评分偏差对于推荐质量的影响

胡必云 李舟军 王 君

(北京航空航天大学 软件开发环境国家重点实验室, 北京 100191) (北京航空航天大学 计算机学院, 北京 100191)

巢文涵

(北京航空航天大学 北京市网络技术重点实验室, 北京 100191)

摘 要: 从理论上分析了评分偏差对于推荐质量的影响;基于潜在偏好及已知评分对评分偏差进行度量,其中潜在偏好通过心理测量学模型计算得出;通过设定不同的评分偏差水平,对评分偏差的影响进行了实验验证.理论分析及实验验证表明:评分偏差可导致推荐准确度及覆盖度下降;基于高质量的评分数据,协同过滤算法可为用户作出好的推荐.

关 键 词: 人工智能; 信号过滤与预测; 信息检索; 评分偏差; 数据质量; 协同过滤; 推荐准确度; 覆盖度

中图分类号: TP 182

文献标识码: A **文 章 编 号**: 1001-5965(2012)06-0823-06

Effect of rating residual on recommendation quality

Hu Biyun Li Zhoujun

(State Key Laboratory of Software Development Environment, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Wang Jun

(School of Computer Science and Technology, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Chao Wenhan

(Key Laboratory of Network Technology of Beijing, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: The effect of the rating residual on recommendation quality was analyzed. The rating residual was measured through user ratings and latent preferences. Latent preferences were computed with psychometric models. With different levels of rating residual, the effect of the rating residual was experimentally evaluated on real world datasets. Theoretical analysis and experimental results show that rating residual has negative effects on recommendation accuracy and coverage. Based on high quality of data, collaborative filtering algorithms can make precise recommendations for users.

Key words: artificial intelligence; signal filtering and prediction; information retrieval; rating residual; data quality; collaborative filtering; recommendation accuracy; coverage

个性化推荐是解决信息过载问题的行之有效的方法之一,并已广泛应用于大型的信息服务提供商,它旨在根据用户的历史偏好信息,向用户主动推荐他/她可能感兴趣的项目,如新闻、书籍等.为了提高推荐质量,研究者提出了许多推荐算法,这些算法可分为3类[1]:基于内容的推荐、协同过滤(CF, Collaborative Filtering)及混合的推荐算法.在这些算法中,CF 算法得到了广泛的研究与应用.CF 算法可进一步分为基于用户、基于项目及基于模型的算法,它们通常基于用户评分为用

收稿日期: 2011-03-18; **网络出版时间**: 2012-06-15 15:43

网络出版地址: www.cnki.net/kcms/detail/11.2625.V.20120615.1543.024.html

基金项目: 国家自然科学基金资助项目(61170189,60973105);软件开发环境国家重点实验室自主研究课题资助项目(SKLSDE-2011ZX-03)

作者简介: 胡必云(1982 -),女,安徽六安人,博士生,hubiyun@cse.buaa.edu.cn.

户进行推荐,其中广泛使用的数据集包括 MovieLens,EachMovie,Netflix,Jester 等. 一般认为,评分数据稀疏会导致 CF 算法的推荐质量下降[2],因此,研究者提出了很多方法以解决该问题. 然而,对于评分数据的另一重要特征——数据质量问题的研究却较少,这可能是由于数据质量相对于数据稀疏更加难以衡量造成的.

在少数涉及到数据质量问题的研究中,文献[3]将推荐算法不能准确预测的评分看成是噪声数据并将它们删除,这使得推荐准确度有了一定的提高,但却造成了推荐覆盖度(coverage)下降(推荐覆盖度指算法可给出的评分预测数占待预测评分总数的比率),后者是容易理解的,因为数据删除加剧了数据稀疏问题;文献[4]将符合随机噪声模式(CNP,Causal Noise Pattern)的用户评分系列删除,使得推荐准确度有所提高,文中未报道评分删除对推荐覆盖度的影响;文献[5]尝试了使用专家评分对普通用户进行推荐,因为专家的评分质量更高. 以上文献虽然对数据质量问题进行了一定研究,并取得了一些初步的研究成果,但存在以下问题:①对数据质量问题的定义过于模糊,即将某些评分[3]或用户评分系列[4]整体地看作是噪声. 然而,用户的评分可能只是在某种程度上偏离了其真实偏好,即一个评分数据可能含有噪声而不全部是噪声;②对数据质量对于 CF 算法推荐质量的影响机制,特别是对于推荐覆盖度的影响,没有做明确的分析与验证;③对于噪声数据的处理方式太极端,即将噪声数据整体删除,这虽然可使推荐准确度有一定的提高,但却导致推荐覆盖度下降[3].

针对上述问题,本文首先从心理学角度对用户评分过程进行分析,并提出评分偏差的概念,以对数据质量问题进行明确的定义. 同时,分析评分偏差对于 CF 算法每一步骤的影响,以初步研究数据质量问题对于推荐质量的影响. 其次,基于心理测量学模型对评分偏差加以度量. 再次,通过设定不同的评分偏差水平,进一步在现实世界数据集上验证评分偏差对于 CF 算法推荐质量的影响. 本文最后对实验结果进行了详细分析并对未来工作进行了说明.

1 评分偏差对于 CF 的影响

1.1 评分偏差

CF 算法常依据用户评分为用户作出推荐,而从心理学角度分析,用户根据不同项目满足他/她的潜在兴趣的程度对不同项目进行评分,因此观测到的评分只是用户潜在兴趣的一种外在表现. 在评分过程中,用户可能会受到各种因素的干扰,导致用户的评分偏离于他/她的真实偏好,即产生了评分偏差. 造成评分偏差的因素可以是错误地使用了评分标准,如对于一个 5 级评分量表来说,一个用户偏向于使用 1 ~ 3 的评分(可称该用户为一个较严的评分人),那么该用户的评分 2 代表其真实偏好为 3(2 和 3 分别为错误的评分量表 1 ~ 3 和正确的评分量表 1 ~ 5 的中间值),即产生了评分偏差 -1(观测值 2 - 真实值 3). 其它可以造成评分偏差的因素包括:对某些项目不在意而给出的随意评分、输入错误等. 更为具体的例子如 Jester 数据集,它使用了[-10,10]的连续评分,用户通过在评分条上点击以给出评分,这种评分方式则更倾向于产生评分偏差[6].

1.2 协同过滤(CF)

下文所提及的 CF 算法主要指基于用户的 CF 算法(user-based CF)[7]. CF 算法主要包含 3 个步骤:相似度计算、邻居选取及预测. CF 算法将用户对于项目的评分看成用户评分向量,然后基于两个用户共同评分的项目(co-rated items)计算用户之间的相似度. 常用的相似度计算方法有皮尔森(Pearson)相关系数、余弦相似度等[2]. 在邻居选取阶段,CF 算法通常选择和当前用户(需要为其作出推荐的用户)最相似的若干个用户作为他/她的邻居;然后根据邻居对于目标项目(需要为其作出评分预测的项目)的评分情况预测当前用户对于目标项目的评分;常用的预测公式如式(1)所示:

$$p_{a,i} = \bar{r}_a + \frac{\sum_{n=1}^{k} s(a,u_n)(r_{u_n,i} - \bar{r}_{u_n})}{\sum_{n=1}^{k} |s(a,u_n)|} \quad (1)$$

其中,$p_{a,i}$是算法给出的当前用户 a 对于目标项目 i 的评分预测值;$\bar{r}_a$为 a 的平均评分;k 为邻居个数;$s(a,u_n)$为 a 和其邻居 u_n 之间的相似度;$r_{u_n,i}$为 u_n 对于 i 的评分. 值得注意的是,在邻居选取阶段选出的邻居并非全部是有效的.

定义 1 有效邻居. 在预测阶段实际使用到的邻居. 即对于 CF 算法,有效用户邻居对于目标项目的评分 $r_{u_n,i}$(见式(1))是存在的.

1.3 评分偏差对于 CF 的影响

1.3.1 评分偏差对于用户相似度计算的影响

直观地,评分偏差可导致两个具有低相似度值的用户变得更为相似,两个具有高相似度值的用户变得较不相似,即评分偏差可导致两个用户

u_i 和 u_j 之间的相似度值 $s(u_i,u_j)$ 升高或降低.

例 1　如表 1 所示,对于 1 ~ 5 的 5 级评分,其中用户 u_1 是一个理想的评分人,她的评分 r_{u_1} 代表了她的真实偏好 t_{u_1},而用户 u_2 是一个较严的评分人,他的观测评分 r_{u_2} 集中于评分等级 1 ~ 3. 如果 u_2 正确地使用了评分等级 1 ~ 5,则他的真实偏好很可能为表中第 4 列所列的评分 t_{u_2}. 根据真实偏好计算得到的 u_1 和 u_2 之间的 Pearson 相关系数是 0.4,而根据观测评分计算得到的 u_1 和 u_2 之间的相关系数为 0.2,即因为评分偏差的影响,u_1 和 u_2 之间的相似程度 $s(u_1,u_2)$ 降低;同样地,用户 u_3 的真实偏好 t_{u_3} 列于表 1 中最后一列,但 u_3 在评分过程中,总有值为 1 或 -1 的评分偏差,即观测到的 u_3 的评分为表中第 5 列数据 r_{u_3}. 当使用真实偏好时,u_1 和 u_3 之间的 Pearson 相关系数为 -0.1,但是当使用观测评分时,两者的相关系数为 0.1,即因为评分偏差的影响,用户 u_1 和 u_3 之间的相似程度 $s(u_1,u_3)$ 升高.

表 1　用户 u 对于项目 i 的评分 r 及真实偏好 t

项目	用户评分以及真实偏好				
	r_{u_1}/t_{u_1}	r_{u_2}	t_{u_2}	r_{u_3}	t_{u_3}
i_1	1	2	3	1	2
i_2	5	3	5	2	3
i_3	3	1	2	4	5
i_4	4	2	3	2	1
i_5	2	3	4	3	4

1.3.2　*评分偏差对于邻居选取的影响*

为了便于分析,首先给出关于相似用户的假设及其推论.

假设 1　两个用户共同评分的项目越多,则这两个用户倾向于越相似.

假设 1 是符合直觉的,即两个用户之间的共同点越多,则倾向于越相似;同时,本文依据用户评分计算得出的用户平均相似度随共评项目数变化的实验结果也反应了该假设.

基于假设 1,可以得出以下推论:

推论 1　两个用户之间的相似性越小,则这两个用户共同评分的项目倾向于越少.

在邻居选取阶段,评分偏差对于用户相似度计算的影响会导致具有低相似度值的用户被选择为邻居,而根据推论 1,这些具有低相似度的邻居是无效邻居的可能性增大.

例 2　如表 2 所示,由于评分偏差的影响,具有低相似度值的用户 u_2 被选择为用户 u_1 的邻居,CF 算法目前需要为 u_1 对于项目 i_4 的评分进行预测(以? 表示). 因为 u_1 和 u_2 相似程度低,根据推论 1,u_2 对于 i_4 没有进行评分,即 u_2 是无效邻居的可能性增大.

表 2　被用户 u_1 和 u_2 共同评分的项目

用户	项目						
	i_1	i_2	i_3	i_4	i_5	i_6	i_7
u_1		*	*	?		*	
u_2	*	*			*		*

注:* —评分.

1.3.3　*评分偏差对于推荐质量的影响*

具有低相似度值的邻居被使用会导致推荐准确度下降,更进一步地,由于低相似性邻居是有效邻居的可能性小,导致推荐覆盖度下降. 评分偏差对于推荐质量的影响分析可总结为图 1.

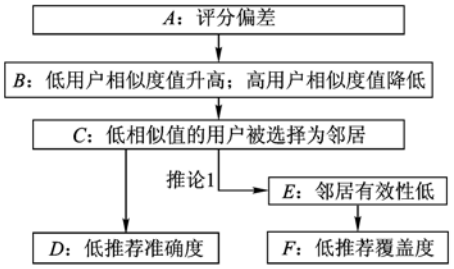


图 1　评分偏差对于推荐质量的影响分析

2　基于心理测量学模型的偏差度量

2.1　心理测量学模型

在心理测量学研究领域,潜在特质模型(latent trait models)或称项目反应理论模型(item response theory models),是一系列旨在通过人对项目如试题、调查问卷等的反应数据,测量人的潜在特质如能力、兴趣、满意度等的模型[8]. 典型的项目反应理论模型有 Rasch 模型,如式(2)所示. Rasch 模型最初用于教育测量学领域,随着模型及其参数估计技术的发展,Rasch 模型被用于越来越多的心理测量领域[9].

$$\ln(p(r_{u,i}=1)/p(r_{u,i}=0)) = \theta_u - b_i \quad (2)$$

在不同的应用领域中,Rasch 模型的参数意义是不同的. 在传统的教育测量领域,式(2)建模了学生 u 答对题目 i 的概率 $p(r_{u,i}=1)$ 与学生能力 θ_u 及题目难度 b_i 之间的关系[8];因为本文主要考虑通过用户对于项目的评分数据测量用户潜在兴趣,所以将 $p(r_{u,i}=1)$ 解释为用户 u 对于项目 i 的反应为积极(如表示感兴趣或评分为 1)的概率;将 θ_u 解释为用户的潜在兴趣(latent interest);将 b_i 解释为项目的可接受程度(agreeability). 因此,式(2)直观上表示了用户的潜在兴趣越强,则越可能表现出积极的反应.

Rasch 模型只能处理二元反应数据(如 0 和 1),为了处理多元反应数据,如 1 ~ 5 级的评分数

据,心理学家对其进行了扩展,典型的扩展后的模型如式(3)或式(4)所示[10],其中 $p(r_{u,i}=x)$ 表示用户 u 对项目 i 评 x 分的概率,x 取自评分等级 $1,2,\cdots,m$;θ_u 和 b_i 的意义同式(2);τ_x 为评分等级 $x-1$ 相对于 x 被选择的可能性;$\kappa_0=0,\kappa_x=-\sum_{k=1}^{x}\tau_k(x=1,2,\cdots,m-1,\kappa_m=0)$ 为评分等级系数(category coefficient).

$$\ln(p(r_{u,i}=x)/p(r_{u,i}=x-1))=\theta_u-b_i-\tau_x \tag{3}$$

$$p(r_{u,i}=x)=\frac{e^{\kappa_x+x(\theta_u-b_i)}}{1+\sum_{k=1}^{m}e^{\kappa_k+k(\theta_u-b_i)}} \tag{4}$$

2.2 心理测量学模型与个性化推荐

文献[11]将心理测量学模型引入个性化推荐算法的研究中,并提出了基于潜在兴趣的 CF 算法,实验结果表明基于潜在兴趣的 CF 算法相比于传统 CF 算法可提高推荐准确度.

上述研究结果表明,将心理测量学引入个性化推荐算法研究中是有益的. 提出具有更高推荐准确度的算法虽然重要,但对于评分数据质量问题的研究同样重要,因为低质量的数据可能会造成研究结果产生偏差甚至误导. 因此,鉴于用户在评分过程中倾向于产生评分偏差,前期工作[12]提出了使用潜在偏好(latent preference)表示用户偏好,其中潜在偏好通过心理测量学模型计算得出,并实验验证了潜在偏好相比于观测评分可更准确地表示用户偏好. 然而,文中并没有对评分偏差对于推荐质量的影响进行实验分析. 本文在前期研究的基础上,提出评分偏差的度量方法,并分析评分偏差对于推荐质量的影响,从而对推荐算法作出更为客观的评价.

2.3 评分偏差度量

具体地,评分偏差的度量分为 3 个步骤:

1) 依据用户评分及心理测量学模型估计用户潜在兴趣 θ_u. 本文用 Rasch 模型参数估计软件 Winsteps[13]. Winsteps 首先将 θ_u,b_i 及 τ_x 初始化为 0,然后遍历用户评分数据集,通过正态逼近算法(PROX, Normal Approximation Estimation Algorithm)迭代更新 θ_u 及 b_i,直至 θ_u 及 b_i 的变化程度小于某一阈值(详细的推导过程可参见文献[14]). Winsteps 然后通过 JML 进行第 2 阶段的参数估计. 首先依据当前 θ_u,b_i,τ_x 值及 Rasch 模型计算用户评分期望值,然后依据评分期望值与评分的差别程度,通过 Newton-Raphson 算法对 θ_u,b_i 及 τ_x 进行迭代更新. 例如,如果依据目前参数计算得到的用户期望评分总分小于实际总分,则提高用户的潜在兴趣值 θ_u. 具体的参数迭代更新公式可参见 Wright 等的推导[15].

2) 对于每一用户评分,依据用户潜在兴趣,计算相应的潜在偏好,如式(5),其中 $l_{u,i}$ 为计算得到的用户 u 对于项目 i 的潜在偏好:

$$l_{u,i}=\sum_{x=1}^{m}x(p(r_{u,i}=x)) \tag{5}$$

3) 依据潜在偏好,计算用户的评分偏差 R. 计算公式见式(6),其中 $r_{u,i}$ 为用户 u 对于项目 i 的评分. 前期工作表明,潜在偏好 $l_{u,i}$ 相比于观测评分 $r_{u,i}$ 可更准确地表示用户偏好[12],因此,由式(6)度量用户评分偏差是可行的.

$$R_{u,i}=r_{u,i}-l_{u,i} \tag{6}$$

3 实验验证

3.1 数据集

实验采用了广泛使用的数据集之一 MovieLens[2]. 该数据集含有 943 个用户对于 1 682 个电影项目的 100 000 个评分. 评分可取值为 1 ~ 5. 实验随机选取了 80% 的用户评分作为训练集,其余的评分作为测试集.

3.2 推荐质量度量标准

实验采用广泛使用的平均绝对误差(MAE, Mean Absolute Error)[2]作为推荐准确度的度量标准,用 E 表示:

$$E=\text{avg}\,|p_{u,i}-r_{u,i}| \tag{7}$$

其中,$p_{u,i}$ 为算法给出的用户 u 对于项目 i 的评分预测值;$r_{u,i}$ 为测试集中的对应评分. MAE 值越小,则推荐准确度越高;推荐覆盖度(coverage)定义为算法可给出的预测值数目占测试评分总数 N 的比例[5].

3.3 实验过程及结果分析

实验首先通过训练集数据使用 Rasch 模型参数估计软件 Winsteps[13]估计用户潜在兴趣;然后针对训练集中的每一用户评分 $r_{u,i}$,计算其对应的潜在偏好 $l_{u,i}$ 及评分偏差 $R_{u,i}=r_{u,i}-l_{u,i}$,接着,通过设定不同的评分偏差水平 a(a 分别取值 0, 0.2, 0.4, …, 1.0)形成不同的训练数据集 $t_{u,i}=aR_{u,i}+l_{u,i}$. 通过由不同评分偏差水平形成的训练集,分别分析评分偏差对于用户相似度计算、邻居选取及推荐质量的影响.

实验采用 Pearson 相关系数[2]计算用户相似度,邻居数 k 分别设置为 5, 10, …, 60,评分预测公式为式(1).

3.3.1　评分偏差对于相似度计算的影响实验分析

实验计算了在不同评分偏差水平 a 下,含有不同共评项目数 $c>x$(x 分别取值为 2,4,…,10,15,25,…,105)的用户之间的平均相似度,结果如图 2 所示(为了清楚的呈现实验结果,本文省略了某些参数取值下的结果,或将不同参数下的结果分别呈现).

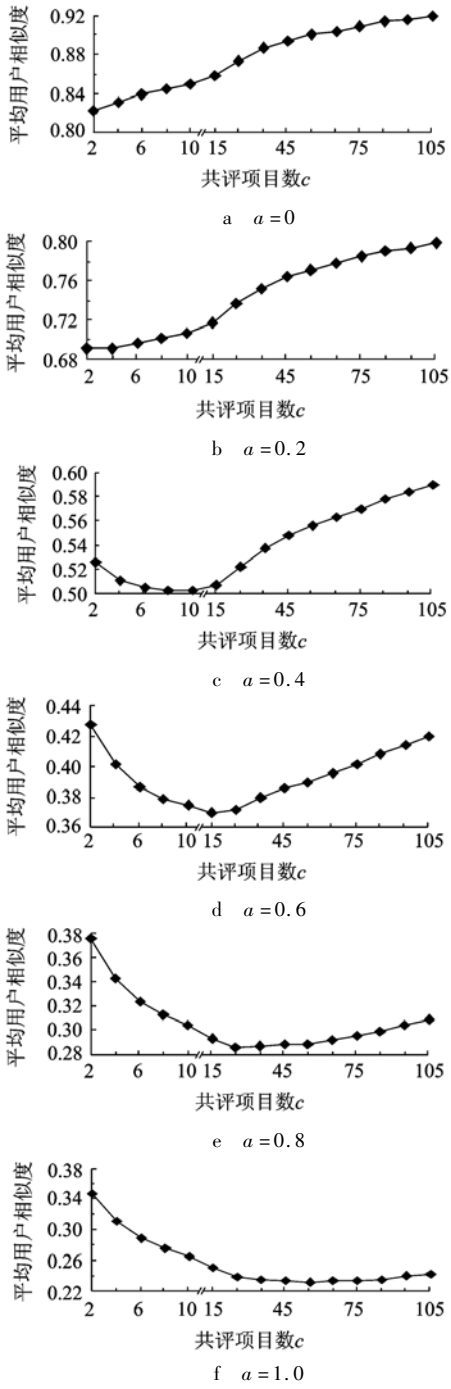


图 2　平均用户相似度随共评项目数 c 的变化情况

图 2 的结果显示,随着偏差水平的增大,具有低相似度的用户之间的相似水平相对提高;具有高相似度的用户之间的相似水平相对降低. 例如,对于 $a=0.2, c<10$ 的这些低相似性用户(图 2b),随着偏差水平的升高,这些用户之间的相似度值逐渐相对提高(相对于同一偏差水平下,$c>10$ 的那些用户);而对于 $a=0.2, c>10$ 的这些高相似性用户,随着偏差水平的升高,这些用户之间的相似度值逐渐相对下降. 这些相似度变化与本文第 1.3.1 节的分析结果一致(图 1 中 $A\rightarrow B$).

3.3.2　评分偏差对于邻居选取的影响实验分析

图 2 的结果显示,随着偏差水平的升高,具有低相似度值的用户在邻居中占的比率增大(图 1 中 $B\rightarrow C$). 例如,对于 $a=0.2, c<10$ 的这些低相似性用户(图 2b),当偏差水平升至 0.8 及 1.0 时,CF 算法完全优先选择这些用户作为邻居,而这些用户之间的共评项目数相对较少,即是有效邻居的可能性小.

为了进一步分析邻居的有效性,定义平均有效邻居比(average effective neighbor ratio),用 e 表示,如式(8),其中 N 为测试集中的评分个数,n_{i_1} 为预测任一测试集中的评分时,有效邻居的个数,n_{i_2} 为可选邻居的个数.

$$e=\frac{\sum_{i=1}^{N}\frac{n_{i_1}}{n_{i_2}}}{N} \tag{8}$$

当邻居数 k 设为不同值时,平均有效邻居比的实验结果报告于图 3. 图 3 显示,随着偏差水平的升高,邻居的有效性降低;而前述分析表明:偏差水平增大,邻居中低相似度值的用户比率增大;据此两点可得出结论:具有低相似度值的用户邻居的有效性低(图 1:$C\rightarrow E$).

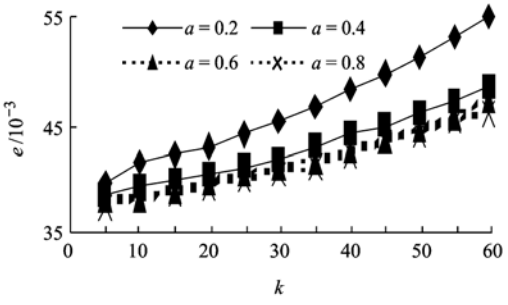


图 3　不同评分偏差水平 a 下的平均邻居有效比

3.3.3　评分偏差对于推荐质量影响实验分析

图 4 显示了不同偏差水平下的推荐准确度结果,可见,平均绝对误差随偏差水平 a 的增加而增加,即随着偏差水平的提高,推荐准确度下降. 最大降幅达 20%(当邻居数 k 为 5 时,将偏差水平平均绝对误差从0.2提高至 0.8,a 值从 0.785 增至 0.945).

不同偏差水平下的推荐覆盖度结果见图 5,可见,推荐覆盖度随偏差水平 a 的增加而下降,即

随着偏差水平的提高,推荐覆盖度下降. 最大降幅达12%(当邻居数 $k=60$ 时,将偏差水平 a 从0.2提高至0.8,推荐覆盖度从0.648降至0.572).

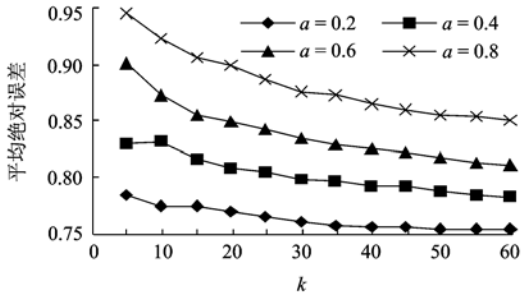


图4 不同评分偏差水平 a 下的推荐准确度

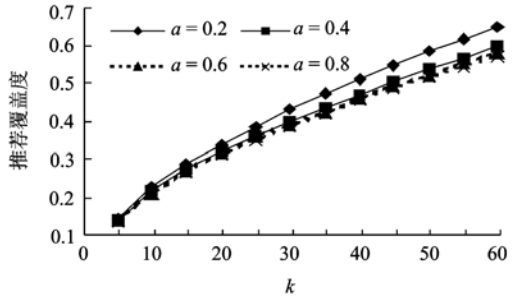


图5 不同评分偏差水平 a 下的推荐覆盖度

推荐准确度主要与用户相似度值有关,即所选邻居是否具有真实可靠的高相似性;推荐覆盖度主要与邻居有效性有关,即针对每一测试评分,是否可找到有效的用户邻居. 因此,在实验分析了评分偏差对于用户相似度计算及邻居选取的影响后,不难得出评分偏差对于推荐准确度及覆盖度的影响,即评分偏差使得相似度值低的用户被选择为邻居,导致推荐准确度下降(图1中 $A\rightarrow B\rightarrow C\rightarrow D$);同时,相似度值低的邻居的有效性低,导致推荐覆盖度下降(图1中 $A\rightarrow B\rightarrow C\rightarrow E\rightarrow F$).

4 结 论

本文的主要贡献在于:①给出了评分数据质量问题的明确定义,即评分偏差;②对评分偏差对于推荐质量(包括推荐准确度及覆盖度)的影响机制进行了分析,并基于心理测量学模型对用户评分偏差进行了度量,通过实验验证了评分偏差的影响:评分偏差导致具有低相似度值的用户被选择为邻居,从而导致推荐准确度下降;同时,由于具有低相似度值的邻居的有效性低,导致推荐覆盖度下降;③结果表明,基于高质量的数据,CF算法可获得好的推荐准确度及覆盖度.

参考文献(References)

[1] Gediminas A, Alexander T. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions[J]. IEEE Trans on Knowledge and Data Engineering (TKDE), 2005, 17(6): 734 – 749

[2] Badrul S, George K, Joseph K, et al. Item-based collaborative filtering recommendation algorithms[C]//Proc of 10th International World Wide Web Conference (WWW'01). New York: ACM Press, 2001: 285 – 295

[3] O'Mahony M P, Hurley N J, Silvestre G C M. Detecting noise in recommender system databases[C]//Proc of the 10th International Conference on Intelligent User Interfaces (IUI '06). New York: ACM Press, 2006: 109 – 115

[4] Cao Huanhuan, Chen Enhong, Yang Jie, et al. Enhancing recommender systems under volatile user interest drifts[C]//Proc of the 18th ACM Conference on Information and Knowledge Management (CIKM'09). New York: ACM Press, 2009: 1257 – 1266

[5] Xavier A, Neal L, Pujol J M, et al. The wisdom of the few: a collaborative filtering approach based on expert opinions from the web[C]//Proc of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'09). New York: ACM Press, 2009: 532 – 539

[6] Herlocker J L, Konstan J A, Terveen L G, et al. Evaluating collaborative filtering recommender systems[J]. Transactions on Information Systems (TOIS), 2004, 22(1): 5 – 53

[7] Wang Jun, de Vries A P, Reinders M J T. Unifying user-based and item-based collaborative filtering approaches by similarity fusion[C]//Proc of the 29th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'06). New York: ACM Press, 2006: 501 – 508

[8] 杜文久. 高等项目反应理论[M]. 重庆: 西南师范大学出版社, 2007: 71 – 88
Du Wenjiu. Advanced item response theory[M]. Chongqing: Southwest Normal University Press, 2007: 71 – 88 (in Chinese)

[9] Cheng Yunghsiang. Exploring passenger anxiety associated with train travel[J]. Transportation, 2010, 37(6): 875 – 896

[10] David Andrich. A rating formulation for ordered response categories[J]. Psychometrikia, 1978, 43(4): 561 – 573

[11] Hu Biyun, Li Zhoujun, Wang Jun. User's latent interest-based collaborative filtering[C]//Proc 32nd European Conference on Information Retrieval (ECIR'10). Berlin: Springer-Verlag, 2010: 619 – 622

[12] Hu Biyun, Li Zhoujun, Chao Wenhan, et al. User preference representation based on psychometric models[C]//Proc 22nd Australia Database Conference (ADC'11). Sydney: ACS, 2011: 57 – 64

[13] Linacre Mike. WINSTEPS Rasch measurement computer program[EB/OL]. Chicago: Winsteps.com, 2007[2011-05-15]. http://www.winsteps.com

[14] Linacre Mike. PROX for polytomous data[J]. Rasch Measurement Transactions, 1995, 8(4): 400

[15] Wright B D, Masters G N. Rating scale analysis[M]. Chicago: MESA Press, 1982: 100

(编 辑:文丽芳)