

一种信任关系强度敏感的社会化推荐算法

郭磊 马军 陈竹敏

(山东大学计算机科学与技术学院 济南 250101)

(leiguo@y7mail.com)

Trust Strength Aware Social Recommendation Method

Guo Lei, Ma Jun, and Chen Zhumin

(School of Computer Science and Technology, Shandong University, Jinan 250101)

Abstract With the advent of social networks, trust-aware recommendation methods have been well studied. Most of these algorithms assume that trusted users will have similar tastes. However, this assumption ignores the fact that two users may establish a trust connection for the social purpose or simply for etiquette, which may not result in similar opinions on the same item. Motivated by this observation, a novel trust strength aware social recommendation method, StrengthMF, is firstly proposed. Compared with previous methods, this new approach assumes that a trust relation does not necessarily guarantee the similarity in preferences between two users. Specifically, StrengthMF learns the trust strength and distinguishes users with more similar interests through the shared user latent feature space, i. e., the user latent feature space in trust network is the same in the rating matrix. This will allow us to acquire a better understanding of the relationship between trust relation and rating similarity. To validate the learned trust strength, InfluenceMF method is then proposed, which retrains SocialMF with estimated trust relations. Experimental results on real world product rating data set Epinions show that the proposed approaches outperform the state-of-the-art algorithms in terms of *RMSE* and *MAE*, and the learned trust strength can further improve traditional recommendation methods.

Key words recommender system; probabilistic matrix factorization; trust strength; social recommendation; collaborative filtering

摘要 为了进一步提高推荐算法的准确率,更好地对用户间的信任关系进行建模,首先提出了一种信任关系强度敏感的社会化推荐算法(StrengthMF).与以往的算法相比,该算法假设建立信任关系的两个用户之间并不一定存在着相似的兴趣爱好.在推荐过程中,StrengthMF算法通过共享的潜在用户特征空间来对信任关系强度和用户兴趣进行建模,通过进一步识别出那些与目标用户有着共同爱好的朋友来对求解的过程进行优化.为了验证算法所估计出的信任关系强度的准确性,接着又在SocialMF算法的基础上,提出了一种使用所估计的信任关系对其重新训练和学习的InfluenceMF算法.实验结果表明,与目前较为流行的方法相比,新方法能在*RMSE*和*MAE*上取得更好的推荐结果,其所推导出的信任关系强度能进一步提高已有推荐算法的性能.

收稿日期:2013-03-30;修回日期:2013-06-17

基金项目:国家自然科学基金项目(61272240,60970047,61103151);教育部高等学校博士学科点专项科研基金项目(20110131110028);山东省自然科学基金项目(ZR2012FM037);山东省优秀中青年科学家科研奖励基金项目(BS2012DX017)

通信作者:马军(majun@sdu.edu.cn)

关键词 推荐系统; 概率矩阵分解; 信任强度; 社会化推荐; 协同过滤

中图法分类号 TP18

随着用户和网络规模的不断扩大, 信息资源的数量呈现几何式增长, 无论是信息的消费者还是生产者都面临着严重的信息过载问题. 而推荐系统作为一种有效的信息过滤技术, 是解决这一问题的的重要手段, 因其能够通过对用户的历史行为数据进行分析, 在对用户的兴趣偏好进行建模的基础上, 为用户主动推荐满足他们需求的资源, 受到了研究者的广泛关注. 目前这类方法已经在很多领域得到了应用, 并带来了巨大的商业价值.

近年来, 随着诸如 Facebook^①, Twitter^② 等社会媒体的兴起, 利用用户间的社会关系进行推荐的方法, 逐渐成为推荐领域的研究热点. 这类方法在推荐过程中假设用户的决策过程容易受到信任关系或朋友关系的影响, 相互信任的朋友之间具有相近的兴趣爱好. 例如, Ma 等人^[1-2] 研究了如何利用用户间的信任关系来进一步提高传统推荐算法的性能, 并给出了一种能够将信任关系信息进行融合的概率矩阵分解框架.

以上这些已有的基于信任关系进行推荐的算法, 虽然能够从模拟现实世界中社会化推荐的角度出发, 对推荐的社会化过程进行建模, 但它们在推荐过程中大都将用户间的信任关系同等对待, 而忽略了信任关系的多样性特点^[3-4]. 然而, 在很多应用场景中, 具有信任关系的两个用户之间并不一定有着共同的兴趣爱好. 例如, 有的用户仅仅是出于社交或者礼貌的目的才同另外一个用户建立信任关系(他们将信任自己的用户之间建立信任关系看作是一种礼貌行为). 此外, 社会网络中链接关系建立的随意性(不需要得到目标用户的同意)也是导致信任关系多样性的重要因素之一. 因此, 如果简单地将所有的信任关系都同等对待, 认为信任关系表示的只是一个用户对另外一个用户兴趣偏好的认同, 将很难发现隐藏在信任关系背后的用户特征和关联关系.

为了进一步提高推荐系统的性能, 针对以上存在的问题, 本文研究了如何对用户间的信任关系强度进行建模, 以及将其融入到传统推荐系统中的方法. 本文首先使用概率矩阵分解的方法对用户间的信任网络进行建模, 从中推导出表示用户间信任关

系的主要特征, 然后在现有推荐方法^[5] (SocialMF) 的基础上, 提出了一种信任关系强度敏感的社会化推荐算法 (StrengthMF). StrengthMF 通过将用户间的信任关系和评分信息通过共享的用户特征空间进行关联, 使得它能在学习过程中进一步识别出与目标用户具有信任关系并且兴趣比较接近的朋友, 从而达到优化推荐结果的目的. 为了验证算法对信任关系估计的准确性, 以及对已有推荐算法的影响, 本文又进一步以 SocialMF 为例, 利用所估计的信任关系值对其重新进行训练和优化. 实验结果表明, 与其他同类型的推荐方法相比, 本文所提出的方法能够在 RMSE 和 MAE 两种评价指标上取得更好的结果.

1 相关工作

本节主要介绍传统的协同过滤算法和与本文相关的几种社会化推荐算法.

传统的协同过滤算法主要是对用户的历史评分数据进行分析, 并且假设用户对某一物品的喜好同与其有着相似兴趣的用户相一致(本文所指的物品可以是电影、音乐和商品等). 一般来说, 协同过滤算法可以分为基于记忆和基于模型两种类型. 基于记忆的算法, 需要将数据全部加载到内存中, 每次推荐都需要使用启发式的规则来重新计算用户或物品^[6-9] 在评分数据上的相似性, 并进而对用户可能的评分进行预测. 然而这种类型的方法随着用户和物品数量的增多, 往往具有较高的计算复杂度.

与基于记忆的推荐算法相比, 基于模型的推荐算法需要事先使用机器学习的方法训练出一个有效的推荐模型, 然后在未知数据上使用训练好的模型进行推荐, 无论是在推荐效果还是预测效率上都有明显提升. 这种类型的方法主要包括潜在特征模型^[10]、聚类模型^[7] 和 aspect 模型^[11-12] 等. 其中, 基于潜在特征的矩阵分解模型^[13-14], 由于能在训练过程中分别为用户和物品推导出与其对应的低维潜在特征向量, 以及在大数据集上所具有的高可扩展性和准确性, 受到了研究者的青睐. 例如, Salakhutdinov

① Facebook: <http://www.facebook.com>

② Twitter: <https://twitter.com>

等人^[15]提出的具有高斯噪声的概率线性模型,研究了如何使用概率的方法从已知评分数据中推导出表示用户和物品的潜在特征向量。

以上这些传统的协同过滤算法虽然已经取得了较好的效果,但它们在推荐过程中都假设用户之间是相互独立的,忽略了用户间的关联关系在许多应用场景中对推荐结果产生的影响。特别是随着社交网络的出现,如何利用更为丰富的上下文信息^[16]和用户间的信任、朋友等社会关系信息^[17-19]进行推荐逐渐成为研究的热点。例如, Ma 等人在文献^[18]中研究了基于信任关系的推荐算法,通过使用信任关系矩阵对目标函数进行约束,在学习过程中使目标用户的特征向量尽可能与其朋友的特征向量接近,从而达到利用社会关系信息进行推荐的目的。Jamali 等人在文献^[19]中提出了一种使用随机游走模型对信任关系进行整合的方法。在随后的工作中,他们又在文献^[1]的基础上,提出了一种基于信任传播的推荐算法^[5]。

目前,已有的社会化推荐算法大都假设用户间的信任关系是单一和同质的,而对信任关系进行显式区分的相关研究还较少。例如, Matsuo 等人^[20]对用户间的信任关系与用户评分之间产生的相互影响力进行建模,并进而使用该影响力作为衡量用户是否加入某个社区的重要指标。另外一个相关工作是 Yeung 等人^[3]研究了 Epinions 网络中用户间的信任关系对他们的最终评分所造成的影响,在实验中他们发现并不是所有建立信任关系的两个用户都会对某一产品有相近的评分。然而,他们的工作更多的是强调如何去发现用户间信任关系的多样性,而不是以提高用户对产品评分预测的准确率为主要优化目标。

2 基于信任关系的推荐方法

本节首先对基于信任关系进行推荐的问题进行说明,接着介绍了本文对用户间的信任关系进行建模的方法。然后,在已有的 SocialMF^[5]方法的基础上,提出了一种能够反映用户间信任关系强度的推荐算法。最后,以 SocialMF 为例,给出了一种利用所估计出的信任关系值对传统的社会化推荐算法进行改进的方法。

2.1 问题描述

本节以社会网络中的物品推荐场景为例对问题进行说明。这里的物品是对电影、音乐和商品等推荐

对象的一般性称谓。在社会网络中,用户总是倾向于选择他们所信任的朋友为其推荐的物品,并且用户之间的信任关系越强,受到朋友影响的可能性也就越大。图 1 展示了基于信任关系的物品推荐场景,从图 1 中可以看出,这一应用场景主要包括两个核心要素:1)用户之间的信任关系;2)用户对物品的历史评分信息。

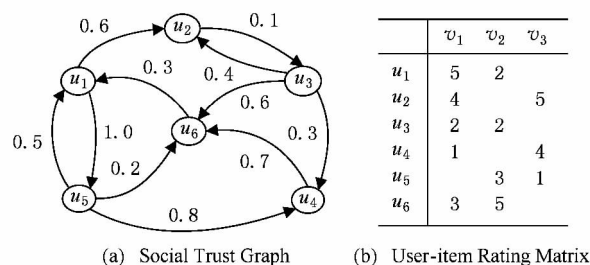


Fig. 1 Example for trust-aware recommendation.

图 1 基于信任关系的推荐场景举例

图 1(a)定义了用户间的信任关系图 $\mathcal{G}=(\mathcal{U}, \mathcal{E})$, 其中 \mathcal{U} 表示用户的集合; $\forall u_1, u_2 \in \mathcal{U}, (u_1, u_2) \in \mathcal{E}$, 表示用户 u_1 对 u_2 的信任关系, 每条信任关系 $(u_1, u_2) \in \mathcal{E}$ 的权重 $s_{ij} \in (0, 1]$ 则表示用户之间信任关系的强度。值得注意的是, 在大多数具有信任关系的网络中, 信任关系通常只以二值的形式出现(0 或 1), 用户间的信任强度很难通过给定的表面值进行区分。图 1(b)定义了用户对物品的评分矩阵 $R=U \times V$; 其中, $\forall u_i \in \mathcal{U}, \forall v_j \in \mathcal{V}, R(u_i, v_j) \in [1, 5]$ 表示 u_i 对 v_j 的喜好程度, 该信息是对用户的兴趣进行建模的重要依据。在这样一个基于信任关系进行推荐的场景中, 算法的目标就是通过所观察到的信任关系网络和历史评分信息, 挖掘出隐藏在信任关系背后的用户特征, 来进一步提高推荐的准确率。

2.2 信任关系的建模方法

为了进一步挖掘出隐藏在信任关系背后的用户特征, 本文引入矩阵分解的方法对用户间的信任关系网络进行分解, 从而推导出一个能表示用户的低维潜在特征向量, 这些特征是刻画两个用户之间信任关系的关键性因素。

假设用户间的信任关系网络用图 $\mathcal{G}=(\mathcal{U}, \mathcal{E})$ 表示, $S=\{s_{ij}\}$ 表示图 \mathcal{G} 的邻接矩阵 $(m \times m)$, $U \in \mathbb{R}^{l \times m}$ 表示矩阵分解模型所推导出的用户特征矩阵, 列向量 U_i 是与其相对应的潜在特征向量。信任关系矩阵的条件概率分布定义为

$$p(S | U, \sigma_s^2) = \prod_{i=1}^m \prod_{k=1}^m [\mathcal{M} s_{ik} | g(U_i^T U_k), \sigma_s^2]^{s_{ik}}, \quad (1)$$

其中, $\mathcal{N}(x|\mu, \sigma^2)$ 表示 x 服从均值为 μ , 方差为 σ^2 的高斯分布, I_{ik}^s 是指示函数, 它表示如果用户 u_i 对 u_k 有信任关系, 则它的值为 1, 否则为 0. $g(x)$ 是逻辑回归函数^①, 将预测值 $(U_i^T U_k)$ 限定在区间 $[0, 1]$ 之内, 并且假设用户的特征矩阵服从均值为 0 的球形高斯先验分布:

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

经过贝叶斯推断, 可以得到:

$$p(U | S, \sigma_S^2, \sigma_U^2) \propto p(S | U, \sigma_S^2) p(U | \sigma_U^2) = \prod_{i=1}^m \prod_{k=1}^m [\mathcal{N}(s_{ik} | g(U_i^T U_k), \sigma_S^2)]^{I_{ik}^s} \times \prod_{i=1}^m \mathcal{N}(U_i | 0, \sigma_U^2 I). \quad (3)$$

在式(3)给出的方法中, 只通过用户间的信任关系来对用户的特征进行估计, 并没有考虑到用户对物品的评分信息. 第 2.3 节将会引入 SocialMF^[5] 方法, 通过对朋友间的兴趣传播进行建模, 来提高推荐的准确率.

2.3 SocialMF 方法

假设用户对物品的评分矩阵 $R = \{r_{ij}\}$ 中有 m 个用户、 n 个物品, 其中 $r_{ij} \in [0, 1]$ 表示相应的评分. 在大多数数据集中, 用户对物品的评分大都使用 1 到 r_{\max} 的整数来表示, 为了不失一般性, 本文使用函数 $f(x) = (x-1)/(r_{\max}-1)$ 将评分区间映射到 $[0, 1]$ 上. 所要推导的用户和物品的 t -维潜在特征矩阵分别用 $U \in \mathbb{R}^{l \times m}$ 和 $V \in \mathbb{R}^{l \times n}$ 表示, 列向量 U_i 和 V_j 则是与其相对应的特征向量.

Jamali 等人提出的 SocialMF 算法假设用户的兴趣爱好更容易受到其所信任的朋友的影响. 也就是说, 用户的潜在特征向量与其朋友(所信任的用户)的潜在特征向量是相关的, 对于任意一个用户 u_i 来说, 可以将其特征向量表示成:

$$\hat{U}_i = \sum_{t \in N_i} s_{it} U_t, \quad (4)$$

其中, \hat{U}_i 是要估计的用户 u_i 的特征向量, $t \in N_i$ 是 u_i 所信任的朋友集合, s_{it} 是用户 u_i 对 u_t 的信任值.

需要注意的是, 这里虽然引入了朋友的兴趣爱好, 但并没有改变评分矩阵 R 的条件分布, 朋友的兴趣爱好信息只对用户的特征向量产生影响. R 的条件概率分布仍然定义为

$$p(R | U, V, \sigma_R^2) = \prod_{i=1}^m \prod_{j=1}^n [\mathcal{N}(r_{ij} | g(U_i^T V_j), \sigma_R^2)]^{I_{ij}^R}, \quad (5)$$

① $g(x) = 1/(1 + \exp(-x))$

其中, I_{ij}^R 是指示函数, 如果用户 u_i 对物品 v_j 进行了评分, 它的值就等于 1, 否则为 0.

用户特征 U 的信息主要表现在以下 2 个方面:

1) 为了防止过拟合, 所采用的均值为 0 的高斯先验;
2) 能反映出用户所信任朋友的特征向量的条件概率分布. 因此, 可以得到:

$$p(U | S, \sigma_U^2, \sigma_T^2) \propto p(U | \sigma_U^2) p(U | S, \sigma_T^2) = \prod_{i=1}^m \mathcal{N}(U_i | 0, \sigma_U^2 I) \times \prod_{i=1}^m \mathcal{N}(U_i | \sum_{t \in N_i} s_{it} U_t, \sigma_T^2 I). \quad (6)$$

假设物品特征矩阵 V 同样服从均值为 0 的高斯分布:

$$p(V | \sigma_V^2) = \prod_{j=1}^n \mathcal{N}(V_j | 0, \sigma_V^2 I). \quad (7)$$

经过贝叶斯推断, 可以进一步得出:

$$p(U, V | R, S, \sigma_R^2, \sigma_U^2, \sigma_V^2) \propto p(R | U, V, \sigma_R^2) p(U | S, \sigma_U^2, \sigma_T^2) p(V | \sigma_V^2) = \prod_{i=1}^m \prod_{j=1}^n [\mathcal{N}(r_{ij} | g(U_i^T V_j), \sigma_R^2)]^{I_{ij}^R} \times \prod_{i=1}^m \mathcal{N}(U_i | \sum_{t \in N_i} s_{it} U_t, \sigma_T^2 I) \times \prod_{i=1}^m \mathcal{N}(U_i | 0, \sigma_U^2 I) \times \prod_{j=1}^n \mathcal{N}(V_j | 0, \sigma_V^2 I). \quad (8)$$

SocialMF 所对应的概率图模型可以用图 2 来表示:

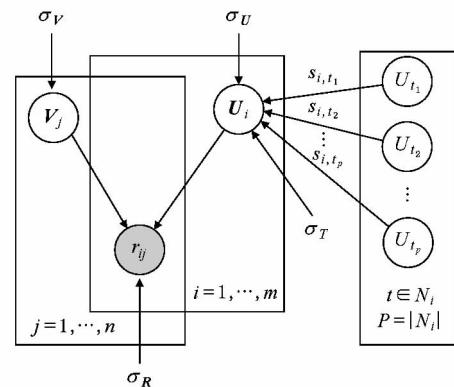


Fig. 2 Graphic model for SocialMF.

图 2 SocialMF 对应的图模型

2.4 信任关系强度敏感的推荐方法

SocialMF 方法通过利用条件概率对每个用户的潜在特征向量进行约束, 来学习用户与其朋友的信任关系对推荐结果造成的影响. 但是, 这种方法在推荐过程中认为用户间的信任关系对用户具有相同

的影响力,并没有考虑到用户间信任关系的形成往往是由于多种因素综合作用的结果,相互信任的两个用户之间并不一定具有相同的兴趣爱好.并且仅仅使用信任关系的表面信息进行推荐,也很难估计出用户真正的兴趣爱好(尤其是在二值信任关系网络中).

为了能对用户间不同的信任关系和兴趣进行建模,更加充分地利用信任关系来进一步提高推荐算法的准确率.本文将用户对物品的评分矩阵信息和用户间的信任关系网络通过共享的用户特征空间进行关联,并提出了一种基于潜在特征向量进行推荐的矩阵分解算法 StrengthMF.在该算法中,共享的潜在用户特征同时受到信任关系网络和评分信息的影响,而信任关系网络反映了用户间不同的信任关系强度,通过对这两部分信息进行联合分解,可以将信任关系的多样性特点通过用户的潜在特征体现出来,从而达到优化推荐结果的目的.算法对应的概率图模型如图 3 所示:

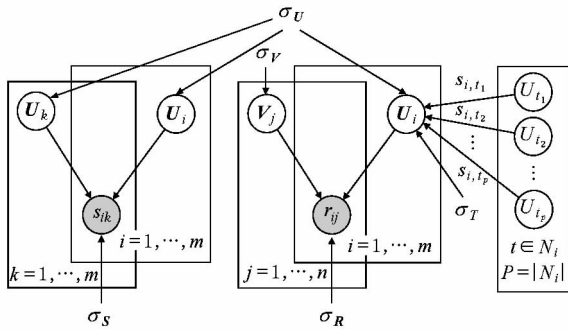


Fig. 3 Graphic model for StrengthMF.

图 3 StrengthMF 对应的概率图模型

参数 U, V 的对数联合后验概率分布可以表示为

$$\begin{aligned} \ln p(U, V | S, R, \sigma_S^2, \sigma_R^2, \sigma_T^2, \sigma_U^2, \sigma_V^2) = & -\frac{1}{2\sigma_R^2} \sum_{i=1}^m \sum_{j=1}^n I_{ij}^R (r_{ij} - g(U_i^T V_j))^2 - \\ & \frac{1}{2\sigma_S^2} \sum_{i=1}^m \sum_{k=1}^m I_{ik}^S (s_{ik} - g(U_i^T U_k))^2 - \\ & \frac{1}{2\sigma_T^2} \sum_{i=1}^m (U_i - \sum_{t \in N_i} s_{it} U_t)^T (U_i - \sum_{t \in N_i} s_{it} U_t) - \\ & \frac{1}{2\sigma_U^2} \sum_{i=1}^m U_i^T U_i - \frac{1}{2\sigma_V^2} \sum_{j=1}^n V_j^T V_j - \\ & \frac{1}{2} ((\sum_{i=1}^m \sum_{j=1}^n I_{ij}^R) \ln \sigma_R^2 + (\sum_{i=1}^m \sum_{k=1}^m I_{ik}^S) \ln \sigma_S^2) - \\ & \frac{1}{2} (ml \ln \sigma_U^2 + nl \ln \sigma_V^2 + ml \ln \sigma_T^2) + C, \quad (9) \end{aligned}$$

其中, C 是与参数无关的常量. 求参数固定时 U, V 的极大后验概率, 相当于最小化以下带正则项的误差平方和函数:

$$\begin{aligned} E(R, S, U, V) = & \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n I_{ij}^R (r_{ij} - g(U_i^T V_j))^2 + \\ & \frac{\lambda_S}{2} \sum_{i=1}^m \sum_{k=1}^m I_{ik}^S (s_{ik} - g(U_i^T U_k))^2 + \\ & \frac{\lambda_T}{2} \sum_{i=1}^m (U_i - \sum_{t \in N_i} s_{it} U_t)^T (U_i - \sum_{t \in N_i} s_{it} U_t) + \\ & \frac{\lambda_U}{2} \|U\|_F^2 + \frac{\lambda_V}{2} \|V\|_F^2, \quad (10) \end{aligned}$$

其中, $\lambda_S = \sigma_R^2 / \sigma_S^2$, $\lambda_T = \sigma_R^2 / \sigma_T^2$, $\lambda_U = \sigma_R^2 / \sigma_U^2$, $\lambda_V = \sigma_R^2 / \sigma_V^2$, $\|\cdot\|_F^2$ 表示 Frobenius 范数. 对于式(10)所示的目标函数, 通过在 U, V 上使用梯度下降的方法进行求解, 可以使目标函数达到局部极小值:

$$\begin{aligned} \frac{\partial E}{\partial U_i} = & \sum_{j=1}^n I_{ij}^R g'(U_i^T V_j) (g(U_i^T V_j) - r_{ij}) V_j + \\ & \lambda_S \sum_{k=1}^m I_{ik}^S g'(U_i^T U_k) (g(U_i^T U_k) - s_{ik}) U_k + \lambda_U U_i + \\ & \lambda_T (U_i - \sum_{t \in N_i} s_{it} U_t) - \lambda_T \sum_{\{t \in N_i\}} s_{it} (U_t - \sum_{w \in N_t} s_{tw} U_w), \\ \frac{\partial E}{\partial V_j} = & \sum_{i=1}^m I_{ij}^R g'(U_i^T V_j) (g(U_i^T V_j) - r_{ij}) U_i + \lambda_V V_j, \quad (11) \end{aligned}$$

其中, $g'(x) = \exp(x) / (1 + \exp(x))^2$ 是逻辑函数 $g(x)$ 的导数.

2.5 InfluenceMF 方法

为了进一步验证算法所估计出的信任关系强度及其对已有推荐算法的影响, 本文以 SocialMF 算法为例, 使用推导出的信任关系矩阵 $S^* = \{s_{ij}^*\}$ 来代替用户间的直接信任关系 S , 并将该算法定义为 InfluenceMF. 由式(10)可知, 通过对信任网络和评分矩阵进行联合分解, 可以得到反映用户间真实信任关系的用户特征矩阵 U . 算法所估计的信任关系矩阵 S^* 可由式(12)来计算:

$$S^* = \sum_{i=1}^m \sum_{k=1}^m I_{ik}^S g(U_i^T U_k). \quad (12)$$

与 SocialMF 算法类似, InfluenceMF 使用如下所示的目标函数重新进行训练:

$$\begin{aligned} E(R, S^*, U^*, V^*) = & \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n I_{ij}^R (r_{ij} - \\ & g(U_i^{*T} V_j^*))^2 + \frac{\lambda_T^*}{2} \sum_{i=1}^m (U_i^* - \sum_{t \in N_i} s_{it}^* U_t^*)^T (U_i^* - \\ & \sum_{t \in N_i} s_{it}^* U_t^*) + \frac{\lambda_U^*}{2} \|U^*\|_F^2 + \frac{\lambda_V^*}{2} \|V^*\|_F^2, \quad (13) \end{aligned}$$

其中, U^* 和 V^* 表示推导出的用户和物品的特征, $\lambda_T^* = \sigma_R^2 / \sigma_T^{*2}$, $\lambda_U^* = \sigma_R^2 / \sigma_U^{*2}$, $\lambda_V^* = \sigma_R^2 / \sigma_V^{*2}$. 通过在 U^*, V^* 上使用梯度下降的方法进行求解, 可以使式(13)所示的目标函数达到局部最小值:

$$\begin{aligned} \frac{\partial E}{\partial \mathbf{U}_i^*} &= \sum_{j=1}^n I_{ij}^R g'(\mathbf{U}_i^{*T} \mathbf{V}_j^*) (g(\mathbf{U}_i^{*T} \mathbf{V}_j^*) - r_{ij}) \mathbf{V}_j^* + \\ &\lambda_r^* (\mathbf{U}_i^* - \sum_{t \in N_i} s_{it}^* \mathbf{U}_t^*) - \lambda_r^* \sum_{\{t \in N_i\}} s_{it}^* (\mathbf{U}_t^* - \sum_{w \in N_t} s_{tw}^* \mathbf{U}_w^*), \\ \frac{\partial E}{\partial \mathbf{V}_j^*} &= \sum_{i=1}^m I_{ij}^R g'(\mathbf{U}_i^{*T} \mathbf{V}_j^*) \cdot \\ &(g(\mathbf{U}_i^{*T} \mathbf{V}_j^*) - r_{ij}) \mathbf{U}_i^* + \lambda_v^* \mathbf{V}_j^*. \end{aligned} \quad (14)$$

3 实验结果与分析

本节首先通过几组实验比较了本文所提出的方法与另外几种方法在推荐性能上的差异,然后进一步分析了 StrengthMF 方法中实验参数 λ_s 对推荐结果的影响。

3.1 数据集和评价方法

在实验中,本文选择 Epinions^① 数据集对算法进行验证. Epinions.com 成立于 1999 年,是一个知名的产品评论网站,它的主要目标是帮助用户根据其他人对产品的评论做出更加客观的选择. 在这个网站上,用户既可以对产品进行打分(1~5 之间的整数),也可以对产品使用文字进行评论. 另外,每个用户都可以通过维护一个信任关系列表来表达对其他用户的看法,并进而同自己所信任的用户之间建立起一种信任关系网络. 在 Epinions.com 中,两个用户之间信任关系是二值的(0 或 1),即要么信任,要么不信任。

实验中所使用的 Epinions 数据集是由 Massa 等人在文献[21]中发布的,其统计特性如表 1 所示. 该数据集共包含了 65 269 个用户对 175 156 个产品的 664 824 条评分信息(至少评过一次分)和用户之间的 487 181 条信任关系信息(该数据集没有包括不信任关系信息)。

为了验证算法的有效性,本文使用两种较为常见的指标^[22] 均方根误差(root mean square error, RMSE)和平均绝对误差(mean absolute error, MAE)进行评价. RMSE 和 MAE 是衡量评分预测准确性的标准,它们反映的是算法的预测评分与用户实际评分的贴近程度,其值越低,表示算法的推荐性能越好。

RMSE 的公式可以表示为

$$RMSE = \sqrt{\frac{1}{N} \sum_{i,j} (R_{ij} - \hat{R}_{ij})^2},$$

其中, R_{ij} 和 \hat{R}_{ij} 分别表示用户 u_i 对产品 v_j 评分的真实值和预测值, N 是测试样例的数量. 与 RMSE 的

定义类似, MAE 的公式可以表示为:

$$MAE = \frac{1}{N} \sum_{i,j} |R_{ij} - \hat{R}_{ij}|.$$

Table 1 Statistics of Epinions Dataset

表 1 Epinions 数据集的统计特性

| Statistics | Training | Test |
|--------------|----------|----------|
| User Count | 38 494 | 26 775 |
| Item Count | 123 083 | 52 073 |
| Rating Count | 531 988 | 132 836 |
| Sparsity/% | 99.988 7 | 99.990 4 |

3.2 结果比较

为了展示 StrengthMF 算法在推荐性能上的提升,本文同以下方法进行了比较:

1) GlobalAverage: 该方法使用所有评分的平均值对用户的评分进行预测。

2) UserAverage: 该方法分别使用每个用户的评分均值对未知的评分数据进行预测。

3) UserItemBias: 该方法是 Koren 等人在文献[23]中提出的,它通过对用户的历史评分数据进行学习,推导出用户和物品对于评分的偏好信息,并利用这些偏好信息进行推荐。

4) PMF: 该方法是由 Salakhutdinov 等人在文献[15]中提出的一种基于概率的矩阵分解方法,在推荐过程中,它只使用了用户对物品的评分信息进行推荐。

5) SocialMF: 该方法是由 Jamali 等人在文献[5]中提出的一种对用户之间的兴趣传播进行建模的算法,与本文的工作比较相关,但是该算法只利用了用户间信任关系的表面信息,而忽略了隐藏在信任关系背后的更深层次的用户特征。

实验随机选择 90% 的 Epinions 评分数据作为训练集,10% 作为测试集. 方法中用到的参数设置如下: $\lambda_s=10$, $\lambda_r=5$, $\lambda_u=\lambda_v=15$. 表 2 给出了特征向量的维度分别为 5 和 10 时的实验结果. 从表 2 中可以看出,基于概率矩阵分解的方法在 RMSE 指标上要明显优于另外 3 种基本的推荐方法. 但是,其中只考虑用户评分信息的 PMF 方法比基于信任关系进行推荐的性能要差. 这个结果表明,仅仅使用用户的评分信息进行推荐是不够的,用户间的信任关系信息是影响推荐结果的重要因素. 从表 2 中还可以看出,与 SocialMF 方法相比,本文提出的 StrengthMF

① Epinions: <http://www.epinions.com>

方法在 *RMSE* 和 *MAE* 两种评价指标上都能取得更好的结果,与简单的使用用户间信任关系的表面

信息进行推荐相比,本文提出的方法更能体现出社会网络的特点,能更好地对用户兴趣信息进行建模.

Table 2 Performance Comparisons

表 2 性能比较

| Dimensionality | Metrics | Global Average | User Average | UserItem Bias | PMF | SocialMF | StrengthMF |
|----------------|---------|----------------|--------------|---------------|---------|----------|------------|
| 5 | MAE | 0.9163843 | 0.9313696 | 0.81784608 | 0.82978 | 0.81316 | 0.81161 |
| | RMSE | 1.205837 | 1.2060234 | 1.0513043 | 1.04997 | 1.04262 | 1.03978 |
| 10 | MAE | 0.9163843 | 0.9313696 | 0.81784608 | 0.83025 | 0.81261 | 0.81058 |
| | RMSE | 1.20583 | 1.2060234 | 1.0513043 | 1.05 | 1.04249 | 1.03967 |

3.3 参数的影响

在 StrengthMF 方法中,参数 λ_s 起到了重要的作用,它控制着该方法中信任关系在整个推荐过程中所占的比重,也就是对信任关系的依赖程度.如果 $\lambda_s=0$,该方法将只使用信任关系的表面信息进行推荐,而忽略了不同用户间信任关系的差异;当 $\lambda_s \rightarrow \infty$ 时,该方法将仅使用用户间的信任关系网络去推断目标用户的特征向量;在其他情况下,该方法将在训练过程中综合使用用户间的评分信息和信任关系强度去推导用户和物品的特征向量.

图 4 给出了参数 λ_s 的变化对评价指标 *RMSE* 和 *MAE* 的影响.从该图中可以发现 λ_s 的取值对实验结果有较为显著的影响,随着 λ_s 取值的不同,推荐的结果也发生了较为明显的变化.当 λ_s 从 0.1 开始增加时,*RMSE* 和 *MAE* 的值逐渐降低(推荐的性能越好),但是当 λ_s 的值超过某个阈值之后(在本文中是 10),它们的值开始逐渐增加(说明推荐的性能开始降低).推荐结果随着参数 λ_s 的变化情况,也正好反映了算法所增加的信任关系信息对推荐结果产生的影响,这部分信息在推荐算法中所占的比重,直接决定了最终推荐结果的好坏.

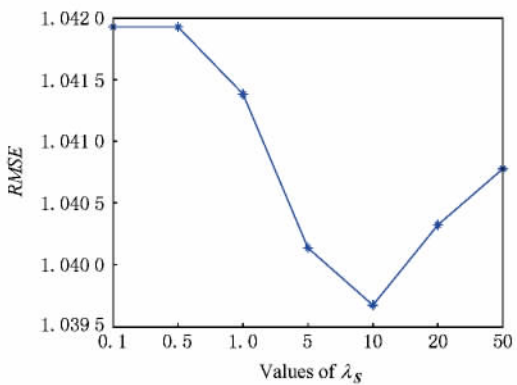
3.4 验证信任关系的准确性

为了验证 StrengthMF 方法所估计出的信任关系的准确性和对已有推荐方法的影响,本节进一步比较了 InfluenceMF 和 SocialMF 在推荐性能上的差异.实验中 InfluenceMF 方法的参数设置如下: $\lambda_{s^*}=5, \lambda_{u^*}=\lambda_{v^*}=15$.表 3 给出了这两种方法在特征向量的维度分别取 5 和 10 的情况下的实验结果.从表 3 中可以看出,在这两种不同的维度下,与 SocialMF 方法相比,InfluenceMF 方法在 *RMSE* 和 *MAE* 指标上都能取得更好的推荐结果.这个实验结果表明,在 StrengthMF 方法中所估计出的信任关系能够反映用户间信任关系的紧密程度和兴趣爱好,并且利用这部分信息能够进一步改善传统推荐算法的性能.

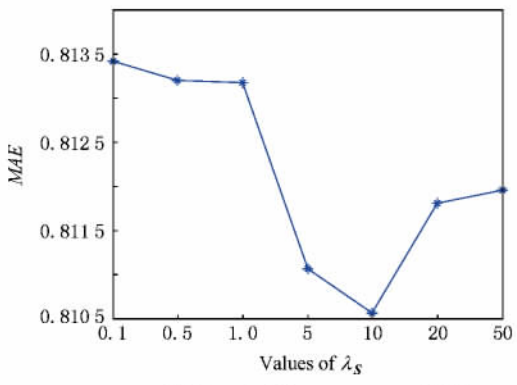
Table 3 Accuracy of Trust Strength

表 3 信任关系强度的准确性

| Dimensionality | Metrics | SocialMF | InfluenceMF |
|----------------|---------|----------|-------------|
| 5 | MAE | 0.81316 | 0.81201 |
| | RMSE | 1.04262 | 1.04148 |
| 10 | MAE | 0.81261 | 0.81143 |
| | RMSE | 1.04249 | 1.04137 |



(a) Impact of λ_s parameter on *RMSE*



(b) Impact of λ_s parameter on *MAE*

Fig. 4 Impact of parameter λ_s (Dimensionality=10).

图 4 参数 λ_s 对实验结果的影响(维度=10)

4 结论和工作展望

本文针对已有的推荐算法不能很好地挖掘用户间真实信任关系的缺点,首先提出了一种对用户间的信任关系进行建模的方法,然后在已有工作的基础上,给出了一种共享用户特征空间的联合概率矩阵分解方法.本文所提出的方法假设用户间信任关系的建立是由于多种因素综合作用的结果,相互信任的两个用户之间并不一定具有相似的兴趣爱好.在 Epinions 数据集上的实验结果表明,用户间真实的信任关系强度在推荐过程中起到了重要作用,是用户选择推荐物品的重要依据之一.另外,本文的方法虽然是在信任关系网络的环境中提出的,但是其对信任关系进行建模的思路,同样也可以应用到其他类型的社会网络和研究课题中,例如,社会化搜索和个性化广告推荐等.

为了更加突出问题本身,本文只考虑了用户间信任关系网络的结构信息,而忽略了用户的其他上下文信息对推荐结果的影响,如用户的地理位置、评论信息和人口统计属性等.未来的工作将集中在如何在已有的推荐算法中使用更多的上下文信息,以便进一步改善推荐的效果.

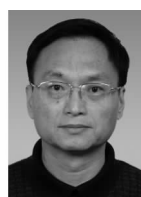
参 考 文 献

- [1] Ma H, King I, Lyu M. Learning to recommend with social trust ensemble [C] //Proc of the 32nd ACM Int Conf on Research and Development in Information Retrieval. New York: ACM, 2009: 203-210
- [2] Ma H, Yang H, Lyu M, et al. Sorec: Social recommendation using probabilistic matrix factorization [C] //Proc of the 17th ACM Conf on Information and Knowledge Management. New York: ACM, 2008: 931-940
- [3] Yeung A C, Iwata T. Strength of social influence in trust networks in product review sites [C] //Proc of the 4th ACM Int Conf on Web Search and Data Mining. New York: ACM, 2011: 495-504
- [4] Yang X, Steck H, Liu Y. Circle-based recommendation in online social networks [C] //Proc of the 18th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2013: 1267-1275
- [5] Jamali M, Ester M. A matrix factorization technique with trust propagation for recommendation in social networks [C] //Proc of the 4th ACM Conf on Recommender Systems. New York: ACM, 2010: 135-142
- [6] Jin R, Chai J, Si L. An automatic weighting scheme for collaborative filtering [C] //Proc of the 27th Annual Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2004: 337-344
- [7] Xue G, Lin C, Yang Q, et al. Scalable collaborative filtering using cluster-based smoothing [C] //Proc of the 28th Annual Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2005: 114-121
- [8] Li Cong, Liang Changyong, Ma Li. A collaborative filtering recommendation algorithm based on domain nearest neighbor [J]. Journal of Computer Research and Development, 2008, 45(9): 1532-1538 (in Chinese)
(李聪, 梁昌勇, 马丽. 基于领域最近邻的协同过滤推荐算法 [J]. 计算机研究与发展, 2008, 45(9): 1532-1538)
- [9] Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms [C] //Proc of the 10th Int Conf on World Wide Web. New York: ACM, 2001: 285-295
- [10] Canny J. Collaborative filtering with privacy via factor analysis [C] //Proc of the 25th Annual Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2002: 238-245
- [11] Hofmann T. Collaborative filtering via gaussian probabilistic latent semantic analysis [C] //Proc of the 26th Annual Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2003: 259-266
- [12] Si L, Jin R. Flexible mixture model for collaborative filtering [C] //Proc of the 20th Int Conf on Machine Learning. Palo Alto: AAAI, 2003: 704-711
- [13] Rennie J, Srebro N. Fast maximum margin matrix factorization for collaborative prediction [C] //Proc of the 22nd Int Conf on Machine Learning. New York: ACM, 2005: 713-719
- [14] Salakhutdinov R, Mnih A. Bayesian probabilistic matrix factorization using markov chain monte carlo [C] //Proc of the 25th Int Conf on Machine Learning. New York: ACM, 2008: 880-887
- [15] Salakhutdinov R, Mnih A. Probabilistic matrix factorization [C] //Proc of the 21st Annual Conf on Neural Information Processing Systems. New York: Curran Associates Inc, 2008: 1257-1264
- [16] Wang Licai, Meng Xiangwu, Zhang Yujie. Context-aware recommender systems [J]. Journal of Software, 2012, 23(1): 1-20 (in Chinese)
(王立才, 孟祥武, 张玉洁. 上下文感知推荐系统 [J]. 软件学报, 2012, 23(1): 1-20)
- [17] Liu F, Lee H. Use of social network information to enhance collaborative filtering performance [J]. Expert Systems with Applications, 2010, 37(7): 4772-4778
- [18] Ma H, Zhou D, Liu C, et al. Recommender systems with social regularization [C] //Proc of the 4th ACM Int Conf on Web Search and Data Mining. New York: ACM, 2011: 287-296

- [19] Jamali M, Ester M. Trustwalker: A random walk model for combining trust-based and item-based recommendation [C] // Proc of the 15th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2009: 397-406
- [20] Matsuo Y, Yamamoto H. Community gravity: Measuring bidirectional effects by trust and rating on online social networks [C] // Proc of the 18th Int Conf on World Wide Web. New York: ACM, 2009: 751-760
- [21] Massa P, Avesani P. Trust-aware recommender systems [C] // Proc of the 2007 ACM Conf on Recommender Systems. New York: ACM, 2007: 17-24
- [22] Zhu Yuxiao, Lü Linyuan. Evaluation metrics for recommender systems [J]. Journal of University of Electronic Science and Technology of China, 2012, 41(2): 164-175 (in Chinese)
(朱郁筱, 吕琳媛. 推荐系统评价指标综述[J]. 电子科技大学学报, 2012, 41(2): 164-175)
- [23] Koren Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model [C] // Proc of the 14th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2008: 426-434



Guo Lei, born in 1983. PhD candidate in Shandong University. Student member of China Computer Federation. His research interests include information retrieval, web data mining and recommender systems.



Ma Jun, born in 1956. Professor and PhD supervisor in Shandong University. Senior Member of China Computer Federation. His research interests include information retrieval, data mining, parallel computing and natural language processing.



Chen Zhumin, born in 1977. Associate professor and master supervisor in Shandong University. Senior Member of China Computer Federation. His research interests include web information retrieval, data mining and social computing (chenzhumin@sdu.edu.cn).