

DOI: 10.7652/xjtub201304017

结合评分和信任的协同推荐算法

秦继伟^{1,2,3}, 郑庆华^{1,2}, 郑德立^{1,2}, 田锋^{1,2}

(1. 陕西省天地网技术重点实验室, 710049, 西安; 2. 西安交通大学计算机科学与技术系, 710049, 西安;

3. 新疆大学现代教育技术中心, 830046, 乌鲁木齐)

摘要: 针对现有基于信任的推荐系统虽能缓解冷启动和虚假评价但较难获取用户之间的信任关系,难以建立用户彼此之间的偏好关系的问题,提出了基于评分-信任协同的推荐算法并给出了相关数学表达式和实现流程。该算法充分利用推荐系统中的共同评分,协同用户间的信任关系,有策略地选择用户评分的相似度和用户间信任值,建立用户之间的偏好关系,进而实现推荐。随着共同评分数目下限值的增加带来推荐准确度提高的同时将造成覆盖率的下降,因而关键是选取合适的下限值。实验结果表明,这种混合推荐的方法相比传统协作推荐方法与信任推荐方法,在精度损失极小的情况下,较大地提升了覆盖率。评分覆盖率指标分别提高了3%和32.1%,用户覆盖率指标分别提高了8.2%和15.1%,从而获得了精度与覆盖率的良好平衡。

关键词: 协同推荐;信任值;偏好关系;推荐系统

中图分类号: TP391 **文献标志码:** A **文章编号:** 0253-987X(2013)04-0100-05

A Collaborative Recommendation Algorithm Based on Ratings and Trust

QIN Jiwei^{1,2,3}, ZHENG Qinghua^{1,2}, ZHENG Deli^{1,2}, TIAN Feng^{1,2}

(1. Shaanxi Province Key Laboratory of Satellite and Terrestrial Network Tech. R&D, Xi'an 710049, China; 2. Department of Computer Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China; 3. Center for Educational Technology, Xinjiang University, Urumqi 830046, China)

Abstract: Trust is used for recommendation which can solve the cold start and the cheating rates problem in conventional recommender system, but it is difficult to build the trust network and the preference relations among users. A collaborative recommendation algorithm is proposed based on ratings and trust, and the correlation expressions and the flow of algorithm are also presented. The similarity weight is calculated by the rating value and the trust value. The algorithm couples the ratings with the trust to establish the similarity weight, and the predicted ratings produce the candidate set for the target users. Experimental results and comparisons with the traditional collaborative recommendation and the trust recommendation show that the proposed algorithm greatly improves the coverage with a tiny loss in accuracy. The rating coverage is much higher than that of the traditional collaborative recommendation and the trust recommendation by 3% and 32.1%, respectively, and much higher than the traditional collaborative recommendation and the trust recommendation on the user coverage by 8.2% and 15.1%, respectively. And a perfect balance between the accuracy and the coverage is obtained.

Keywords: collaborative recommendation; trust value; preference relations; recommender system

收稿日期: 2012-07-10。 作者简介: 秦继伟(1978—),女,工程师;郑庆华(通信作者),男,教授,博士生导师。 基金项目: 国家杰出青年基金资助项目(60825202);国家自然科学基金资助项目(61070072,61103160);“十一五”国家科技支撑计划重点资助项目(2009BAH51B02)。

网络出版时间: 2013-02-19

网络出版地址: <http://www.cnki.net/kcms/detail/61.1069.T.20130219.1010.007.html>

推荐系统作为一种信息过滤的手段,通过挖掘、分析用户与资源之间的关系,帮助用户准确、高效地从海量数据中发现其可能感兴趣的资源,使其尽可能满足用户的需求,有效地解决了信息过载的问题,被广泛地应用于各大门户网站和电子商务领域。然而,在传统推荐中存在数据稀疏而造成冷启动问题,同时在电子商务中一些商家为了各自的利益,做出虚假评价等,易形成欺骗问题^[1]。随着社会网络的出现,信任作为人际关系的核心概念被引入到推荐系统。它作为用户对推荐者推荐资源的满意程度的一种期望,取代传统推荐中的用户之间偏好的相似权重,在某种程度上解决了冷启动问题和由虚假评分引起的欺骗问题。

基于信任的推荐系统利用了社会化网络的基本理论,即人们更倾向于相信熟人或朋友的推荐,目前基于信任的推荐系统已得到了广泛应用。Epinions^[2]是比较早的将信任应用到商品推荐的系统,此系统由用户直接标识出是否信任其他用户,由此建立用户之间的信任关系。Cosme^[3]基于信任推荐的电子商务系统,用户可以标识与其他用户的信任关系,同时也可以发布对某种商品的评分。随着社会化网络不断深入,研究者们将目光关注到推荐系统中信任关系的研究,即如何构建用户之间的信任网络^[4-5]。文献[6]验证了基于信任的推荐系统符合小世界理论,将信任网络的深度和信任路径相结合,综合考虑用户之间的相似权重,改进了基于信任的推荐。同时文献[7]表明其具有小世界网络的无尺度特性,通过推荐网络验证了相比陌生人而言,用户更信任“密切关系的人”。除此以外,信任作为一种推荐因素被研究者应用到博客推荐^[8]、电影推荐^[9]等。文献[10]指出,基于信任的朋友或熟人的推荐比传统的基于评分的推荐,更能符合用户的需要,并指出综合评价会优于仅基于信任的评价,因为在推荐过程中综合评价可能发现一些珍惜的资源。综上所述,基于信任的推荐更能满足用户的需要,因此成为推荐系统研究领域中的热点,并被应用到许多领域。

本文在已有研究的基础上,从用户评价的时态特征出发,根据推荐系统中用户对资源项目的评分及用于与用户间的信任值,建立一种评分-信任的协同推荐算法,并给出此算法的相关数学表达式及实现的流程,为用户尽可能推荐符合用户需求的资源项目。在信任推荐常用数据集 Epinions 上的实验表明,该算法相比传统协同过滤推荐系统算法,推荐

的质量得到明显的改善,准确率和覆盖率都有所提高,与基于信任推荐算法相比,覆盖率有所提高,即为用户发现一些珍惜资源提供了可能的机会。

1 相关定义

推荐系统(Recommender System)是利用从其他用户体验中学习获得的知识库,帮助目标用户(Active User)在一组候选集中挑选出最符合目标用户最需要的资源项目(Item)的智能系统^[11]。传统推荐系统是一个用户-项目的二维效用模型,形式化定义如下: $u:U \times I \rightarrow R$,这里 R 表示一组候选集序列。

基于信任推荐系统(Trust-aware Recommender System)是将信任融入推荐过程,结合信任网络向目标用户推荐项目资源的推荐系统。因此,扩展推荐系统的二维效用模型,基于信任的推荐系统是一个用户-信任网络-项目的三维效用模型, $v:U \times T \times I \rightarrow R$,则每个用户 $u \in u_{\text{trustnet}}, u_{\text{trustnet}} \subseteq U$,在用户的信任网络中通过其信任用户的经验,为用户 u 尽可能找到对用户效用最大的资源 $i', i' \in I$,形式化定义为

$$\forall u \in u_{\text{trustnet}}, u_{\text{trustnet}} \subseteq U' \quad i'_u = \underset{i \in I}{\operatorname{argmax}} v(u, t_u, i)$$

$u_{\text{trustnet}} = t_u$

本文基于评分-信任协同的推荐(RTCR)是将传统的协同过滤算法中用户对资源项目的评分与信任推荐中的用户之间的信任值相结合,以此作为目标用户选择邻居的依据,帮助目标用户在候选资源项目集合中挑选出最符合目标用户需要的资源项目。鉴于此,评分-信任协同混合推荐是一个用户-评分加信任-项目的三维效用模型, $v:U \times R_T \times I \rightarrow R$,则每个用户 $u \in U$,根据用户对项目资源的偏好经验和用户间的信任关系,为用户 u 尽可能找到对用户效用最大的资源项目 $i', i' \in I$,此时形式化定义为

$$\forall u \in U, \quad i'_u = \underset{i \in I}{\operatorname{argmax}} v(u, r_i, i)$$

显然,在推荐系统中用户之间的协作和交互需要一定的评价机制,包括常用的基于评分的评价机制和才兴起的基于信任的评价机制。通常基于评分的评价采用静态的评价机制,由用户给出的评分计算用户之间的偏好相似关系,并为目标用户计算预测评分,以此决定推荐资源项目。基于信任的评价通过用户的历史行为、社会关系及用户自身的属性特征,建立用户之间的信任程度,并以信任程度为依据,为目标用户计算预测评分,决定推荐资源项目。

由于推荐环境的开放性、动态性、共享性,使得推荐系统中用户之间的相似偏好建立不可能仅仅依靠评分这种静态方式,必须充分考虑到用户间的可信程度,协同推荐资源项目。

2 结合评分和信任的协同推荐

将信任作为推荐决策的重要因素,并结合传统的协同过滤方法,本文基于传统协同过滤方法和信任网络,提出了基于评分-信任的协同推荐(RTCR)算法。

在此推荐系统中,历史交互数据包含用户对项目的评分数据和用户间信任网络的数据。这里,用户-项目评分数据集可以用一个 $n \times m$ 矩阵表示,其中行表示用户,列表示项目,元素 (i, j) 表示用户 i 对项目 j 的评分;用户间信任网络数据用一个 $n \times n$ 矩阵表示,其中元素 (u, v) 是用户 u 对用户 v 的信任值,表示用户 u 对用户 v 的信任程度。

然后,基于共同评分项目的数目设置阈值,有策略地选择基于用户评分的相似度与信任值相结合来计算用户之间的推荐权重,建立用户之间的偏好关系,生成邻居集合;通过邻居集合对未评价过项目的评分进行预测,最后产生推荐集合。其基本处理流程如图1所示。

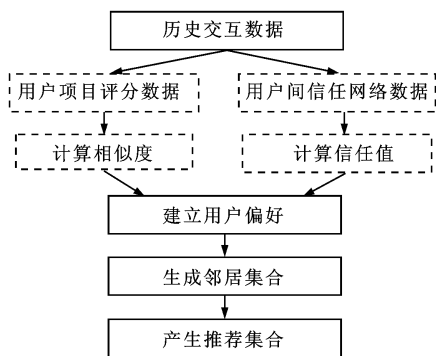


图1 基于评分-信任的协同推荐算法的基本流程

2.1 建立用户偏好

相似度作为项目推荐的主要考虑因素,被用来衡量实体之间的偏好关系,从而作为推荐权重进行推荐预测。本文所提 RTCR 算法,借鉴用户普遍的认知心理,将信任作为实体决策选择时的一个主观概念,结合客观存在的评分,共同评价用户之间的偏好关系。本文基于用户间共同评价项目的数目,设置上下阈值,有策略地选择评分相似度和信任值来计算推荐权重,建立用户之间的偏好关系。其中,评分相似度的计算采用 Pearson 相关因子的方法

$$s(u_a, u_b) = \frac{\sum_{i \in I_{a,b}} (R_{a,i} - \bar{R}_a)(R_{b,i} - \bar{R}_b)}{\left(\sum_{i \in I_{a,b}} (R_{a,i} - \bar{R}_a)^2 \right)^{1/2} \left(\sum_{i \in I_{a,b}} (R_{b,i} - \bar{R}_b)^2 \right)^{1/2}} \quad (1)$$

式中: $s(u_a, u_b)$ 为基于评分的用户 u_a 和用户 u_b 间相似度; $R_{a,i}$ 是用户 u_a 对项目 i 的评分; $I_{a,b}$ 是用户 u_a 和用户 u_b 共同评价过的项目集合; $\bar{R}_a$ 是用户 u_a 所评价过项目的评分平均值。

假设 n 为用户 u_a 和用户 u_b 共同评分的项目数目, n_1 是共同评价项目数目的下限阈值, n_2 是共同评价项目数目的上限阈值, $t_{a,b}$ 为用户 u_a 和用户 u_b 间的信任值, $w_{a,b}$ 为推荐权重,则

$$w_{a,b} = \begin{cases} 0, n < n_1 \text{ 且 } t_{a,b} = 0 \\ t_{a,b}, n < n_1 \text{ 且 } t_{a,b} \neq 0 \\ \frac{n}{n_2} s(u_a, u_b), n_1 \leq n \leq n_2 \text{ 且 } t_{a,b} = 0 \\ \frac{n}{n_2} s(u_a, u_b) + \left(1 - \frac{n}{n_2}\right) t_{a,b}, n_1 \leq n \leq n_2 \\ \text{且 } t_{a,b} \neq 0 \\ s(u_a, u_b), n > n_2 \end{cases} \quad (2)$$

2.2 生成邻居集合

邻居是与当前目标用户兴趣、偏好相近的用户集合。这里,我们采用 Best-n-neighbors 技术选取邻居,综合考虑相似度和信任值以此为推荐权重,选取推荐数量最高的 N 个用户作为当前目标用户的邻居集合。

2.3 产生推荐集

目标用户邻居集合选定后,将当前目标用户邻居对项目的评分与用户之间的偏好关系相结合,预测当前目标用户对待评价项目的评分值,以 Top-N 为原则,形成目标用户的推荐集合。假设 u_a 为目标用户, N_a 为邻居集合,则 u_a 对项目 j 的预测评分为

$$P_{a,j} = \bar{R}_a + \frac{\sum_{b \in N_a} w_{a,b} (R_{b,j} - \bar{R}_b)}{\sum_{b \in N_a} |w_{a,b}|} \quad (3)$$

3 实验结果与分析

通过实验来验证本文提出的 RTCR 算法。通过对公共数据集 Epinions 的分析,由用户之间的信任标签建立信任关系网络,并从中获取用户之间的信任值,同时基于用户对项目的评分获取用户之间的相似度。在此基础上,运用 java 语言设计实现了

本文所提算法,并传统协同过滤推荐(CF)算法和信任推荐(TR)算法进行比较,通过分析实验结果,验证了本文提出的 RTCR 算法的有效性。

3.1 实验数据集

实验数据集 Epinions 来自 TrustLet 组织发布的数据集。该数据集由 Paolo Massa 和 Paolo Avesani^[12]提供,爬取自“Epinions.com”的网站。此数据集有 49 290 个用户和 139 738 个商品,每个用户至少给出过一个评分(1~5 分),有 487 181 个信任评价,信任值为[1],评分矩阵稀疏度为 99.991 35%。

3.2 评价指标

平均绝对偏差(Mean Absolute Error,MAE)是最常用的一种推荐质量度量方法,因此本文采用平均绝对偏差 M_{AE} 作为推荐性能评价指标。平均绝对偏差通过计算预测的用户评分与实际的用户评分之间的偏差来度量预测的准确性,该值越小,推荐的质量越高。 M_{AE} 的定义为:设预测用户评分集合为 $\{p_1, p_2, \dots, p_n\}$,对应的实际用户评分集合为 $\{r_1, r_2, \dots, r_n\}$,则

$$M_{AE} = \sum_{1 \leq i \leq n} |p_i - r_i| / N \tag{4}$$

由于 MAE 指标对评分数据不平衡非常敏感,且没有很好地反映对冷启动用户的推荐效果^[13],因此引入平均绝对用户误差(Mean Absolute User Error,MAUE)来考察系统对冷启动用户的推荐效果^[12],定义为 M_{AUE} :设所有用户的集合为 U ,用户数量为 M ,则

$$M_{AUE} = \sum_{u_i \in U} A_{u_i} / M \tag{5}$$

式中: $A_{u_i} = \sum_{1 \leq i \leq n} |p_i - r_i| / N$ 。算法对用户 u_i 的预测评分集合表示为 $\{p_1, p_2, \dots, p_n\}$,对应的用户 u_i 的实际评分集合为 $\{r_1, r_2, \dots, r_n\}$ 。

另外,由于实验所用的评分矩阵非常稀疏,可预测评分占有所有评分的百分比,即评分覆盖率(Rating Coverage)就成为一个推荐可用性的重要指标^[13],定义为 C_r :设所有用户的总评分的数量为 S ,其中算法可以进行预测的评分数量为 P ,则

$$C_r = P / S \tag{6}$$

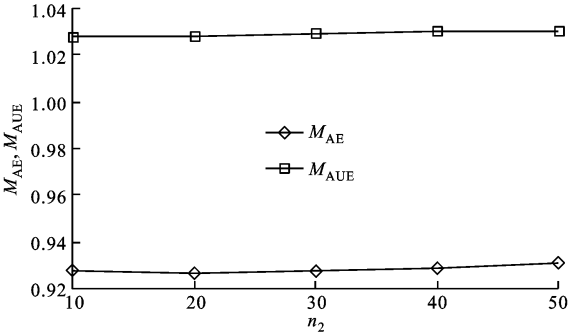
同样地,评分覆盖率没有很好地反映冷启动用户的覆盖情况,因此引入用户覆盖率(User Coverage)来考察系统对冷启动用户的推荐覆盖率^[12],定义为 C_u :设用户总数为 N ,算法可以为其预测至少一个评分的用户数量为 M ,则

$$C_u = M / N \tag{7}$$

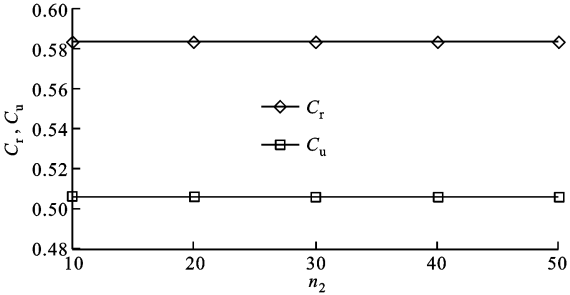
3.3 实验结果与分析

实验 1 本实验是为了确定推荐权重计算表达式中用户共同评价项目的数目的上、下限阈值。分别通过上、下限阈值的不同取值,计算推荐权重,进行预测推荐。通过实验结果,分析评价指标,讨论选定合适的上、下限阈值。

首先,假定下限阈值不变($n_1=0$),上限阈值 n_2 从 10 开始取值,逐步增加,依次为 10、20、30、40、50,从图 2 所示实验结果可以看出,随着 n_2 的增加,MAE 会发生变化,先降低再逐步增加,当共同评分数目的上限阈值达到 20,即 $n_2=20$ 时,MAE 降到最低,而推荐的覆盖率则无变化。利用式(1)计算用户评分相似度 $s(u_a, u_b)$ 时,如果两个用户的共同评分数目大于或等于 2,一定存在 $s(u_a, u_b)$ 。因此,在推荐权重 $w_{a,b}$ 的计算表达式中,当 $n_2 \geq 2$ 时,对 $n_1 \leq n \leq n_2$ 或 $n > n_2$,无论在何种条件下, $s(u_a, u_b)$ 均存在。可见, n_2 的变化不会引起 $w_{a,b}$ 是否存在的变化,进而不会引起覆盖率的变化。由此得出结论,在此数据集上存在一个共同评分数目的上限值($N=20$),使得推荐的误差最低,并且上限的取值与覆盖率无关。



(a)精度

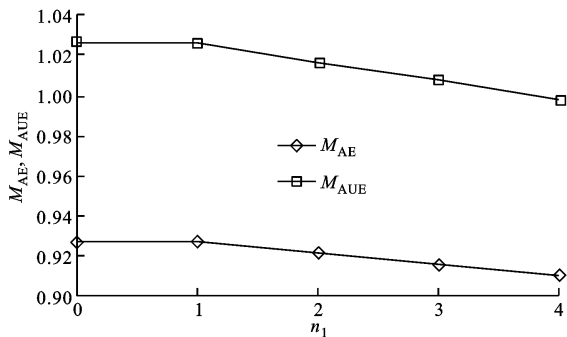


(b)覆盖率

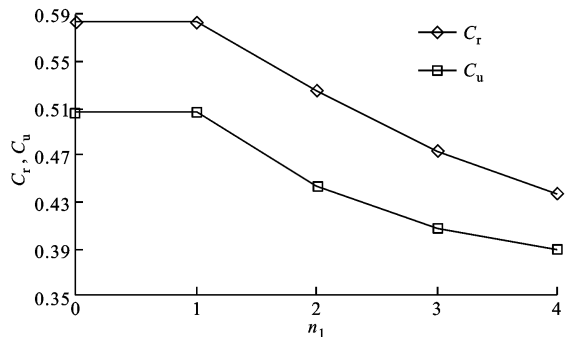
图 2 不同上限阈值 n_2 的实验结果

接着,假定上限阈值不变化($n_2=20$),下限阈值 n_1 从 0 开始取值,逐步增加,依次为 0、1、2、3、4,从

图 3 所示实验结果可以看出,随着 n_1 的增加,推荐的准确性逐步增加,而覆盖率则逐渐降低。对推荐权重表达式(2)分析可知,在信任值数目不变的情况下, n_1 设置得越大,计算 $s(u_a, u_b)$ 需要的共同评分数目就越多,可靠性就越高;同时, n_1 越大,满足共同评分数目大于 n_1 的用户数目将减少,因此覆盖率逐步降低。在 Epinions 数据集上, $n_1 = 1$,覆盖率取得最大值。由此得出结论,随着共同评分数目下限阈值 n_1 的增加,推荐的准确性提高,同时造成覆盖率降低, n_1 的取值在 RTCR 算法中至关重要。



(a) 精度



(b) 覆盖率

图 3 不同下限阈值 n_1 的实验结果

实验 2 本实验对 RTCR 算法与 CF 算法、TR 算法推荐的准确率和覆盖率进行比较,以验证本文所提 RTCR 算法的性能。

实验结果如图 4 所示,在覆盖率方面,RTCR 算法的 Rating Coverage 指标优于 CF 和 TR 算法,分别提高了 3% 和 32.1%。RTCR 算法的 User Coverage 指标也优于 CF 和 TR 算法,分别提高了 8.2% 和 15.1%。RTCR 算法在精度损失极小的情况下,取得了较大的覆盖率提升,从而获得了精度与覆盖率的良好平衡。

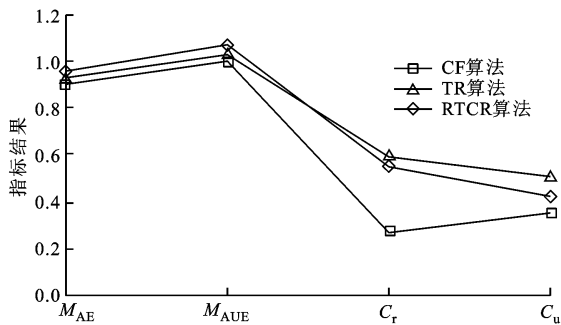


图 4 3 种(CF、TR、RTCR)算法性能的比较

4 结 论

传统协同过滤算法在处理稀疏评分矩阵时,难以保证足够的预测精度,而基于信任推荐算法又没有考虑用户评分之间的相似关系,信息利用不充分,造成覆盖率的损失。针对上述问题,本文提出了基于评分-信任协同推荐算法,将用户共同评分相似度和信任值相结合,提高了覆盖率,实现了精度与覆盖率的良好平衡。

在推荐系统的研究中,由于所涉及的领域较多,不应只考虑某单一的领域,而应该联合其他推荐领域中用户的行为和数据,结合社会化网络相关理论,建立用户的偏好模型,对目标用户做出更准确、可靠的项目推荐。这也是我们下一步将继续研究的工作。

参考文献:

- [1] LEE D H. Pittcult: trust-based cultural event recommender [C]// Proceedings of the 2008 ACM Conference on Recommender Systems. New York, USA: ACM, 2008: 311-314.
- [2] MASSA P, AVESANI P. Trust metrics in recommender systems [M]// Computing with Social Trust. Berlin, Germany: Springer, 2009: 259-285.
- [3] MATSUO Y, YAMAMOTO H. Community gravity: measuring bidirectional effects by trust and rating on online social networks [C]// Proceedings of the 18th International Conference on World Wide Web. New York, USA: ACM, 2009: 751-760.
- [4] VICTOR P, CORNELIS C, COCK M D, et al. Key figure impact in trust-enhanced recommender systems [J]. AI Communications, 2008, 21(2/3): 127-143.
- [5] MA Nan, LIM E P, NGUYEN V A, et al. Trust relationship prediction using online product review data

(下转第 124 页)