

基于项目评分预测的协同过滤推荐算法*

邓爱林^{1,2+}, 朱扬勇¹, 施伯乐¹

¹(复旦大学 计算机与信息技术系,上海 200433)

²(上海电信技术研究院,上海 200122)

A Collaborative Filtering Recommendation Algorithm Based on Item Rating Prediction

DENG Ai-Lin^{1,2+}, ZHU Yang-Yong¹, SHI Bai-Le¹

¹(Department of Computer and Information Technology, Fudan University, Shanghai 200433, China)

²(Shanghai Telecom Technology Research Institute, Shanghai 200122, China)

+Corresponding author: Phn: 86-21-58755256, E-mail: dal5758@sina.com; dal5758@etang.com

<http://dmgroup.org.cn>

Received 2002-08-08; Accepted 2002-09-30

Deng AL, Zhu YY, Shi BL. A collaborative filtering recommendation algorithm based on item rating prediction. *Journal of Software*, 2003,14(9):1621~1628.

<http://www.jos.org.cn/1000-9825/14/1621.htm>

Abstract: Recommendation system is one of the most important technologies in E-commerce. With the development of E-commerce, the magnitudes of users and commodities grow rapidly, resulted in the extreme sparsity of user rating data. Traditional similarity measure methods work poor in this situation, make the quality of recommendation system decreased dramatically. To address this issue a novel collaborative filtering algorithm based on item rating prediction is proposed. This method predicts item ratings that users have not rated by the similarity of items, then uses a new similarity measure to find the target users' neighbors. The experimental results show that this method can efficiently improve the extreme sparsity of user rating data, and provide better recommendation results than traditional collaborative filtering algorithms.

Key words: E-commerce; recommendation system; collaborative filtering; item similarity; recommendation algorithm; MAE (mean absolute error)

摘要: 推荐系统是电子商务系统中最重要的技术之一。随着电子商务系统用户数目和商品数目的日益增加,在整个商品空间上用户评分数据极端稀疏,传统的相似性度量方法均存在各自的弊端,导致推荐系统的推荐质量急剧下降。针对用户评分数据极端稀疏情况下传统相似性度量方法的不足,提出了一种基于项目评分预测的协同过滤推荐算法,根据项目之间的相似性初步预测用户对未评分项目的评分,在此基础上,采用一种新颖的相似性度量方法计算目标用户的最近邻居。实验结果表明,该算法可以有效地解决用户评分数据极端稀疏情况下

* Supported by the National High Technology Development 863 Program of China under Grant Nos.863-317-01-04-99, 863-306-ZT06-07-02 (国家高科技发展计划(863))

第一作者简介: 邓爱林(1975—),男,湖北黄冈人,博士,工程师,主要研究领域为数据库,知识库。

传统相似性度量方法存在的问题,显著地提高推荐系统的推荐质量.

关键词: 电子商务;推荐系统;协同过滤;项目相似性;推荐算法;平均绝对偏差

中图法分类号: TP391 文献标识码: A

随着互联网的普及和电子商务的发展,推荐系统逐渐成为电子商务 IT 技术的一个重要研究内容,越来越多地得到研究者的关注.目前,几乎所有大型的电子商务系统,如 Amazon,CDNOW,eBay,当当网上书店等,都不同程度地使用了各种形式的推荐系统.

最近邻协同过滤推荐是当前最成功的推荐技术^[1],其基本思想就是基于评分相似的最近邻居的评分数据向目标用户产生推荐.由于最近邻居对项目(电子商务中的商品、电影、音乐等)的评分与目标用户非常相似,因此目标用户对未评分项目的评分可以通过最近邻居对该项目评分的加权平均值逼近.

随着电子商务系统规模的进一步扩大,用户数目和项目数据急剧增加,导致用户评分数据的极端稀疏性^[1],在用户评分数据极端稀疏的情况下,传统相似性度量方法均存在各自的弊端,使得计算得到的目标用户的最近邻居不准确,推荐系统的推荐质量急剧下降.针对用户评分数据的极端稀疏性,本文提出了一种基于项目评分预测的最近邻协同过滤推荐算法,通过计算项目之间的相似性,初步预测用户对未评分项目的评分,然后采用一种新颖的相似性度量方法计算用户的最近邻居.实验结果表明,本方法可以有效地解决用户评分数据极端稀疏情况下传统相似性度量方法存在的不足,显著地提高推荐系统的推荐质量.

1 相关工作

为了产生精确的推荐,保证推荐系统的实时性要求,研究者提出了各种不同的推荐算法,如协同过滤推荐系统、基于项目的协同过滤推荐、Bayesian 网络技术、聚类技术、关联规则技术以及基于图的 Horting 图技术等.

Typestry^[2]是最早提出来的基于协同过滤的推荐系统,目标用户需要明确指出与自己行为比较类似的其他用户.GroupLens^[3]是基于用户评分的自动化协同过滤推荐系统,用于推荐电影和新闻.Ringo 推荐系统^[4]和 Video 推荐系统^[5]通过电子邮件的方式分别推荐音乐和电影.Breese^[1]等人对各种协同过滤推荐算法及其改进进行了深入分析.

传统的协同过滤推荐通过用户的最近邻居产生最终的推荐,基于项目的协同过滤推荐首先计算项目之间的相关性,然后通过用户对相关项目的评分预测用户对未评分项目的评分^[6].

Bayesian 网络技术利用训练集创建相应的模型^[7],模型用决策树表示,节点和边表示用户信息.训练得到的模型非常小,所以对模型的应用非常快.这种方法适合于用户的兴趣爱好变化比较慢的场合.

聚类技术将具有相似兴趣爱好的用户分配到相同的簇中^[8,9],聚类产生之后,根据簇中其他用户对商品的评价预测目标用户对该商品的评价.由于聚类过程离线进行,所以在线的推荐算法产生推荐的速度比较快.

关联规则技术在零售业得到了广泛的应用,关联规则挖掘可以发现不同商品在销售过程中的相关性.基于关联规则的推荐算法根据生成的关联规则模型和用户当前的购买行为向用户产生推荐^[10].关联规则模型的生成可以离线进行,因此可以保证有效地推荐系统的实时性要求.

Horting 图技术是一种基于图的方法^[11],节点代表用户,边代表两个用户之间的相似度.在图中搜索近邻节点,然后综合近邻节点的评分形成最后的推荐.Horting 图技术可以跳过中间节点寻找最近邻居,考虑了节点之间的传递相似关系.因此,推荐精度优于最近邻协同过滤技术.

针对数据的极端稀疏性问题,文献[12]提出通过奇异值分解(SVD)减少项目空间的维数,使得用户在降维后的项目空间上对每一个项目均有评分,实验结果表明,这种方法可以有效地解决同义词(synonymy)问题,显著地提高推荐系统的伸缩能力.但降维会导致信息损失,降维效果与数据集密切相关,在项目空间维数很高的情况下,降维的效果难以保证^[13].

2 传统的相似性度量方法及其分析

协同过滤推荐根据其他用户的观点产生对目标用户的推荐列表,它基于这样一个假设^[1]:如果用户对一些项目的评分比较相似,则他们对其他项目的评分也比较相似.协同过滤推荐系统使用统计技术搜索目标用户的若干最近邻居,然后根据最近邻居对项目的评分预测目标用户对项目的评分,产生对应的推荐列表.

为了找到目标用户的最近邻居,必须度量用户之间的相似性,然后选择相似性最高的若干用户作为目标用户的最近邻居.目标用户的最近邻居查询是否准确,直接关系到整个推荐系统的推荐质量.准确查询目标用户的最近邻居是整个协同过滤推荐成功的关键.

用户评分数据可以用一个 $m \times n$ 阶矩阵 $A(m,n)$ 表示, m 行代表 m 个用户, n 列代表 n 个项目,第 i 行第 j 列的元素 R_{ij} 代表用户 i 对项目 j 的评分.用户评分数据矩阵如图 1 所示.

度量用户 i 和用户 j 之间相似性的方法是,首先得到经用户 i 和用户 j 评分的所有项目,然后通过不同的相似性度量方法计算用户 i 和用户 j 之间的相似性,记为 $sim(i,j)$.

	Item ₁	...	Item _k	...	Item _n
User ₁	$R_{1,1}$...	$R_{1,k}$...	/
...
User _j	$R_{j,1}$...	/	...	$R_{j,n}$
...
User _m	/	...	$R_{m,k}$...	$R_{m,n}$

Fig.1 User rating data matrix

图 1 用户评分数据矩阵

2.1 传统的相似性度量方法

度量用户间相似性的方法有多种,主要包括如下 3 种方法^[6]:余弦相似性、相关相似性以及修正的余弦相似性.

- 余弦相似性(cosine):用户评分被看做是 n 维项目空间上的向量,如果用户对项目没有进行评分,则将用户对该项目的评分设为 0,用户间的相似性通过向量间的余弦夹角度量.设用户 i 和用户 j 在 n 维项目空间上的评分分别表示为向量 \vec{i}, \vec{j} ,则用户 i 和用户 j 之间的相似性 $sim(i,j)$ 为

$$sim(i, j) = \cos(\vec{i}, \vec{j}) = \frac{\vec{i} \cdot \vec{j}}{\|\vec{i}\| \|\vec{j}\|}$$

分子为两个用户评分向量的内积,分母为两个用户向量模的乘积.

- 相关相似性(correlation):设经用户 i 和用户 j 共同评分的项目集合用 I_{ij} 表示,则用户 i 和用户 j 之间的相似性 $sim(i,j)$ 通过 Pearson 相关系数度量:

$$sim(i, j) = \frac{\sum_{c \in I_{ij}} (R_{i,c} - \bar{R}_i)(R_{j,c} - \bar{R}_j)}{\sqrt{\sum_{c \in I_{ij}} (R_{i,c} - \bar{R}_i)^2} \sqrt{\sum_{c \in I_{ij}} (R_{j,c} - \bar{R}_j)^2}}$$

$R_{i,c}$ 表示用户 i 对项目 c 的评分, \bar{R}_i 和 \bar{R}_j 分别表示用户 i 和用户 j 对项目的平均评分.

- 修正的余弦相似性(adjusted cosine):在余弦相似性度量方法中没有考虑不同用户的评分尺度问题,修正的余弦相似性度量方法通过减去用户对项目的平均评分来改善上述缺陷,设经用户 i 和用户 j 共同评分的项目集合用 I_{ij} 表示, I_i 和 I_j 分别表示经用户 i 和用户 j 评分的项目集合,则用户 i 和用户 j 之间的相似性 $sim(i,j)$ 为

$$sim(i, j) = \frac{\sum_{c \in I_{ij}} (R_{i,c} - \bar{R}_i)(R_{j,c} - \bar{R}_j)}{\sqrt{\sum_{c \in I_i} (R_{i,c} - \bar{R}_i)^2} \sqrt{\sum_{c \in I_j} (R_{j,c} - \bar{R}_j)^2}}$$

$R_{i,c}$ 表示用户 i 对项目 c 的评分, \bar{R}_i 和 \bar{R}_j 分别表示用户 i 和用户 j 对项目的平均评分.

2.2 传统相似性度量方法分析

随着电子商务系统规模的扩大,用户数目和项目数目呈指数级增长,用户评分数据极端稀疏.在大型电子商

务系统中,用户评分的项目一般不会超过项目总数的 1%^[6],经两个用户共同评分的项目则更少.在用户评分数据极端稀疏的情况下,传统的相似性度量方法存在相应的弊端.下面我们详细分析传统的相似性度量法在用户评分数据极端稀疏情况下存在的问题.

在余弦相似性度量方法中,用户没有评分的项目均将评分假设为 0,设用户 i 对项目 j 的评分为 $r_{i,j}$,则在构造用户评分数据矩阵 $A(m,n)$ 时,对任意的项目 j ,用户 i 对项目的评分 $R_{i,j}$ 为

$$R_{i,j} = \begin{cases} r_{i,j}, & \text{if user } i \text{ rated item } j \\ 0, & \text{if user } i \text{ not rated item } j \end{cases}$$

这样做的好处是可以有效地提高计算性能,但在项目数量及其巨大、用户评分数据极端稀疏的情况下,上述假设的可信度并不高,因为用户对所有未评分项目的评分均假设为 0.事实上,用户对未评分商品的喜好程度不可能完全相同,对这些项目的评分也不可能完全相同(全部为 0),由此可见,在用户评分数据极端稀疏的情况下,余弦相似性并不能有效地度量用户之间的相似性,修正的余弦相似性存在同样的问题.

在相关相似性度量方法中,设经用户 A 评分的项目集合用 I_A 表示,则在计算用户 i 和用户 j 之间的相似性时,首先需要计算经用户 i 和用户 j 评分的项目集合的交集 I_{ij} :

$$I_{ij} = I_i \cap I_j.$$

然后在项目集合 I_{ij} 上通过相关相似性度量方法计算用户 i 和用户 j 之间的相似性.常识告诉我们,用户只有在比较多的项目上评分比较相似,我们对用户之间的相似性的确定才比较高.在用户评分数据极端稀疏的情况下,经两个用户共同评分的项目集合 I_{ij} 更小,经常只有一两个项目,即使用户在这样小的项目集合上评分非常相似,根据常识我们也不能肯定它们之间的相似性很高,因此在用户评分数据极端稀疏的情况下,相关相似性度量方法也存在一定的弊端.

综上所述,传统的 3 种相似性度量方法在用户评分数据极端稀疏的情况下并不能有效地度量用户之间的相似性,从而使得计算出来的目标用户的最近邻居不准确,导致整个推荐算法的推荐质量急剧下降.

3 基于项目评分预测的协同过滤推荐算法

由于用户评分数据的极端稀疏性,传统的相似性度量方法不能有效地计算目标用户的最近邻居,协同过滤推荐系统的推荐质量难以保证.为了解决用户评分数据的极端稀疏性,最简单的办法就是将用户对未评分项目的评分设为一个固定的缺省值(一般设为评分域的中间值,如在 5 分制评分中设为 3),或者设为用户的平均评分,实验表明,这种改进方法可以有效地提高协同过滤推荐系统的推荐精度^[1].

显然,用户对未评分项目的评分不可能完全相同,因此上述改进方法并不能从根本上解决用户评分数据极端稀疏情况下传统相似性度量方法存在的问题.本文通过计算项目间的相似性,由用户对相似项目的评分预测用户对未评分项目的评分,使得用户之间共同评分的项目比较多,从而可以有效地解决用户评分数据极端稀疏情况下传统相似性度量方法存在的不足,使得计算得到的目标用户的最近邻居比较准确.

下面我们来详细介绍基于项目评分预测的协同过滤推荐算法.算法主要分为两步:寻找最近邻居和产生推荐.第 3.1 节给出了基于项目评分预测的邻居查找方法,第 3.2 节给出了根据最近邻居产生推荐的计算方法.

3.1 寻找最近邻居

为了有效地解决用户评分数据极端稀疏情况下传统相似性度量方法存在的问题,本文提出在计算用户 i 和用户 j 之间的相似性时,首先计算经用户 i 和用户 j 评分的项目集合的并集 U_{ij} ,设经用户 A 评分的项目集合用 I_A 表示,则 U_{ij} 为

$$U_{ij} = I_i \cup I_j.$$

用户 i 和用户 j 在项目集合 U_{ij} 中未评分的项目通过用户对相似项目的评分预测出来,然后在商品集合 U_{ij} 上计算用户 i 和用户 j 之间的相似性.这种方法不仅能有效地解决相关相似性度量方法中用户共同评分数据比较少,而且可以有效地解决余弦相似性度量方法和修正的余弦相似性度量方法中对所有未评分商品的评分均相同的问题(均为 0),使得计算出来的目标用户的最近邻居比较准确,从而有效地提高推荐算法的推荐

质量.

预测用户 i 对项目集合 U_{ij} 中未评分项目的评分是基于项目相似性的协同过滤推荐算法的关键.设用户 i 在项目空间 U_{ij} 中未评分的项目集合用 N_i 表示,即

$$N_i = U_{ij} - I_i.$$

对任意项目 $p \in N_i$,使用如下方法预测用户 i 对项目 p 的评分 $P_{i,p}$ ^[6]:

- 计算项目 p 与其他项目之间的相似性,与计算用户间相似性类似,首先需要得到对项目 i 和项目 j 评分的所有用户评分,然后通过第 2.1 节中介绍的各种相似性度量方法计算项目 i 和项目 j 之间的相似性.
- 将相似性最高的若干项目作为项目 p 的邻居项目集合,即在整个项目空间中查找项目集合 $M_p = \{I_1, I_2, \dots, I_v\}$,使得 $p \notin M_p$,并且项目 I_1 与项目 p 的相似性 $\text{sim}(p, I_1)$ 最高,项目 I_2 与项目 p 的相似性 $\text{sim}(p, I_2)$ 次之,依此类推.

- 得到 M_p 后,采用文献[6]中提出的方法预测用户 i 对项目 p 的评分 $P_{i,p}$:

$$P_{i,p} = \frac{\sum_{n \in M_p} \text{sim}_{p,n} \times R_{i,n}}{\sum_{n \in M_p} (|\text{sim}_{p,n}|)}.$$

通过上述方法处理后,用户 i 和用户 j 对项目集合 U_{ij} 中的所有项目均有评分,即对任意的项目 $p \in U_{ij}$,用户 i 对项目 p 的评分为

$$R_{i,p} = \begin{cases} r_{i,p}, & \text{if user } i \text{ rated item } p \\ P_{i,p}, & \text{if user } i \text{ not rated item } p \end{cases}$$

然后基于项目集合 U_{ij} 通过第 2.1 节中介绍的 3 种相似性度量方法计算用户 i 和用户 j 之间的相似性,在我们提出的方法中,相关相似性和修正的余弦相似性是一致的,具体为:

- 余弦相似性

$$\text{sim}(i, j) = \frac{\sum_{u \in U_{ij}} (R_{i,u} \times R_{j,u})}{\sqrt{\sum_{u \in U_{ij}} R_{i,u}^2} \sqrt{\sum_{u \in U_{ij}} R_{j,u}^2}}.$$

- 相关相似性(修正的余弦相似性)

$$\text{sim}(i, j) = \frac{\sum_{u \in U_{ij}} (R_{i,u} - \bar{R}_i)(R_{j,u} - \bar{R}_j)}{\sqrt{\sum_{u \in U_{ij}} (R_{i,u} - \bar{R}_i)^2} \sqrt{\sum_{u \in U_{ij}} (R_{j,u} - \bar{R}_j)^2}}.$$

寻找最近邻居的目标就是对每一个用户 u ,在整个用户空间中查找用户集合 $C = \{C_1, C_2, \dots, C_k\}$,使得 $u \notin C$,并且 C_1 与 u 的相似性 $\text{sim}(u, C_1)$ 最高, C_2 与 u 的相似性 $\text{sim}(u, C_2)$ 次之,依此类推.

3.2 产生推荐

通过本文提出的相似性度量方法得到目标用户的最近邻居,下一步需要产生相应的推荐.设用户 u 的最近邻居集合用 NBS_u 表示,则用户 u 对项目 i 的预测评分 $P_{u,i}$ 可以通过用户 u 对最近邻居集合 NBS_u 中项目的评分得到,计算方法如下^[1]:

$$P_{u,i} = \bar{R}_u + \frac{\sum_{n \in NBS_u} \text{sim}(u, n) \times (R_{n,i} - \bar{R}_n)}{\sum_{n \in NBS_u} (|\text{sim}(u, n)|)}.$$

$\text{sim}(u, n)$ 表示用户 u 与用户 n 之间的相似性, $R_{n,i}$ 表示用户 n 对项目 i 的评分, \bar{R}_n 分别表示用户 u 和用户 n 对项目的平均评分.

4 实验结果及其分析

4.1 数据集

我们采用 MovieLens 站点提供的数据集(<http://movielens.umn.edu/>).MovieLens 是一个基于 Web 的研究型

推荐系统,用于接收用户对电影的评分并提供相应的电影推荐列表.目前,该 Web 站点的用户已经超过 43 000 人,用户评分的电影超过 3 500 部.

我们从用户评分数据库中选择 6 000 条评分数据作为实验数据集,实验数据集中共包含 145 个用户和 805 部电影,其中每个用户至少对 20 部电影进行了评分.

整个实验数据集需要进一步划分为训练集和测试集,为此,我们引入变量 x 表示训练集占整个数据集的百分比.例如, $x=0.8$ 表示整个数据集的 80% 作为训练集,20% 作为测试集.在本文的所有实验中,均采用 $x=0.8$ 作为实验基础.

为了度量整个数据集的稀疏性,我们引入稀疏等级的概念,其定义为用户评分数据矩阵中未评分条目所占的百分比.我们选择的电影数据集的稀疏等级为

$$1-6000/(145 \times 805)=0.9486.$$

4.2 度量标准

评价推荐系统推荐质量的度量标准主要包括统计精度度量方法和决策支持精度度量方法两类^[6].统计精度度量方法中的平均绝对偏差 MAE(mean absolute error)易于理解,可以直观地对推荐质量进行度量,是最常用的一种推荐质量度量方法,本文采用平均绝对偏差 MAE 作为度量标准.平均绝对偏差 MAE 通过计算预测的用户评分与实际的用户评分之间的偏差度量预测的准确性,MAE 越小,推荐质量越高.

设预测的用户评分集合表示为 $\{p_1, p_2, \dots, p_N\}$, 对应的实际用户评分集合为 $\{q_1, q_2, \dots, q_N\}$, 则平均绝对偏差 MAE 定义为^[6]

$$MAE = \frac{\sum_{i=1}^N |p_i - q_i|}{N}.$$

4.3 实验结果

首先对各种不同的相似性度量标准进行实验,选择最佳的相似性度量标准;然后在相同的数据集上进行对比实验,比较本文提出的基于项目评分预测的协同过滤推荐算法与经典的协同过滤推荐算法的推荐质量,并对实验结果进行分析.

4.3.1 相似性度量标准比较

在计算项目间的相似性时,由于余弦相似性度量方法比较容易实现,预测速度比较快,而且预测精度也比较高,所以本文选择余弦相似性作为项目间相似性的度量标准.

在基于项目评分预测的协同过滤推荐算法中,修正的余弦相似性度量方法与相关相似性度量方法一致,统称为相关相似性.原始的余弦相似性称为余弦相似性.以这两种相似性度量方法分别对数据集进行实验,计算其 MAE,邻居个数从 4 增加到 20,间隔为 4.实验结果如图 2 所示.由图 2 可知,在各种实验条件下,余弦相似性度量方法的 MAE 均低于相关相似性度量方法,因此,在后续的实验,我们选择余弦相似性作为基于项目评分预测的协同过滤推荐算法中用户间相似性的度量方法.

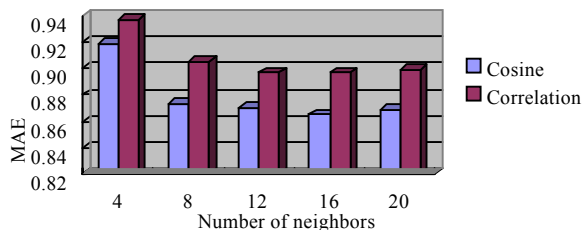


Fig.2 Comparison of similarity measure methods

图 2 相似性度量标准比较

4.3.2 推荐算法推荐质量比较

为了检验本文提出的算法的有效性,我们以传统的协同过滤推荐算法作为对照,在传统的协同过滤推荐算

法中,分别以余弦相似性和相关相似性作为相似性度量标准,计算其 MAE,邻居个数从 4 增加到 20,间隔为 4,然后与本文提出的基于项目评分预测的协同过滤推荐算法作比较,实验结果如图 3 所示。

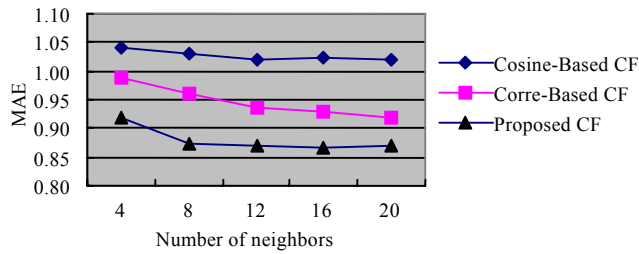


Fig.3 Comparison of accuracy of recommendation algorithms

图 3 推荐算法推荐精度比较

由图 3 可知,在各种实验条件下,本文提出的基于项目评分预测的协同过滤推荐算法均具有最小的 MAE.由此可知,与传统的协同过滤推荐算法相比,我们提出的基于项目评分预测的协同过滤推荐算法可以显著地提高推荐系统的推荐质量.

4.3.3 实验结果分析

本文提出的基于项目评分预测的协同过滤推荐算法与传统的协同过滤推荐算法的最大不同就在于寻找最近邻居的方法不一样.在传统的协同过滤推荐算法中,如果采用相关相似性,两个用户之间的相似性由经用户共同评分的项目决定,在用户评分数据极端稀疏的情况下,经两个用户共同评分的项目非常少,经常只有一两个项目,用户即使在项目上的评分非常相似,也不一定就是最近邻居,实验结果表明,这种相似性度量方法的推荐质量最低.如果采用余弦相似性,则用户对所有未评分项目的评分均相同(均为 0),这显然是不精确的,由实验结果可知,这种相似性度量方法的推荐质量比相关相似性度量方法更好.在基于项目评分预测的协同过滤推荐算法中,首先通过项目之间的相似性初步预测了用户对未评分项目的评分,使得用户之间共同评分的项目比较多,然后通过余弦相似性度量方法计算用户之间的相似性,实验结果证明,基于项目评分预测的协同过滤推荐算法推荐质量最高.

5 小 结

本文首先深入分析了在用户评分数据极端稀疏情况下,余弦相似性、相关相似性和修正的余弦相似性度量方法在计算目标用户的最近邻居时存在的问题.针对上述问题,提出了一种基于项目评分预测的协同过滤推荐算法,这种方法可以有效地解决上述度量方法存在的不足,使得计算得到的目标用户的最近邻居比较准确.实验结果表明,基于项目评分预测的协同过滤推荐算法可以有效地解决用户评分数据的极端稀疏性以及用户评分数据极端稀疏情况下,传统的相似性度量方法存在的弊端,显著地提高推荐系统的推荐质量.

References:

- [1] Breese J, Hecherman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering. In: Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI'98). 1998. 43~52.
- [2] Goldberg D, Nichols D, Oki BM, Terry D. Using collaborative filtering to weave an information tapestry. Communications of the ACM, 1992,35(12):61~70.
- [3] Resnick P, Iacovou N, Suchak M, Bergstrom P, Riedl J. Grouplens: An open architecture for collaborative filtering of netnews. In: Proceedings of the ACM CSCW'94 Conference on Computer-Supported Cooperative Work. 1994. 175~186.
- [4] Shardanand U, Maes P. Social information filtering: Algorithms for automating "Word of Mouth". In: Proceedings of the ACM CHI'95 Conference on Human Factors in Computing Systems. 1995. 210~217.
- [5] Hill W, Stead L, Rosenstein M, Furnas G. Recommending and evaluating choices in a virtual community of use. In: Proceedings of the CHI'95. 1995. 194~201.

[6] Sarwar B, Karypis G, Konstan J, Riedl J. Item-Based collaborative filtering recommendation algorithms. In: Proceedings of the 10th International World Wide Web Conference. 2001. 285~295.

[7] Chickering D, Hecherman D. Efficient approximations for the marginal likelihood of Bayesian networks with hidden variables. *Machine Learning*, 1997,29(2/3):181~212.

[8] Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 1977,B39:1~38.

[9] Thiesson B, Meek C, Chickering D, Heckerman D. Learning mixture of DAG models. Technical Report, MSR-TR-97-30, Redmond: Microsoft Research, 1997.

[10] Sarwar B, Karypis G, Konstan J, Riedl J. Analysis of recommendation algorithms for E-commerce. In: ACM Conference on Electronic Commerce. 2000. 158~167.

[11] Wolf J, Aggarwal C, Wu K-L, Yu P. Horting hatches an egg: A new graph-theoretic approach to collaborative filtering. In: Proceedings of the ACM SIGMOD International Conference on Knowledge Discovery and Data Mining. San Diego, 1999. 201~212.

[12] Sarwar BM, Karypis G, Konstan JA, Riedl J. Application of dimensionality reduction in recommender system—A case study. In: ACM WebKDD 2000 Workshop. 2000.

[13] Aggarwal CC. On the effects of dimensionality reduction on high dimensional similarity search. In: ACM PODS Conference. 2001.



几何造型与处理 2004 国际会议 征文通知

Geometric Modeling and Processing(GMP)是国际上有重要影响的学术会议，每两年举办一次，会议论文集由 IEEE Computer Society Press 出版，优秀论文还将在 *Computer Aided Design* 和 *Computer Aided Geometric Design* 两个代表该领域国际最高学术水平的刊物上以专辑形式发表。GMP2004 将于 2004 年 4 月 13 日~15 日在清华大学召开。大会主席孙家广院士，程序委员会主席是维也纳技术大学的 Helmut Pottmann 教授和清华大学计算机系胡事民教授。

征稿范围：

- | | | |
|--------|------------|---------------|
| 曲线曲面 | CAGD 的数学基础 | 实体造型 |
| 多分辨率造型 | 形状表示技术 | 数字几何处理 |
| 几何约束求解 | 几何不确定性 | 几何计算中的偏微分方程方法 |

同时,也欢迎讨论以下应用领域的几何计算问题的论文:

- | | | |
|-------|---------|-------|
| 反向工程 | 图像处理 | 计算机视觉 |
| 机器人技术 | 快速成型 | 网格生成 |
| 医学成像 | 科学计算可视化 | |

投稿截止时间:2003 年 9 月 30 日

有关投稿要求, 请浏览会议网页: <http://cg.cs.tsinghua.edu.cn/GMP2004/>

联系人: 严寒冰、蔡晓华, 电话: 010-62782052, Email: gmp2004@tsinghua.edu.cn