不同数据集上的性能比较

Tab. 2 Performance comparison of various recommendation algorithms on different data sets

数据集	特征维度	指标	UserMean	ItemMean	PMF	NMF	SVD	SocialMF	TTR
Epinions	5	MAE	0.937	0.925	0.911	0.887	0.872	0.862	0.846
		RMSE	1.234	1.223	1.153	1.123	1.102	1.091	1.081
	10	MAE	0.937	0.925	0.909	0.884	0.875	0.860	0.846
		RMSE	1.244	1.229	1.149	1.120	1.110	1.093	1.080
	20	MAE	0.937	0.925	0.916	0.883	0.859	0.852	0.846
		RMSE	1.244	1.229	1.158	1.119	1.092	1.088	1.081
MovieLens	5	MAE	0.833	0.805	0.821	0.783	0.769	0.761	0.756
		RMSE	1.130	1.102	1.048	1.009	0.994	0.988	0.982
	10	MAE	0.833	0.805	0.818	0.785	0.769	0.760	0.756
		RMSE	1.130	1.102	1.044	1.010	0.994	0.990	0.982
	20	MAE	0.833	0.805	0.820	0.782	0.767	0.761	0.755
		RMSE	1.130	1.102	1.046	1.006	0.991	0.987	0.981

## 4 结束语

推荐系统主要以用户历史评分信息为基础,对其喜好进行预测,但是如果当新用户没有任何评分数据或信任关系信息时,推荐系统就无法进行有效的预测,因此冷启动用户是影响推荐系统性能高低的关键因素。社会网络中包含着丰富的标签信息,可以被用于提升社会化推荐的精准度。本文研究的贡献在于将社会标签信息与信任关系信息相结合,提出了一种融合社会标签与信任关系的社会网络推荐算法。该算法最大的优势在于对新用户的评分进行了预处理,并且在构建项目属性时加入了项目被标注的标签信息,从而能在一定程度上提高推荐系统的准确性。实验结果表明,将用户信任关系与项目标签信息相融合的推荐方法在平均绝对误差和均方根误差指标上较其他推荐方法有大幅度的提高,特别是对冷启动用户较多的信任网络效果更加明显。由于社会标签信息和信任关系会随着用户兴趣的迁移会不断变化,因此下一步的研究方向是对社会标签及信任关系的动态演化进行分析研究,从而实现更加高效的实时推荐。

## 参考文献:

- [1] Zhao W, Guan Z, Liu Z. Ranking on heterogeneous manifolds for tag recommendation in social tagging services[J]. *Neurocomputing*, 2015, 148:521-534.
- [2] 李亚克, 田青, 高航. 结合类标签关联度的有序核判别回归学习[J]. *数据采集与处理*, 2016, 31(3):532-540.  
Li Yake, Tian Qing, Gao Hang. Kernel discriminant learning for ordinal regression using label membership[J]. *Journal of Data Acquisition and Processing*, 2016, 31(3): 532-540.
- [3] 秦雨, 余正涛, 王炎冰, 等. 基于特征映射的微博用户标签兴趣聚类方法[J]. *数据采集与处理*, 2015, 30(6):1246-1252.  
Qin Yu, Yu Zhengtao, Wang Yanbing, et al. Micro blog user label interest clustering method based on feature mapping[J]. *Journal of Data Acquisition and Processing*, 2015, 30(6): 1246-1252.

- [4] Ma H, Zhou T C, Lyu M R, et al. Improving recommender systems by incorporating social contextual information[J]. ACM Transactions on Information Systems, 2011, 29(2):219-230.
- [5] Wu H, Pei Y, Li B, et al. Item recommendation in collaborative tagging systems via heuristic data fusion[J]. Knowledge-Based Systems, 2015, 75:124-140.
- [6] Heymann P, Koutrika G, Garcia-Molina H. Can social bookmarking improve web search[C]// Proceedings of the International Conference on Web Search and Web Data Mining. New York, USA: ACM, 2008:195-206.
- [7] Li X, Guo L, Zhao Y E. Tag-based social interest discovery[C]// Proceedings of the International Conference on World Wide Web Conference Series. New York, USA: ACM, 2008:675-684.
- [8] Puglisi S, Parra-Arnau J, Forné J, et al. On content-based recommendation and user privacy in social-tagging systems[J]. Computer Standards & Interfaces, 2015, 41:17-27.
- [9] Gabriel H H, Spiliopoulou M, Nanopoulos A. Summarizing dynamic social tagging systems[J]. Expert Systems with Applications, 2014, 41(2):457-469.
- [10] Wu Z Y, Zou M. Modeling social tagging using latent interaction potential[J]. Physica A Statistical Mechanics & Its Applications, 2014, 413(11):125-133.
- [11] Ramage D, Heymann P, Manning C D, et al. Clustering the tagged web[C]// International Conference on Web Intelligence, DBLP 2009. Milan, Italy: ACM, 2009:54-63.
- [12] Wu P, Zhang Z K. Enhancing personalized recommendations on weighted social tagging networks[J]. Physics Procedia, 2010, 3(5):1877-1885.
- [13] Bobadilla J S, Ortega F, Hernando A, et al. A collaborative filtering approach to mitigate the new user cold start problem [J]. Knowledge-Based Systems, 2012, 26: 225-238.
- [14] 印桂生, 张亚楠, 董宇欣, 等. 基于受限信任关系和概率分解矩阵的推荐[J]. 电子学报, 2013, 42(5): 904-911.  
Yin Guisheng, Zhang Yanan, Dong Yuxin, et al. A constrained trust recommendation using probabilistic matrix factorization [J]. Acta Electronica Sinica, 2013, 42(5): 904-911.
- [15] 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[J]. 计算机研究与发展, 2013, 50(9): 1805-1813.  
Guo Lei, Ma Jun, Chen Zhumin. Trust strength aware social recommendation method[J]. Journal of Computer Research and Development, 2013, 50(9): 1805-1813.
- [16] Noel J, Sanner S, Tran K N, et al. New objective functions for social collaborative filtering[C]// Proceedings of the International Conference on World Wide Web. New York, USA: ACM, 2012: 859-868.
- [17] Salakhutdinov R, Mnih A. Probabilistic matrix factorization[C]// Proceedings of the 21st Annual Conference on Neural Information Processing Systems. New York, USA: Curran Associates, 2008: 1257-1264.
- [18] Lee D D, Seung H S. Learning the parts of objects by non-negative matrix factorization[J]. Nature, 1999, 401(6755):788-791.
- [19] Kurucz M, Benczur A. Methods for large scale SVD with missing values[C]//In Proceedings of the KDD Cup and Workshop. New York, USA:ACM,2007: 31-41.
- [20] 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. New York, USA: ACM, 2010: 135-142.

## 作者简介:



胡云(1978-),女,副教授,博士。研究方向:智能信息处理、社会网络分析和个性化推荐算法研究,E-mail: anniyahy@126.com。



张舒(1979-),通信作者,男,讲师,研究方向:数据挖掘、社会网络推荐,E-mail: ppsshufan@126.com。



李慧(1979-),女,博士,副教授,研究方向:个性化推荐、社会网络分析,E-mail: shufanzs@126.com。



施珺(1963-),女,教授,研究方向:智能信息处理。



仲兆满(1978-),男,副教授,研究方向:中文信息处理。