

一种基于标签迁移学习的改进正则化奇异值分解推荐算法

方耀宁^{*} 郭云飞 丁雪涛 兰巨龙

(国家数字交换系统工程技术研究中心 郑州 450002)

(清华大学软件工程学院 北京 100084)

摘 要: 基于正则化奇异值分解(RSVD)的推荐算法在预测准确性上具有明显优势,但存在计算量大的缺点,一定程度上限制了其在实际工程中的应用。针对这个问题,该文提出一种基于标签迁移学习的改进 RSVD 推荐算法,在相对稠密的辅助数据集上根据标签信息来提取用户/项目特征,然后将用户/项目特征应用到 RSVD 算法中,在目标数据集上进行评分预测。在 MovieLens 数据集上的实验结果表明,该算法能够缓解数据稀疏性问题,显著降低预测误差(约 0.01 RMSE),同时减少约 50%的训练时间。

关键词: 计算机网络;推荐系统;协同过滤;正则化奇异值矩阵分解(RSVD);迁移学习

中图分类号: TP393

文献标识码: A

文章编号: 1009-5896(2013)12-3046-05

DOI: 10.3724/SP.J.1146.2013.00290

An Improved Regularized Singular Value Decomposition Recommender Algorithm Based on Tag Transfer Learning

Fang Yao-ning Guo Yun-fei Ding Xue-tao Lan Ju-long

(National Digital Switching System Engineering and Technological R&D Center, Zhengzhou 450002, China)

(School of Software, Tsinghua University, Beijing 100084, China)

Abstract: The recommender algorithm based on Regularized Singular Value Decomposition (RSVD) has significant advantages in predictive accuracy, while it is computationally intensive, which limits greatly its application to engineering projects. To address this issue, an improved algorithm based on tag transfer learning is proposed. It leverages tag information in the relatively denser auxiliary dataset to extract user/item features, which are further used in the RSVD approach in order to make recommendation in the target dataset. Experiments on MovieLens datasets show that the proposed algorithm can handle the sparsity issue effectively, achieve far better prediction results (reducing about 0.01 RMSE), and save about 50% training time at the same time.

Key words: Computer network; Recommender system; Collaborative filtering; Regularized Singular Value Decomposition; Transfer learning

1 引言

随着 Web2.0 技术的发展和应用,极大丰富网络信息的同时也导致了信息过载。推荐系统旨在主动为用户提供个性化的信息过滤服务,其主要分为两大类:基于内容的推荐算法和协同过滤推荐算法。前者直接分析用户的个人信息、项目的内容;后者关注用户的历史行为信息。在 Netflix Prize 中,基于正则化奇异值分解(RSVD)的协同过滤推荐算法在准确性和稳定性上得到了广泛认可^[1]。然而,协同过滤算法面临严重的稀疏性问题,实际应用中数据集往往非常稀疏,无法准确提取用户偏好。基于

RSVD 的推荐算法还面临训练时间偏长、过拟合等问题,很难直接应用到实际项目中^[2]。由于实际应用中用户和项目的数量都是以万为单位的,评分矩阵中元素的个数上亿,所以降低训练时间对于推荐系统有重要意义。

在大数据研究背景下,不同领域数据集的交叉复用成为新的研究热点。迁移学习是从辅助数据集(Auxiliary Dataset, AD)中提取“知识”,用于提高目标数据集(Target Dataset, TD)学习任务的一类方法。辅助数据集在特征空间上与目标数据集具有重叠部分,能够丰富目标数据集的信息。迁移学习算法广泛应用于数据挖掘、计算机视觉、推荐系统等领域,一般分为基于模型、实例和特征 3 类^[3]。

为了缓解稀疏性问题,减少训练时间,本文提

2013-03-07 收到,2013-07-18 改回

国家 973 计划项目(2012CB315901)和国家 863 计划项目(2011AA01A103)资助课题

*通信作者:方耀宁 fyn07@163.com

出一种基于标签迁移学习的改进 RSVD 算法 TT-SVD(Tag Transfer SVD)。TT-SVD 适用于 AD 具有丰富标签信息,而 TD 缺乏标签信息、评分数据较稀疏的情形。例如,豆瓣网含有丰富电影标签和评论信息,但不提供电影服务;优酷网提供电影服务,但电影的标签和评论信息较少;融合两个网站的数据,能够有效提高服务质量。在 MovieLens 1 M 和 10 M 上的仿真结果表明:TT-SVD 只需要约 50% 的训练时间就能达到与 RSVD 相同的预测效果,有效缓解数据稀疏性问题。

2 RSVD 推荐算法

矩阵分解算法从评分矩阵 R 中同时提取用户和项目的特征矩阵 M 和 N 。推荐算法把 R 当作有缺失数据的矩阵,根据 R 中的已知数据估计出 M 和 N ,最后根据 MN^T 预测缺失的数据。特征因子一致性强的用户和项目组合就会得到较高的预测评分。基于矩阵分解的算法主要分为 3 类:奇异值分解^[4]、非负矩阵分解^[5]、概率矩阵分解^[6],其中奇异值分解算法最为常用。

奇异值分解(Singular Value Decomposition, SVD)是一种常用的矩阵分解方法,将目标矩阵 R 分解成为

$$R = X \Sigma Y^T = X \Sigma^{1/2} (Y \Sigma^{1/2})^T \quad (1)$$

其中 X, Y 为正交矩阵, Σ 是对角矩阵。由此可得 R 的低阶近似:

$$R_k = X_k \Sigma_k Y_k^T = MN^T, \quad (2)$$

$$M = X_k \Sigma_k^{1/2}, N = Y_k \Sigma_k^{1/2}$$

RSVD 算法在目标函数中加入了正则化约束项来防止过拟合:

$$\min E = \|MN^T - R\|^2 + \lambda (\|M\|^2 + \|N\|^2) \quad (3)$$

其中 λ 为正则化参数, M 和 N 是 K 维矩阵, K 为特征维度^[1,4]。一种简单的优化方式是在 RSVD 算法基础上增加用户和项目偏置(bias),文献[1]进一步提出利用用户隐含反馈信息的 SVD++ 和 timeSVD++ 算法,取得了较高的预测准确性^[4]。矩阵分解的训练方法主要包括随机梯度下降法、交替最小二乘,马尔科夫链蒙特卡罗。随机梯度下降法比较简单,可以方便应用到较复杂的模型中^[7]。

最近的研究成果表明,迁移学习能够有效缓解协同过滤中的稀疏性问题。文献[8]提出在 AD 中筛选与 TD 特性一致的“实例”来丰富 TD 中信息,文献[9]从“主题层面”来解决 AD 与 TD 连接稀疏的问题,文献[10]提出从 AD 中得到不确定性评分(评分范围或分布)作为评分预测的附加约束条件,

文献[11]使用层次聚类方法处理 AD 中的标签信息,改进了基于用户的协同过滤算法。基于张量分解的推荐算法虽然也可以对辅助信息进行建模^[12-14],但计算量非常大,难以应用到实际项目,并且容易导致过拟合问题。本文提出的 TT-SVD 算法将基于标签的迁移学习算法和矩阵分解算法结合在一起,能够显著缓解数据稀疏性,降低计算复杂度,提高预测准确性。

3 TT-SVD

实际应用中,TD 的评分矩阵往往是非常稀疏,评分密度在 1% 以下,很难直接应用矩阵分解等成熟的协同过滤推荐算法^[1,15]。TT-SVD 算法的原理是:首先,在辅助数据集中用标签来表示项目的特征;然后,根据用户选择的项目集合来计算用户的特征;之后,在目标数据集中根据选择项目的用户群体来确定项目的特征;之后,在目标数据集上对用户/项目特征进行平滑处理,消除不同数据集评分尺度的影响;最后,将用户/项目特征应用到 RSVD 算法,起到提高预测准确性和降低训练时间的效果。

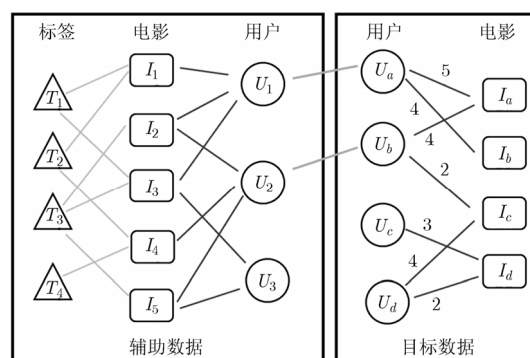


图 1 标签迁移学习示意图

TT-SVD 算法的详细步骤如下所示:

步骤 1 对 AD 中的标签信息进行预处理。理想情况下,AD 中的项目具有明确的分类标签,如电影标签悬疑、动作、剧情等,标签特征矩阵 TF 为单位矩阵,标签 i 的特征向量为 TF_i 。层次聚类的方法有很多,其核心在于对“距离”的定义,LSA 算法的难点在于参数选择和计算复杂度,对层次聚类和 LSA 算法的详细讨论超出了本文的研究范围^[11,16]。

步骤 2 在 AD 中计算用户/项目的特征。AD 中项目 i 的特征向量 IF_i 直接由项目 i 具有的标签特征向量合成,用户 u 的特征向量 M_u 由用户 u 选择过的项目集合的特征分布决定^[4]。

$$\left. \begin{aligned} \mathbf{I}\mathbf{F}_i &= \sum_{j \in T_i} \mathbf{T}\mathbf{F}_j \\ M_u &= \left[\sum_{i \in I_u} \mathbf{I}\mathbf{F}_i \right]^{-1} \sum_{i \in I_u} \mathbf{I}\mathbf{F}_i \end{aligned} \right\} \quad (4)$$

其中 T_i 表示项目 i 的标签集合, I_u 表示用户 u 选择的项目集合, $u \in U_{AD} \cap U_{TD}$, $i \in I_{AD}$ 。

步骤 3 在 TD 中计算项目的特征 目标数据集中项目 i 的特征由选择过该项目的用户集合的特征分布决定。

$$N_i = \left[\sum_{u \in U_i} M_u \right]^{-1} \sum_{u \in U_i} M_u \quad (5)$$

其中 U_i 表示选择过项目 i 的用户集合, $u \in U_{AD} \cap U_{TD}$, $i \in I_{TD}$ 。

步骤 4 在 TD 中对迁移学习的特征进行平滑处理 不同数据集合的评分范围可能是不同的, 不同用户的评分尺度也不同, 因子分解模型的特征维度 K 也会影响用户/项目特征的取值。因此, 迁移学习的用户/项目特征首先要经过平滑处理, 以适应不同的目标数据集合。

首先, 对用户/项目特征向量进行尺度变化:

$$\left. \begin{aligned} M_u &\leftarrow \sqrt{K} \cdot M_u \\ N_i &\leftarrow \sqrt{K} \cdot N_i \end{aligned} \right\} \quad (6)$$

然后, 根据目标数据集中已知交叉用户的评分 R_{ui} , 用一步梯度下降方法对用户特征 N 和项目特征 M 进行平滑处理:

$$\left. \begin{aligned} M_u &\leftarrow M_u + \alpha \cdot ((R_{ui} - M_u N_i^T) \cdot M_u) \\ N_i &\leftarrow N_i + \alpha \cdot ((R_{ui} - M_u N_i^T) \cdot N_i) \end{aligned} \right\} \quad (7)$$

其中 α 为梯度下降步长, 是 TT-SVD 算法引入的唯一附加参数, 在实验中发现 α 在 0.05 附近取值时能够取得最好的效果。

步骤 5 将迁移学习的用户/项目特征应用到 RSVD 从辅助数据集中迁移学习的用户/项目特征不可能完全刻画目标数据集合的特征, 一方面是因为辅助数据集合和目标数据集合只是在特征空间上有部分重叠, 另一方面在迁移学习过程中也可能会丢失有效特征信息。因此, 选择矩阵分解模型特征维度 K 略大于迁移学习的特征维度 Z , 用扩展的 $(K-Z)$ 个维度来捕捉 AD 没有迁移学习到的特征, 经扩展后的用户/项目特征矩阵分别用 M^X 和 N^X 表示。

$$\left. \begin{aligned} M_{uf}^X &\leftarrow M_{uf}, \quad u \in U_{AD} \cap U_{TD}, \quad f \leq Z \\ N_{if}^X &\leftarrow N_{if}, \quad i \in I_{TD}, \quad f \leq Z \end{aligned} \right\} \quad (8)$$

M^X, N^X 中其它的特征因子初始化为随机值。

步骤 6 使用随机梯度下降法(SGD)进行训练

对于训练集中的所有评分 R_{ui} , 用迭代式(9)描述。

$$\left. \begin{aligned} e_{ui} &= R_{ui} - M_u^X (N_i^X)^T \\ M_{uf}^X &\leftarrow M_{uf}^X + lr \cdot (e_{ui} \cdot M_{uf}^X - \lambda N_{if}^X) \\ N_{if}^X &\leftarrow N_{if}^X + lr \cdot (e_{ui} \cdot N_{if}^X - \lambda M_{uf}^X) \end{aligned} \right\} \quad (9)$$

其中 lr 为学习速率, λ 为正则化系数。

步骤 2 利用项目特征来计算用户特征, 而步骤 3 利用用户特征来计算项目特征, 这样就实现了以标签信息为纽带交替计算用户、项目特征。

4 实验设计及结果分析

根据文献[11], 对 MovieLens 数据集合进行预处理模拟构造 AD 和 TD: 用 MovieLens 1 M 数据集作为目标数据集合, 用标签信息相对稠密的 MovieLens 10 M 数据集作为辅助数据集合。MovieLens 10 M 中大约有 3000 个用户与 MovieLens 1 M 中的用户相同, 只利用电影的标签信息和用户的观看记录, 以此来模拟辅助迁移学习过程。本文采用均方根误差 RMSE 作为预测准确性评价指标, 使用半衰期效用指标 HLU 作为排序准确性指标^[17]。

$$\left. \begin{aligned} MAE &= \left| S \right|^{-1} \sum_{r_{ij} \in S} |M_i N_j^T - r_{ij}| \\ RMSE &= \sqrt{\left| S \right|^{-1} \sum_{r_{ij} \in S} (M_i N_j^T - r_{ij})^2} \end{aligned} \right\} \quad (10)$$

$$\left. \begin{aligned} HLU_u &= \sum_{i \in V(u, K)} \frac{\max(r_{ui} - d, 0)}{2^{(l_{ui}-1)/(h-1)}} \\ HLU &= 100 \frac{\sum_{u \in U} HLU_u}{\sum_{u \in U} HLU_u^{\max}} \end{aligned} \right\} \quad (11)$$

其中 d 一般设为系统评分平均值, h 是用户关注度为 50% 的位置, $V(u, L)$ 表示推荐给用户 u 的项目列表, L 为推荐列表长度, l_{ui} 为项目 i 在推荐列表中的位置。

实验中采用的 RSVD 参数: 学习率为 0.005, 正则化系数为 0.02, 特征空间维度 $K=20$ ^[1]。从目标数据集中随机抽取 $x\%$ 样本作为训练集合, 其余 $(1-x\%)$ 作为测试集合。设计了两组实验分别在评分稀疏和稠密条件下对比 TT-SVD 和 RSVD 的性能。

A 组评分稀疏情况下的实验结果如图 2 所示, 图 2(a)和图 2(b)在不同评分密度下对比了 TT-SVD 和 RSVD 算法的 RMSE 和 HLU 指标, 实验结果表明 TT-SVD 的 RMSE 能够降低约 0.02, HLU 提高约 0.02, 明显缓解了数据稀疏性对推荐性能的影响;

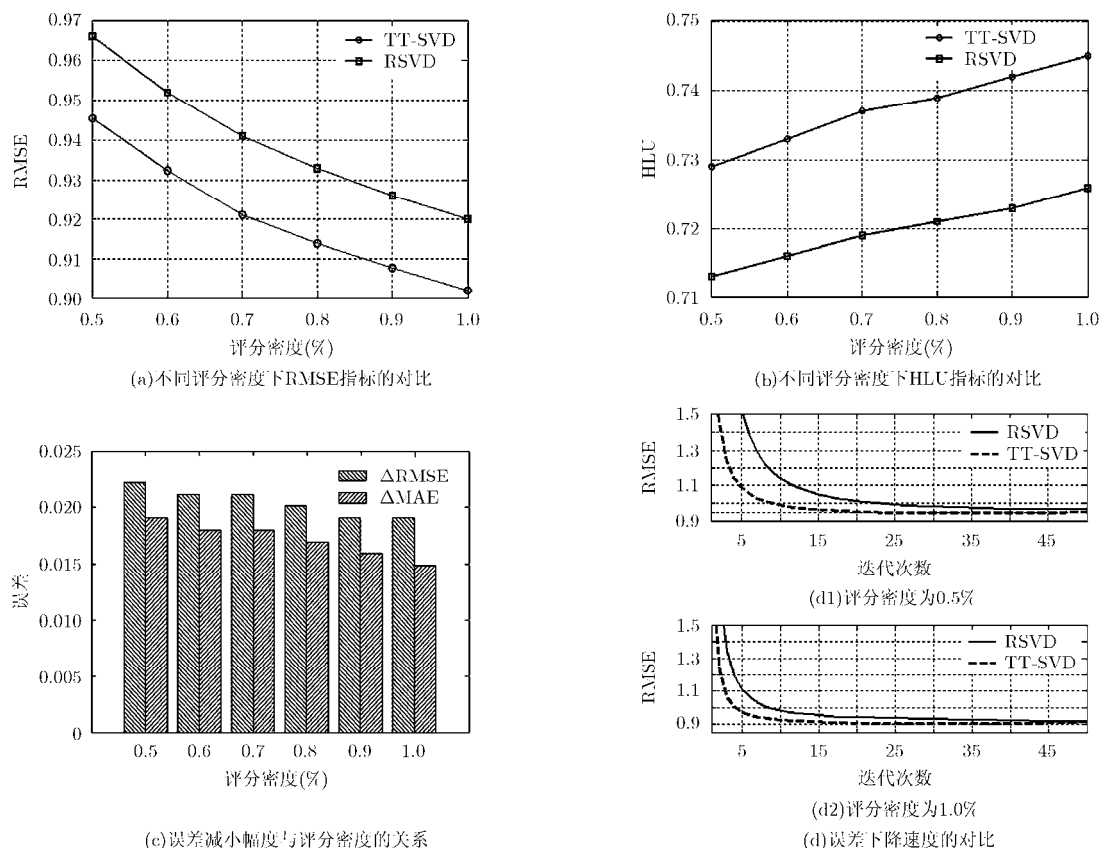
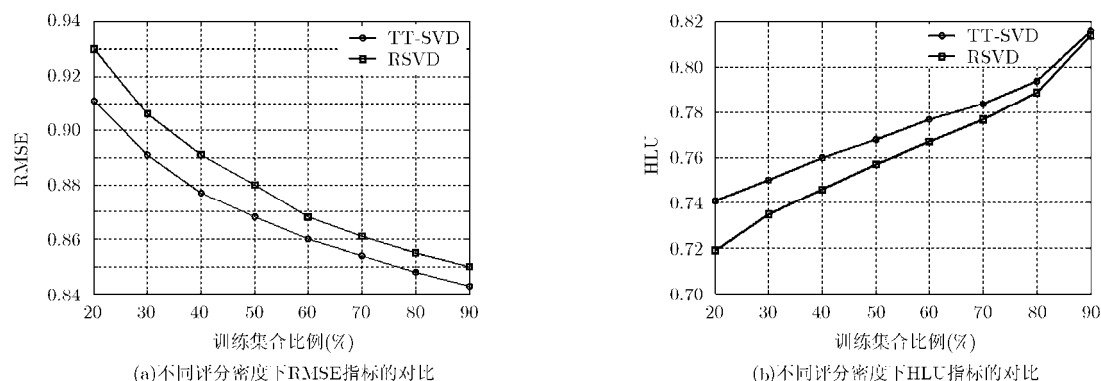


图 2 评分稀疏情况下 TT-SVD 与 RSVD 的性能对比

图 2(c) 中 $\Delta MAE = MAE_{RSVD} - MAE_{TT-SVD}$, $\Delta RMSE = RMSE_{RSVD} - RMSE_{TT-SVD}$, 表示误差下降幅度, 可以看出评分矩阵密度越小, TT-SVD 的性能优势越明显; 图 2(d) 对比了 TT-SVD 和 RSVD 算法的训练时间, TT-SVD 迭代 25 次时的预测误差比 RSVD 迭代 50 次的预测误差还要低, 节省了约 50% 的训练时间, 显著提高了计算效率。

B 组评分稠密情况下的实验结果如图 3 所示, 与 A 组实验结果基本一致, TT-SVD 总是能用更少的训练时间取得更好的预测结果。



5 结束语

本文提出了一种基于标签迁移学习的改进 RSVD 推荐算法, 在 MovieLens 数据集上的仿真结果表明, TT-SVD 算法能够有效缓解数据稀疏性问题, 显著提高预测准确性, 同时减少 50% 左右的训练时间, 具备较高的实际应用价值。TT-SVD 算法依然面临过拟合的问题, 下一步的研究将会探索如何把 TT-SVD 算法扩展到贝叶斯模型, 以消除过拟合问题。

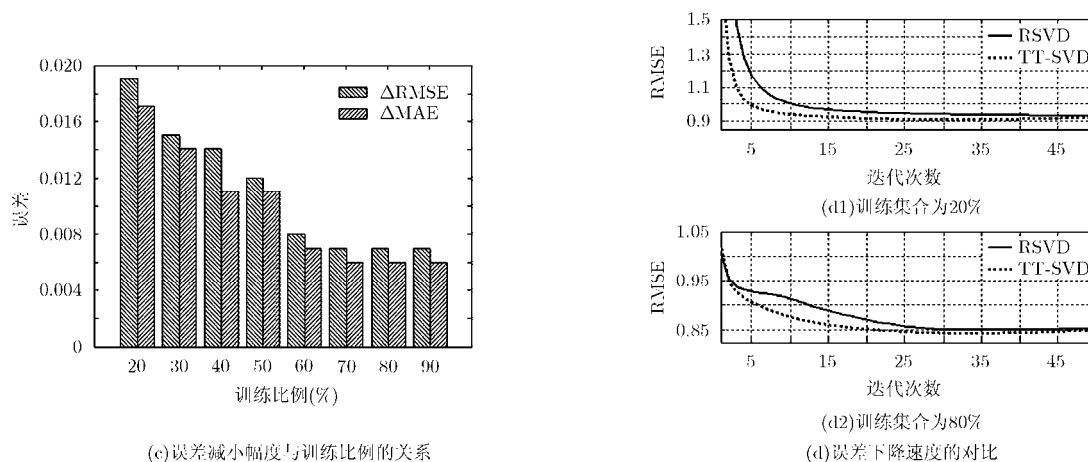


图3 评分稠密情况下 TT-SVD 与 RSVD 的性能对比

参考文献

- [1] Ricci F, Rokach L, Shapira B, *et al.* Recommender Systems Handbook: A Complete Guide for Scientists and Practioners [M]. New York: Springer, 2011: 145-186.
 - [2] Lü Lin-yuan, Medo M, Yeung C H, *et al.* Recommender systems[J]. *Physics Reports*, 2012, 519(1): 1-49.
 - [3] Pan W. Transfer learning in collaborative filtering [D]. [Ph.D. dissertation], Hong Kong University of Science and Technology, 2012.
 - [4] CACHED F, Carneiro V, Fernandez D, *et al.* Comparison of collaborative filtering algorithms: limitations of current techniques and proposals for scalable, high-performance recommender systems[J]. *ACM Transactions on the Web (TWEB)*, 2011, 5(1): 2-33.
 - [5] Cichocki A, Zdunek R, Phan A H, *et al.* Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation [M]. Singapore, Wiley, 2009: 7-18.
 - [6] Salakhutdinov R and Mnih A. Bayesian probabilistic matrix factorization using Markov chain Monte Carlo[C]. Proceedings of the 25th International Conference on Machine Learning, New York, 2008: 880-887.
 - [7] Rendle S. Factorization machines with libFM[J]. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2012, DOI:10.1145/12168752.2168771.
 - [8] Lu Z, Zhong E, Zhao L, *et al.* Selective transfer learning for cross domain recommendation[C]. Proceedings of the 2013 SIAM Conference on Data Mining, Austi, 2013: 641-649.
 - [9] Tang J, Wu S, Sun J, *et al.* Cross-domain collaboration recommendation[C]. Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, 2012: 1285-1293.
 - [10] Pan W, Xiang E W, and Yang Q. Transfer learning in collaborative filtering with uncertain ratings[C]. Twenty-Sixth AAAI Conference on Artificial Intelligence, Toronto, 2012: 662-668.
 - [11] Wang W, Chen Z, Liu J, *et al.* User-based collaborative filtering on cross domain by tag transfer learning [C]. Proceedings of the 1st International Workshop on Cross Domain Knowledge Discovery in Web and Social Network Mining, New York, 2012: 10-17.
 - [12] Zhou Xujuan, Xu Yue, Li Yuefeng, *et al.* The state-of-the-art in personalized recommender systems for social networking[J]. *Artificial Intelligence Review*, 2012, 37(2): 119-132.
 - [13] Qi Y and Edu P. Infinite Tucker decomposition: nonparametric Bayesian models for multi-way data analysis [C]. Proceedings of the 29th International Conference on Machine Learning, Edinburgh, 2012: 1023-1030.
 - [14] Karatzoglou A, Amatriain X, Baltrunas L, *et al.* Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering[C]. Proceedings of the fourth ACM conference on Recommender systems, Barcelona, 2010: 79-86.
 - [15] Rettinger A, Wermser H, Huang Y, *et al.* Context-aware tensor decomposition for relation prediction in social networks[J]. *Social Network Analysis and Mining*, 2012: 1-13.
 - [16] Zhang Z K, Zhou T, and Zhang Y C. Tag-aware recommender systems: a state-of-the-art survey[J]. *Journal of Computer Science and Technology*, 2011, 26(5): 767-777.
 - [17] 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): 163-175.
- Zhu Yu-xiao and Lü Lin-yuan. Evaluation metrics for recommender systems[J]. *Journal of University of Electronic Science and Technology of China*, 2012, 41(2): 163-175.

方耀宁：男，1987年生，硕士生，研究方向为社会化网络、推荐系统。

郭云飞：男，1963年生，教授，博士生导师，研究方向为高性能交换技术、网络安全。

丁雪涛：男，1988年生，硕士生，研究方向为社会化网络、推荐系统。

兰巨龙：男，1962年生，教授，博士生导师，研究方向为宽带信息网络、下一代互联网。