

从图 9 和图 10 中可以看到,在两个数据集中,各个算法的推荐效果随着训练集比例的增加而提高.这是由于训练样本的增大一定程度上减小了评分数据稀疏性的影响.此外可以看到,本文提出的 TPMF 以及 PTPMF 算法在各个比例的训练集下都展现出最佳的推荐性能.这是由于这两个模型在已有的信任数据中进一步挖掘了信任用户对待评分用户的影响机制,从而更好地提升了推荐性能.

3.5 个性化权重的影响

为了更全面地比较个性化权重算法的实验效果,本文在 Epinions1, Epinions2 以及 Ciao1, Ciao2 数据集上分别对 PTPMF 模型当 $D=5$ 以及 $D=10$ 这两种情况进行验证.此外,为了直观地显示个性化权重算法的实验效果,不妨将聚类后所得类别数设为 2.具体的 PTPMF 模型个性化权重的取值以及对比较算法的权重取值情况见表 7.比如,第 2 行表示各个模型在 Epinions1 数据集上的参数取值,其中,PTPMF 模型以及对比较模型 TPMF 的特征维度 D 均取为 5.情形 1 即 $\alpha=0.85/0.35$,表示个性化权重算法(PTPMF)将用户聚类后分成 2 类,每一类的个性化权重 α 分别为 0.85 和 0.35(即,其对应的 TPMF 模型最佳固定权值 α 为 0.60,取参数 $\epsilon=0.25$,从而得到个性化权值 0.85 以及 0.35);而对比较算法 1(TPMF)固定权重 α 为 0.85,对比较算法 2(TPMF)固定权重 α 为 0.35.情形 2~情形 4 类似于情形 1,只是将个性化权重算法(PTPMF)分成的两个类的权重差异逐渐减小(情形 1~情形 4 的参数 ϵ 取值依次沿着 0.25, 0.17, 0.10, 0.05 递减).实验分别针对表 7 中的不同情形进行验证,具体结果如图 11~图 14 所示.而表 7 中的最后一列表示使得 TPMF 取得最佳效果的 α 值.

Table 7 Values for PTPMF model

表 7 PTPMF 模型取值

数据集	情形 1	情形 2	情形 3	情形 4	TPMF 取值
Epininos1 ($D=5$)	$\alpha=0.85/0.35$	$\alpha=0.77/0.43$	$\alpha=0.70/0.50$	$\alpha=0.65/0.55$	$\alpha=0.60$
Epininos1 ($D=10$)	$\alpha=0.80/0.40$	$\alpha=0.73/0.47$	$\alpha=0.70/0.50$	$\alpha=0.65/0.55$	$\alpha=0.60$
Epininos2 ($D=5$)	$\alpha=1.00/0.40$	$\alpha=0.93/0.47$	$\alpha=0.85/0.55$	$\alpha=0.77/0.63$	$\alpha=0.70$
Epininos2 ($D=10$)	$\alpha=1.00/0.60$	$\alpha=0.95/0.65$	$\alpha=0.90/0.70$	$\alpha=0.85/0.75$	$\alpha=0.80$
Ciao1 ($D=5$)	$\alpha=1.00/0.88$	$\alpha=0.99/0.89$	$\alpha=0.98/0.90$	$\alpha=0.97/0.91$	$\alpha=0.94$
Ciao1 ($D=10$)	$\alpha=1.00/0.88$	$\alpha=0.98/0.90$	$\alpha=0.97/0.91$	$\alpha=0.96/0.92$	$\alpha=0.94$
Ciao2 ($D=5$)	$\alpha=1.00/0.88$	$\alpha=0.99/0.89$	$\alpha=0.98/0.90$	$\alpha=0.97/0.91$	$\alpha=0.94$
Ciao2 ($D=10$)	$\alpha=1.00/0.90$	$\alpha=0.99/0.91$	$\alpha=0.98/0.92$	$\alpha=0.97/0.93$	$\alpha=0.95$

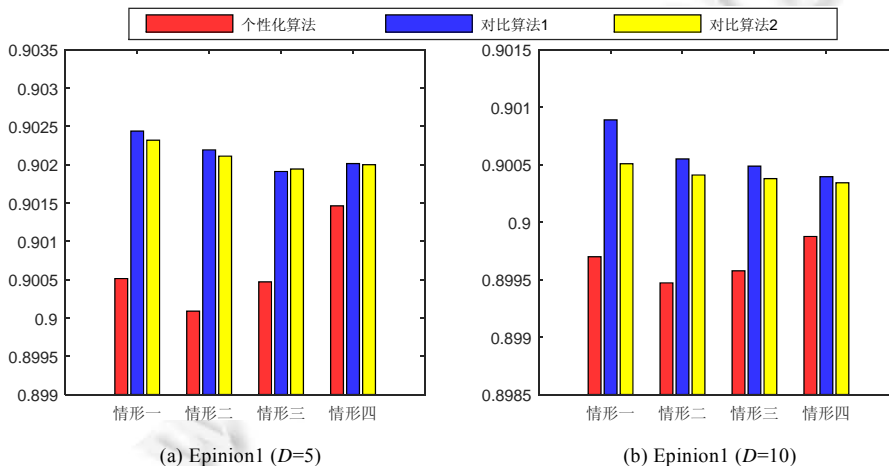


Fig.11 Recommendation results of personalized algorithm in Epinions1

图 11 Epinions1 上个性化算法推荐结果

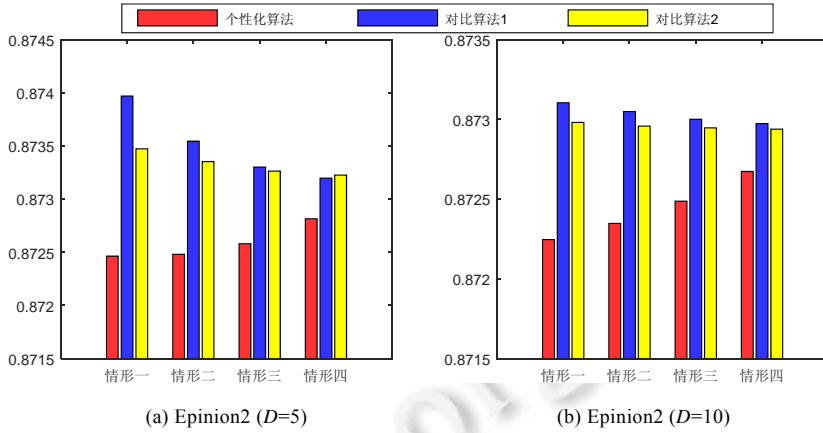


Fig.12 Recommendation results of personalized algorithm in Epinions2

图 12 Epinions2 上个性化算法推荐结果

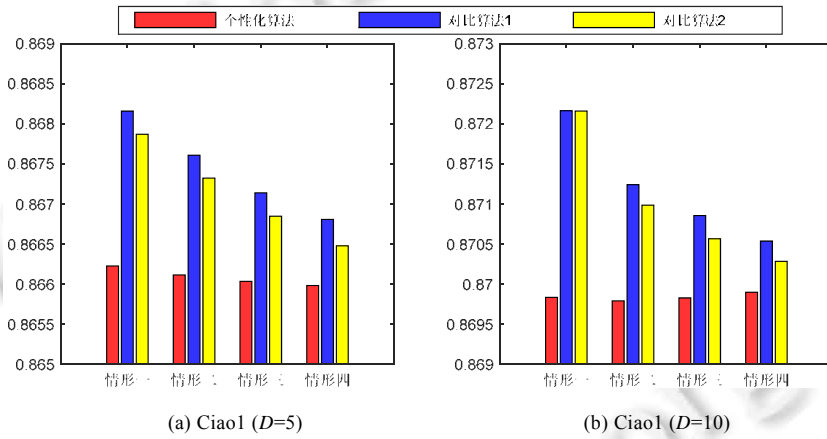


Fig.13 Recommendation results of personalized algorithm in Ciao1

图 13 Ciao1 上个性化算法推荐结果

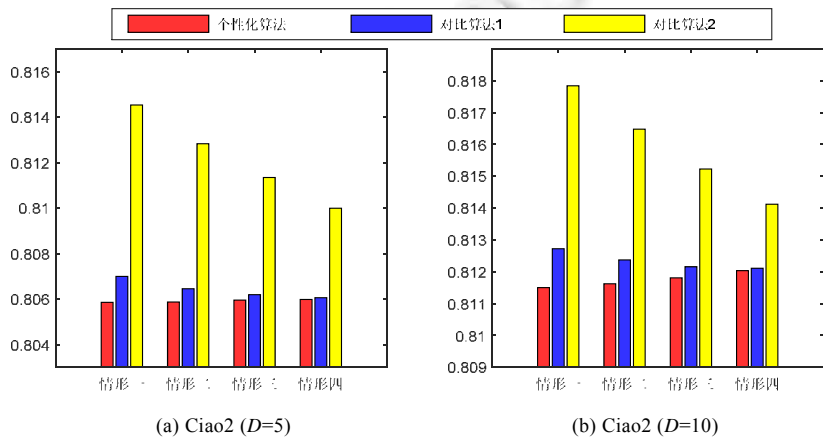


Fig.14 Recommendation results of personalized algorithm in Ciao2

图 14 Ciao2 上个性化算法推荐结果

从图 11~图 14 的实验效果图中可以看到,PTPMF 算法在将用户聚类并对每一类赋予不同权重后,所得 MAE 值不仅均优于固定权重时的情形,同时也优于最佳 TPMF 算法的效果.此外,观察各个数据集中最优 PTPMF 算法的参数 α 时可以发现:对于同一个数据集,在模型维度改变时,最佳参数 α 均比较接近(如 Epinions1 数据集中, $D=5$ 时,最佳 $\alpha=0.77/0.43$; $D=10$ 时,最佳 $\alpha=0.73/0.47$).这是由于同一个数据集中的用户有着确定的最佳 α 值,它们不会随着模型的变化而发生变化;而在不同数据集中,最佳参数 α 之间差异比较明显(如 $D=5$ 时, Epinions1 数据集中最佳 $\alpha=0.77/0.43$,而 Epinions2 以及 Ciao1 数据集中最佳 α 分别为 $1.00/0.40, 0.97/0.91$).这是由于不同数据集中用户偏好往往不同,其最佳 α 分布也不同.

上述结果证明了不同用户对这两种信任机制的权重是不同的,同时也说明了本文提出算法的正确性.

4 相关工作

4.1 传统的推荐算法

传统的推荐算法^[5,6]主要有基于内容的推荐算法^[5]、基于协同过滤的推荐算法^[6]以及混合推荐算法^[24,25].其中,基于内容的推荐算法利用用户历史信息构建用户配置文件,进而结合项目内容信息为用户推荐与其最相似的项目.而基于协同过滤的推荐算法分为基于内存的协同过滤算法和基于模型的协同过滤算法.

- 基于内存的协同过滤算法借助用户对项目的评分信息,寻找与待推荐用户最相似的邻居用户集,并将这些邻居用户对项目的评分作为待推荐用户对项目评分的参考,常见的方法有基于用户的协同过滤算法、基于项目的协同过滤算法等;
- 基于模型的协同过滤算法通过训练集获得模型参数,从而利用该模型为用户进行推荐.常见的方法有矩阵分解模型^[19]、聚类模型^[26]、概率模型^[27]等.其中,Salakhutdinov 等人^[19]提出的 PMF 模型对用户特征、项目特征以及用户评分赋予高斯先验,并利用低维矩阵分解方法获得用户偏好以及项目属性,从而实现了推荐效果的有效性以及推荐算法的高效性.因此,PMF 模型得到了广泛的关注和应用.

然而,传统的推荐算法只考虑了用户对项目的评分信息,忽略了用户间的信任关系对提高推荐性能的巨大作用,许多学者以此为契机,对基于社交信任的推荐算法进行了相应的研究.

4.2 基于社交信任的推荐算法

基于社交信任的推荐算法通过在原有的推荐模型中加入社交网络中的信任信息,使得用户间的兴趣偏好以及相互影响关系得以进一步被挖掘,从而提高了推荐效果^[28].

Ma 等人^[10]将因子分析方法引入到 PMF 模型中,认为用户间的信任关系可以通过用户特征向量与潜在因子特征向量的乘积得到.该作者又将好友信任关系的影响扩展到评分影响上,进而提出了 STE 模型^[2].作者认为信任用户对项目的评分会影响用户对该项目的最终评分,从而综合考虑用户自身评分以及信任用户的评分后给出最终评分.实验证明,该模型优于当时已有的算法.Jamali 等人^[3]指出,文献[2]中提出的 STE 模型所考虑的社交信任关系对待推荐用户的影响仅仅是直接的评分影响,而这种影响并不具备传播性.事实上,信任用户之间的影响可以通过影响用户特征从而间接地影响整个社交网络.作者基于这方面的考虑,提出了 SocialMF 模型.实验结果表明,该模型相比于 STE 模型有着明显的提升.Yang 等人^[14]在文献[3]的基础上,认为用户在不同场景下对好友的信任程度并不相同,在该场景中,更为专业的好友会得到用户更多的信任.基于此,作者提出了多种在不同场景下计算用户间信任关系的方法,并将新的信任权重应用到 SocialMF 模型中.Xu 等人^[29]将社交关系引入推荐用户建模,直接将用户属性表征为好友之间所具有的共同偏好,从而更好地建模不同社交场景下用户所体现出的不同属性.该作者在此基础上,进一步考虑了用户社交关系的主题敏感性与局部稠密性,从而借助快速捕捉社交互动行为中的稠密子结构,提升了算法的性能^[30].

此外,许多工作如在推荐算法中加入时间^[8]等信息,也都是在基于社交信息的推荐算法的基础之上展开的,这在另一方面也印证了研究基于社交信任的推荐算法的重要性.

与上述相关工作不同,本文通过结合信任用户间接影响用户偏好和信任用户直接影响评分结果这两种不

同机制得到 TPMF 模型,并针对不同用户受两种机制影响权重不同的问题,实现了用户模型参数的个性化选择,进而提出了新的推荐框架.

5 结束语

本文基于概率矩阵分解模型,从社交关系影响属性的间接影响和社交关系影响决策的直接影响这两种机制入手,深入研究了用户的信任关系对其评分行为产生的影响,从而提出了基于用户信任关系的概率矩阵分解模型.进而,本文针对上述基本模型对所有用户均使用固定权重融合信任关系的两种不同影响机制这一缺陷,通过分析用户的评分相关性实现用户聚类,并设计个性化权重算法,最终提出了完整的个性化用户评分预测系统框架.通过 Epinions 以及 Ciao 公开数据集上的大量实验表明,本文提出的基于用户信任关系的概率矩阵分解模型能够较好地提升推荐准确率.同时,实验与讨论证实,本文所研究的几个因素(评分影响、特征影响以及个性化权重)均对最终的推荐结果起到了关键作用.

由于用户间的信任关系及用户的兴趣、倾向均随时间发生演变,未来工作将主要围绕如何捕捉、建模这一演变规律并实现准确预测展开.此外,借助同质性原理可以发现,用户间的评分行为(兴趣偏好)会影响用户间的信任关系.因此,借助时间信息同样可以刻画用户兴趣偏好以及用户间的信任关系动态演化的过程,这也是今后工作的核心内容.同时,未来工作还将研究如何引入用户信任关系在不同领域下的不同权重,从而进一步完善本文所提出的基于信任关系的用户评分预测技术.

References:

- [1] Liu Q, Zeng X, Zhu HS, *et al.* Mining indecisiveness in customer behaviors. In: Proc. of the 2015 IEEE Int'l Conf. on Data Mining (ICDM). IEEE, 2015. 281–290.
- [2] Ma H, King I, Lyu MR. Learning to recommend with social trust ensemble. In: Proc. of the 32nd Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. New York: ACM Press, 2009. 203–210.
- [3] Jamali M, Ester M. A transitivity aware matrix factorization model for recommendation in social networks. In: Proc. of the 22th Int'l Joint Conf. on Artificial Intelligence. Burlington: Morgan Kaufmann Publishers, 2011. 2644–2649.
- [4] Liu Q, Ge Y, Li Z, *et al.* Personalized travel package recommendation. In: Proc. of the 2011 IEEE 11th Int'l Conf. on Data Mining (ICDM). IEEE, 2011. 407–416.
- [5] Lops P, De Gemmis M, Semeraro G. Content-Based Recommender Systems: State of the Art and Trends. Springer-Verlag, 2011. 73–105.
- [6] Schafer J, Frankowski D, Herlocker J, *et al.* Collaborative filtering recommender systems. The Adaptive Web, 2007. 291–324.
- [7] Ricci F, Rokach L, Shapira B. Introduction to Recommender Systems Handbook. Springer-Verlag, 2011.
- [8] Guo L, Ma J, Chen ZM. Trust strength aware social recommendation method. Journal of Computer Research and Development, 2013,50(9):1805–1813 (in Chinese with English abstract).
- [9] Li YS, Song MN, Hai-Hong E, *et al.* Social recommendation algorithm fusing user interest social network. Journal of China Universities of Posts & Telecommunications, 2014,21(14):26–33.
- [10] Ma H, Yang H, Lyu MR, *et al.* Sorec: Social recommendation using probabilistic matrix factorization. In: Proc. of the 17th ACM Conf. on Information and Knowledge Management. New York: ACM Press, 2008. 931–940.
- [11] Tang J, Gao H, Hu X, *et al.* Exploiting homophily effect for trust prediction. In: Proc. of the ACM Int'l Conf. on Web Search and Data Mining. 2013. 53–62.
- [12] Xu T, Zhong H, Zhu HS, *et al.* Exploring the impact of dynamic mutual influence on social event participation. In: Proc. of the 2015 SIAM Int'l Conf. on Data Mining. Philadelphia: SIAM, 2015. 262–270.
- [13] Xu T, Zhu HS, Zhao X, *et al.* Taxi driving behavior analysis in latent vehicle-to-vehicle networks: A social influence perspective. In: Proc. of the 22nd ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. ACM Press, 2016. 1285–1294.
- [14] Yang XW, Steck H, Liu Y. Circle-Based recommendation in online social networks. In: Proc. of the 18th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2012. 1267–1275.
- [15] Tang J, Hu X, Gao H, *et al.* Exploiting local and global social context for recommendation. 2013.
- [16] Massa P. A survey of trust use and modeling in real online systems. Trust E-Services: Technologies, Practices and Challenges. Idea Group Inc, 2007. 51–83.
- [17] Leskovec J, Huttenlocher D, Kleinberg J. Predicting positive and negative links in online social networks. In: Proc. of the Int'l Conf. on World Wide Web. ACM Press, 2010. 641–650.
- [18] Page L. The PageRank Citation Ranking: Bringing Order to the Web. Stanford InfoLab, 1998. 1–14.

- [19] Mnih A, Salakhutdinov RR. Probabilistic matrix factorization. In: Proc. of the Advances in Neural Information Processing Systems. 2008. 1257–1264.
- [20] Oestreicher-Singer G, Sundararajan A. Recommendation networks and the long tail of electronic commerce. 2010.
- [21] Grover A, Leskovec J. node2vec: Scalable feature learning for networks. In: Proc. of the ACM SIGKDD Int'l Conf. 2016. 855–864.
- [22] Beigi G, Tang J, Wang S, *et al.* Exploiting emotional information for trust/distrust prediction. In: Proc. of the Siam Int'l Conf. on Data Mining. 2016.
- [23] Jiang M, Cui P, Liu R, *et al.* Social contextual recommendation. In: Proc. of the 21st Int'l Conf. on Information and Knowledge Management. New York: ACM Press, 2012. 45–54.
- [24] Moreno A, Ariza-Porras C, Lago P, *et al.* Hybrid Model Rating Prediction with Linked Open Data for Recommender Systems Semantic Web Evaluation Challenge. Springer Int'l Publishing, 2014. 193–198.
- [25] Son LH. HU-FCF: A hybrid user-based fuzzy collaborative filtering method in recommender systems. Expert Systems with Applications: An Int'l Journal, 2014,41(15):6861–6870.
- [26] Gao M, Cao F, Huang JZ. A cross cluster-based collaborative filtering method for recommendation. In: Proc. of the 2013 IEEE Int'l Conf. on Information and Automation. Piscataway: IEEE, 2013. 447–452.
- [27] Getoor L, Sahami M. Using probabilistic relational models for collaborative filtering. In: Proc. of the Workshop on Web Usage Analysis and User Profiling (WEBKDD'99). 1999.
- [28] Meng XW, Liu SD, Zhang YJ, *et al.* Research on social recommender systems. Ruan Jian Xue Bao/Journal of Software, 2015, 26(6):1356–1372 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4831.htm> [doi: 10.13328/j.cnki.jos.004831]
- [29] Xu T, Liu D, Chen EH, *et al.* Towards annotating media contents through social diffusion analysis. In: Proc. of the 12th Int'l Conf. on Data Mining (ICDM 2012). Brussels, 2012. 1158–1163.
- [30] Xu T, Zhu HS, Chen EH, *et al.* Learning to annotate via social interaction analytics. Knowledge and Information Systems, 2014, 41(2):251–276.

附中文参考文献:

- [8] 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.
- [28] 孟祥武, 刘树栋, 张玉洁, 等. 社会化推荐系统研究. 软件学报, 2015, 26(6):1356–1372. <http://www.jos.org.cn/1000-9825/4831.htm> [doi: 10.13328/j.cnki.jos.004831]



杜东舫(1992—),男,浙江宁波人,硕士,主要研究领域为推荐系统,社交网络.



管楚(1989—),男,博士,主要研究领域为用户数据分析,机器学习.



徐童(1988—),男,博士,副研究员,CCF 专业会员,主要研究领域为数据挖掘.



刘淇(1986—),男,博士,副教授,CCF 专业会员,主要研究领域为数据挖掘与知识发现,机器学习方法与应用.



鲁亚男(1996—),女,硕士,主要研究领域为计算机视觉,数据挖掘.



陈恩红(1968—),男,博士,教授,博士生导师,CCF 会士,主要研究领域为数据挖掘,机器学习.