

# 基于增加相似度系数的加权二部图推荐算法

李镇东 罗琦 施力力

(南京信息工程大学信息与控制学院江苏省气象能源利用与控制工程技术研究中心 南京 210044)  
(南京信息工程大学信息与控制学院江苏省大气环境与装备技术协同创新中心 南京 210044)

**摘要** 基于二部图的推荐算法是个性化推荐领域的一个研究热点,其中,如何科学地利用用户的评分资源,在评分数据不全的情况下对目标用户进行准确高效的推荐是研究难点,也因此受到众多学者的关注。因此,提出了一种以单调饱和函数为权,利用目标用户和其他项目共同评分个数相对用户总数均值的正切值作为传统相似度系数的推荐算法;同时,对调整系数后的相似度进行降序排列,利用前  $K$  个最近邻居集对目标用户进行推荐。实验结果表明,改进后的算法提高了推荐的准确性,降低了复杂度。

**关键词** 个性化推荐,加权二部图,单调饱和,准确性

**中图分类号** TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.7.047

## Weighted Bipartite Network Recommendation Algorithm Based on Increasing Similarity Coefficient

LI Zhen-dong LUO Qi SHI Li-li

(Jiangsu Engineering Research Center on Meteorological Energy Using and Control, School of Information and Control, Nanjing University of Information Science and Technology, Nanjing 210044, China)

(Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology, School of Information and Control, Nanjing University of Information Science and Technology, Nanjing 210044, China)

**Abstract** The recommendation algorithm based on bipartite networks is a research hotspot in the personalized recommendation system, while the difficulty of research is how to make use of the users' evaluation resources scientifically to work out an efficient and accurate recommendation for target users in the absence of rating data. Meanwhile, it has received sufficient attention of scholars. Therefore, a new recommendation algorithm was put forward with the monotonous saturation function as weight, and tangent of target users and other projects' common rating numbers against the mean value of total users is used as traditional similarity coefficient. At the same time, after the coefficient gets adjusted, the similarity will be sorted in descending order, the set of the first  $K$  nearest neighbors of which can be utilized for target users' recommendation. The experimental results prove that the revised algorithm improves the accuracy of recommendation and reduces its complexity.

**Keywords** Personalized recommendation, Weighted bipartite networks, Monotonic saturation, Accuracy

## 1 引言

随着数字信息技术的快速发展,人们每天所接触到的信息量正在以指数级的速度增长,信息过载的问题日益明显。如何让人们在最短的时间内找到他们最需要的信息并且让网络中的暗信息<sup>[1]</sup>能够被用户获取,成为了当今研究的热点。个性化推荐系统应运而生,它能够快速地发现人们的兴趣与需求。推荐系统是指电子商务网站根据用户的购买行为发现用户的兴趣点,进而向用户提出一些参考性的购买信息和建议,以便帮助用户决定购买什么产品,代替了销售人员现场向用户推荐商品的行为<sup>[2,3]</sup>。现有的推荐系统算法有基于内容的推荐算法、协同过滤推荐算法、混合推荐算法以及基于二部图网络结构的推荐算法等。

基于内容的推荐算法<sup>[4,5]</sup>与用户相关联的商品信息为

根据,找出相似的商品推荐给用户。该方法需要分析所有用户或者项目内容的信息,根据用户的兴趣或者商品属性(用途、类别等)的信息来对目标用户进行相关的推荐。基于内容的推荐是信息过滤的一种演化,需要依靠文本信息进行处理,但是对于非文本信息(音乐、图像等)往往无法处理,不能帮助用户发现相关的兴趣点。

协同过滤(Collaborative Filtering, CF)推荐<sup>[6,7]</sup>通过收集用户的信息,利用已有的信息计算用户间的相似度,通过筛选找到目标用户的最近邻居集,将邻居的信息作为预测目标用户可能喜欢的商品的依据。该算法不受数据格式影响,能够快速有效地为用户进行推荐,且可以解决长尾问题,但是该算法存在数据稀疏性、冷启动等方面的问题。

基于内容的推荐算法与协同过滤推荐算法都有各自的局限性,在实际应用中,通常将多种推荐算法组合起来,以期达

到稿日期:2015-06-23 返修日期:2015-09-26 本文受国家自然科学基金资助项目(61174077)资助。

李镇东(1990-),男,硕士生,主要研究方向为数据挖掘、推荐算法、信息检索, E-mail: 703396507@qq.com; 罗琦(1958-),男,教授,博士生导师,主要研究方向为随机动力系统分析; 施力力(1991-),男,硕士生,主要研究方向为随机稳定性。

到取长补短的作用。比较常见的是将上述两种推荐算法进行组合,混合推荐在实际应用中比单一推荐方式具有更高的推荐效率和准确性<sup>[8]</sup>,但同时也增加了时间和空间的开销。

随着推荐系统的发展,基于用户-项目二部图网络结构的推荐算法得到了越来越多学者的关注。受复杂网络物质扩散和热传导思想的启发,Zhou 等人<sup>[9]</sup>和 Liu 等人<sup>[10]</sup>提出利用资源分配的方法计算用户间的相似性,取得了良好的效果。Huang 等人<sup>[11]</sup>在已有的算法中首先加入扩散动力学来填充评分矩阵,在缓解数据稀疏性的基础上进行有效推荐。Zhang 等人<sup>[12]</sup>在 Zhou 等人<sup>[9]</sup>提出的网络推断(Network-Based Inference, NBI)算法的基础上引入评分权重,提出加权网络推断(Weighted Network-Based Inference, WNBI)算法。以用户对项目的评分作为权重,按权重进行资源分配矩阵的计算,该算法在没有增加时间和空间开销的情况下提高了推荐精度。Wang 等人<sup>[13]</sup>在区分高低分的情况下引入项目度与项目权值之和的比值 $\theta$ ,提高了推荐的准确性和多样性。但是上述算法在评分数据不全的情况下不能确保推荐的准确性,设置评分权重时没有更细致地划分梯度并且在预测评分时没有考虑目标用户的最近邻居集,而是利用所有用户的信息,增加了算法的复杂度。

针对上述问题,提出了一种基于增加相似系数系数的加权二部图推荐算法。首先利用评分资源,对评分进行单调饱和和函数加权,再计算用户间相似度,然后对相似度进行降序排列,再查找目标用户的 $K$ 个最近邻居集合来对目标用户进行有效推荐。计算相似度时结合目标用户和其他用户共同评分项目的个数与用户总数比值的正切函数值来动态调整用户间的相似度。实验证明,所提算法不仅提高了预测的精度,而且减小了算法的复杂度。

## 2 基于加权二部图的推荐算法

二部图由 $n$ 个用户节点和 $m$ 个项目节点组成,如果用户 $u_i$ 购买或者选择过项目 $o_j$ ,则用户 $u_i$ 和项目 $o_j$ 就被一条边连接。这就构成了用户-项目二部图 $G(U, O, E)^{[9,10]}$ , $E$ 表示二部图的边,即连接用户和项目的边,定义用户集合 $U = \{u_1, u_2, u_3, \dots, u_n\}$ ,项目集合 $O = \{o_1, o_2, o_3, \dots, o_m\}$ ,因此可以用 $n+m$ 个节点表示整个系统。其中,用户和项目之间的信息用邻接矩阵 $A = \{a_{ij}\} \in R^{n \times m}$ 表示, $a_{ij} = 1$ 表示用户 $u_i$ 购买或选择过项目 $o_j$ ;否则, $a_{ij} = 0$ 。

### 2.1 相似度计算

Zhou 等人<sup>[9]</sup>和 Liu 等人<sup>[10]</sup>受物理学中动力传播原理的启发,将资源分配和热量传播思想融合到计算用户间的相似度当中。假设每个用户在初始时刻被分配了一定数量的资源(推荐能量),然后用户将自己的资源平均分给其已经选择的项目,每个项目再把其收到的资源平均分给选择它的用户。假设任意两个用户 $u_\alpha$ 和 $u_\beta$ 之间存在某种能量关系,权重 $S_{\alpha\beta}$ 表示用户 $u_\beta$ 可能贡献给用户 $u_\alpha$ 的资源额度,即两个用户的相似度,如式(1)所示。

$$S_{\alpha\beta} = \frac{1}{k(u_\beta)} \sum_{i=1}^m \frac{a_{\alpha i} a_{\beta i}}{k(o_i)} \quad (1)$$

其中, $k(u_\beta) = \sum_{i=1}^m a_{\beta i}$ 表示用户 $u_\beta$ 的度(该用户选择过项目的数目); $k(o_i) = \sum_{l=1}^n a_{li}$ 表示项目 $o_i$ 的度(该项目被用户选择过的数

目)。为表述方便,将式(1)刻画两用户相似度的算法简称为算法(A)。

Wang 等人<sup>[13]</sup>在 Zhou 等人<sup>[9]</sup>的算法的基础上提出了加权网络结构的推荐算法。对每条用户-项目的连边赋一个权重 $\omega$ 。若用户 $u_i$ 购买或选择过项目 $o_j$ 且评分大于等于3,则 $\omega_{ij} = 1$ ;若用户 $u_i$ 购买或选择过项目 $o_j$ 但评分小于3,则 $\omega_{ij} = \lambda$ (文献[13]通过实验验证,最后 $\lambda$ 取0.5时推荐精度最优,本文在后续实验中亦使用该最优值);若用户 $u_i$ 未购买或选择过项目 $o_j$ ,则 $\omega_{ij} = 0$ 。在计算用户间的相似性系数时,该算法引入项目度与项目的权值之和的比值 $\theta$ ,即 $\theta(o_i) = \frac{k(o_i)}{d(o_i)}$ , $k(o_i)$ 表示项目 $o_i$ 的度。同时引入函数 $f(\theta) = \delta + e^{\lambda(\theta-1)}$ 来动态调节推荐准确性和多样性(文献[13]通过实验验证,最后确定 $\delta = 0.8$ 时推荐精度最优,本文在后续实验中亦使用该最优值),如式(2)所示。

$$S_{\alpha\beta} = \frac{1}{d(u_\beta)} \sum_{i=1}^m \frac{\omega_{\alpha i} \omega_{\beta i}}{d^{f(\theta)}(o_i)} \quad (2)$$

其中, $d(u_\beta) = \sum_{i=1}^m \omega_{\beta i}$ 表示用户 $u_\beta$ 与所选择的项目之间连边的权重总和; $d(o_i) = \sum_{l=1}^n \omega_{li}$ 表示项目 $o_i$ 与选择它的用户间连边的权重总和。以下将由式(2)刻画两用户相似度的算法简称为算法(B)。

### 2.2 生成推荐

计算目标用户 $u_\alpha$ 对未评分项目 $o_i$ 的预测评分 $v_{\alpha i}$ 的公式为:

$$v_{\alpha i} = \frac{\sum_{\beta=1, \beta \neq \alpha}^n S_{\alpha\beta} \times \omega_{\beta i}}{\sum_{\beta=1, \beta \neq \alpha}^n S_{\alpha\beta}} \quad (3)$$

其中, $S_{\alpha\beta}$ 为目标用户 $u_\alpha$ 与用户 $u_\beta$ 之间的相似度, $\omega_{\beta i}$ 为用户 $u_\beta$ 对项目 $o_i$ 的评分权重。以下将由式(3)进行预测评分的算法简称为算法(C)。

### 2.3 算法分析

在用户评分资源不全的情况下,使用加权二部图推荐算法(B)<sup>[13]</sup>存在几点不足:

1)在推荐系统中,有个共识的规律:在用户数目相同的情况下,如果目标用户与其他用户共同购买或者选择过的项目个数越多,且对共同购买或选择过的项目的评分越接近,则他们的相似度则越高。但是在评分数据不准确的情况下(例如评分数量不充足),就会存在用户间只有一个或较少的共同评分项目的情况,此时直接使用算法(B)计算,他们之间的相似度却非常高。如表1所列,与用户E相似度最高的应该是用户C,他们拥有3个共同评分项目,分数也相对比较接近,但用算法(B)求出E和D的相似度却超过用户E和C之间的相似度,这样的结果是不符合平时的规律的。并且在实际的数据集中,用户间的共同评分往往很多都与表1的这种情况类似,这会导致在预测环节中出现较大的偏差。

表1 数据集1

	$i_1$	$i_2$	$i_3$	$i_4$	$i_5$	$i_6$	$i_7$	$i_8$
A								
B					5			
C		5		4		5	4	5
D	5					2		
E	5	4	4				2	5

2)算法(B)中提出的用户评分权重没有更细致的划分,只有评分3分为分界点。如表2所列,用户A与用户B只有在项目*i*<sub>2</sub>上有1分之差,用户D和用户E在项目*i*<sub>7</sub>上有1分之差,按照算法(B)计算,两用户间相似度是一样的。虽然只有1分之差,但还是可以体现出打分者当时的心境。算法(B)并没有对评分权重进行更细致的划分,导致计算相似度时比较笼统,从而对预测精度产生了一定的影响。

表2 数据集2

	<i>i</i> <sub>1</sub>	<i>i</i> <sub>2</sub>	<i>i</i> <sub>3</sub>	<i>i</i> <sub>4</sub>	<i>i</i> <sub>5</sub>	<i>i</i> <sub>6</sub>	<i>i</i> <sub>7</sub>	<i>i</i> <sub>8</sub>
A	5	4	4	1	1	1	2	5
B	5	3	4	1	1	1	2	5
C	5	4	4	1	1	1	2	5
D	5	4	4	1	1	1	2	5
E	5	4	4	1	1	1	1	5

3)算法(C)在对目标用户进行预测评分时,目标用户对其余每个用户都进行了相似度的对比,见式(3)。如此一来在增加了算法的复杂度的同时降低了推荐精度。

## 2.4 改进思路

针对2.3节中1)和2)的不足,提出了一种以单调饱和函数为权重的评分加权方式,通过对比两种单调饱和函数:反正切函数和hill函数,最终将hill函数作为评分权重。利用目标用户和其他用户共同评分项目的个数与用户总数比值的反正切函数值来动态调整用户间的相似度。选择了几种函数进行对比分析,包括幂函数、对数函数和三角函数等。通过计算发现利用反正切函数进行调整更符合我们的需求,准确度也优于其他改进方法,得到了更准确的最近邻居集,提高了二部图推荐算法的精度。

针对2.3节中3)的不足,在对目标用户进行预测之前,先对目标用户进行相似度的降序排列,完成最近邻居集合的筛选,在最近邻居集的基础上对目标用户进行推荐。实验发现,该算法融合最近邻居集思想之后,算法的复杂度减小了,算法的精度得到了提高。

## 3 改进算法

### 3.1 评分权重及用户相似度的改进算法

在实际推荐系统中,由于评分数量非常少,导致评分数据不准确,共同评分数目极少。因此,在计算用户间的相似度时,使用算法(B)会出现大量类似于算法分析中1)所提到的问题,导致不能精确地为目标用户找到最近邻居集,预测分数和最后的推荐精度也会受到影响。

本文利用目标用户与其他用户共同评分项目越多则其兴趣越相似的原则来动态调整用户间的相似度,并且选用单调饱和和hill函数作为评分权重 $\omega_{ij}$ ,用户 $u_i$ 对项目 $o_j$ 的评分为 $R_{ij}$ ,则 $\omega_{ij} = \frac{R_{ij}^\xi}{(1+R_{ij}^\xi)}$ , $\xi$ 为hill系数。

改进算法采用项目间共同评分数目相对用户数目均值的正切值来动态调整相似度,改进的相似度如下:

$$sim(\alpha, \beta) = \tan\left(\frac{I(\alpha) \cap I(\beta)}{n}\right) * \sum_{i=1}^m \frac{\omega_{\alpha i} \omega_{\beta i}}{d(u_\alpha) d^{f(\theta)}(o_i)} \quad (4)$$

其中,用户 $u_\alpha$ 和 $u_\beta$ 的共同评分项目用 $I(\alpha) \cap I(\beta)$ 表示, $n$ 表示用户的数量。从式(4)中可看出正切值作为相似度的改进系数的作用反映在用户 $u_\alpha$ 与用户 $u_\beta$ 所选择的项目中,共同购买或选择的项目越多,则 $u_\alpha$ 与 $u_\beta$ 的相似度越大。以下将

由式(4)刻画两用户相似度的算法简称为算法(D)。

从表1的数据中可以判断出与用户E相似度最高的应该是用户C,他们有3个共同评分项目,且评分也相对接近;然而从表3的传统相似度数据中可以看出,运用算法(B)计算用户E与用户D的相似度为0.191;用户E与用户C的相似度为0.159,这是不符合常理的。使用改进的算法(D)之后,可以看出用户E与用户C的相似度为0.818,大于用户E与用户D的相似度0.219,符合我们的预期想法。

表3 “表1”相似度对比

传统相似度					
	A	B	C	D	E
A	1	0	0	0	0
B	0	1	0	0	0
C	0	0	1	0.149	0.177
D	0	0	0.044	1	0.064
E	0	0	0.159	0.191	1
改进相似度					
	A	B	C	D	E
A	1	0	0	0	0
B	0	1	0	0	0
C	0	0	1	0.216	0.82
D	0	0	0.086	1	0.087
E	0	0	0.818	0.219	1

从表2的数据中可以看出用户A与用户B之间只有在项目*i*<sub>2</sub>上有1分之差,用户D和用户E只有在项目*i*<sub>7</sub>上有1分之差。但是从表4的数据中可以看出,使用算法(B)计算出的用户A与用户B之间的相似度和用户A与用户C之间的相似度是相同的,都为0.054;同理,用户C与用户D、用户E之间的相似度也是相同的,这显然是不合理的。通过使用改进之后的算法(D)可以看出,虽然只有1分之差,但相似度仍然存在一定的差别,如表4中加粗部分所列。

表4 “表2”相似度对比

传统相似度					
	A	B	C	D	E
A	1	0.054	0.054	0.054	0.054
B	0.054	1	0.054	0.054	0.054
C	0.054	0.054	1	0.054	0.054
D	0.054	0.054	0.054	1	0.054
E	0.054	0.054	0.054	0.054	1
改进相似度					
	A	B	C	D	E
A	1	14.848	14.876	14.846	14.961
B	14.711	1	14.711	14.711	14.796
C	14.876	14.848	1	14.876	14.961
D	14.876	14.848	14.876	1	14.961
E	14.876	14.471	14.503	14.503	1

### 3.2 基于最近邻居集合的预测算法改进

从预测算法(C)可以看出,计算目标用户 $u_\alpha$ 对未评分项目 $o_i$ 的预测评分 $v_{\alpha i}$ 中,利用了目标用户和所有其他用户的相似度信息对目标用户进行预测。而在现实生活中,人们需要买东西但没有目标时,会倾向于询问一些跟他们有一定社会关系、相互认识且具有相同兴趣或者研究领域的人的意见来帮助作出选择。简而言之,当一个用户需要个性化推荐时,可以先找到和他有相似兴趣的其他用户,然后把那些用户喜欢的、而该用户没有听说过的物品推荐给他<sup>[17]</sup>。从上面的描述中可以看出,该方法主要包括两个步骤:

1)找到与目标用户具有相似兴趣的用户集合;

2)找到这个集合中其他用户喜欢的且目标用户没有选择过的项目推荐给目标用户。

根据上述思想,本文在算法(C)的基础上融合最近邻居集来对目标用户未选择的项目进行预测,并对预测结果进行排名,取前  $N$  个项目推荐给目标用户。具体预测算法如下:

$$v_{oi} = \frac{\sum_{\beta \in S(I)} \text{sim}(\alpha, \beta) \times \omega_{\beta i}}{\sum_{\beta \in S(I)} \text{sim}(\alpha, \beta)} \quad (5)$$

其中,  $v_{oi}$  表示用户  $u_{\alpha}$  对未评分的项目  $o_i$  的预测评分,  $\text{sim}(\alpha, \beta)$  为用户  $u_{\alpha}$  和  $u_{\beta}$  的相似度,  $S(I) = \{I_1, I_2, \dots, I_k\}$  为目标用户  $u_{\alpha}$  的  $K$  个最近邻居集合。以下将由式(5)进行预测评分的算法简称为算法(E)。

实验发现(见图2),在预测算法中加入用户评分的期望值可以更真实地体现用户的购买心理,并且可以获得更高的推荐精度,如式(6)所示。为体现优越性,两种预测算法的精度实验对比如图2所示。

$$v_{oi} = R_{\alpha}^{\text{avg}} + \frac{\sum_{\beta \in S(I)} \text{sim}(\alpha, \beta) \times (R_{\beta i} - R_{\beta}^{\text{avg}})}{\sum_{\beta \in S(I)} \text{sim}(\alpha, \beta)} \quad (6)$$

其中,  $v_{oi}$  表示用户  $u_{\alpha}$  对未评分的项目  $o_i$  的预测评分,  $R_{\alpha}^{\text{avg}}$ ,  $R_{\beta}^{\text{avg}}$  分别表示用户  $u_{\alpha}$  和  $u_{\beta}$  的平均评分,  $\text{sim}(\alpha, \beta)$  为用户  $u_{\alpha}$  和  $u_{\beta}$  的相似度,  $S(I) = \{I_1, I_2, \dots, I_k\}$  为目标用户  $u_{\alpha}$  的  $K$  个最近邻居集合。以下将由式(6)进行预测评分的算法简称为算法(F)。

通过分析图2可知,预测算法(F)的精度优于算法(E),因此本文最终采用算法(F)来预测目标用户对项目的评分。

## 4 算法的详细步骤

算法的详细步骤如图1所示。

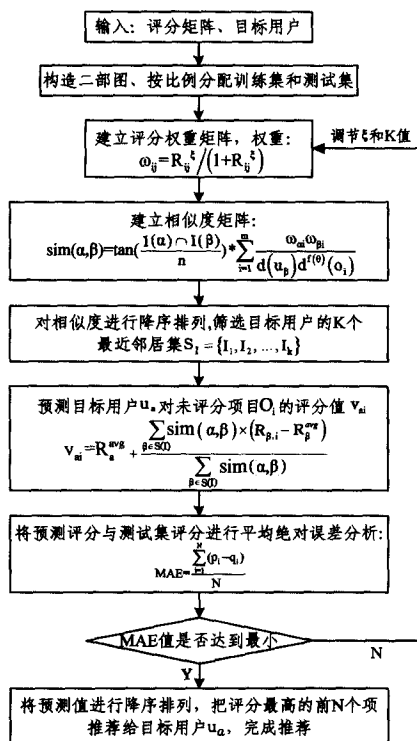


图1 算法详细步骤

下面对本文算法时间复杂度进行分析。在二部图推荐系统中,有  $n$  个用户和  $m$  个项目。算法(D)中  $f(\theta)$  的参数  $\delta, \lambda$

以及  $hill$  的系数  $\xi$  的值可以在线下计算,节省了在线的计算时间;算法(D)中用户间的相似度、用户间共同评分项目个数都可以离线进行计算,时间复杂度为  $O(m^2)$ ;算法(F)用来计算目标用户对未评分项目的预测评分,需要在线计算,其时间复杂度为  $O(K)$ (其中  $K$  为最近邻居集个数,且  $K < n$ )。而在文献[13]中,计算用户间相似性时的时间复杂度为  $O(m^2 \cdot n)$ ,计算用户未评分项目的评分时的时间复杂度为  $O(n)$ 。通过比较可以发现,本文算法降低了复杂度,提高了推荐效率。

## 5 实验分析

### 5.1 数据集

本实验采用的数据来源于 Movie lens 站点提供的公开数据集(<http://movielens.umn.edu/>)。它包含了 943 个用户对 1682 部电影的 100000 条评分数据,每个用户对电影都必须有至少 20 条评价。用户评分范围为 1~5,分值越大表示用户对项目的喜爱程度越大。实验将评分数据集分为 80% 的训练集和 20% 的测试集,评分数据集的稀疏度定义为已评分项目/总评分项目,因此本实验数据集的稀疏度为:

$$1 - \frac{100000}{943 * 1682} = 93.69\%$$

### 5.2 度量标准

度量推荐系统的推荐质量采用的是平均绝对误差(MAE)和均方根误差(RMSE)。MAE 和 RMSE<sup>[22]</sup>都是通过计算预测项目集合与实际项目评分集合的偏差来判定推荐的准确性,误差值越小,推荐精度越好。

设预测用户的评分集合表示为  $\{p_1, p_2, \dots, p_N\}$ ,实际用户评分集合表示为  $\{q_1, q_2, \dots, q_N\}$ 。算法的 MAE 和 RMSE 可通过式(7)和式(8)表示:

$$MAE = \frac{\sum_{i=1}^N (p_i - q_i)}{N} \quad (7)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (p_i - q_i)^2}{N}} \quad (8)$$

### 5.3 预测算法比较

通过实验分析预测算法(E)和预测算法(F),发现算法(F)的精度优于算法(E),比较如图2所示,横坐标为最近邻居个数  $K$ ,纵坐标为 MAE 值。

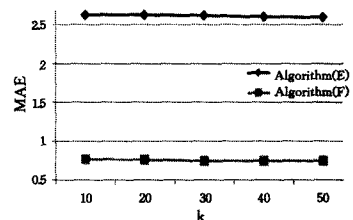


图2 预测算法比较

如图2所示,算法(F)的推荐精度明显优于算法(E)的推荐精度。所以后面的实验均在采用预测算法(F)的环境下进行比较分析。

### 5.4 改进的相似度算法的比较

对文献[13]中的算法进行实验分析时发现,在同时运用算法(C)进行预测(没有运用最近邻居集进行预测)时,文献[13]提出的相似度算法(算法(B))的 MAE 值达到 3.5256,运

用本文提出的相似度算法(算法(D))的 MAE 值为 3.3363。实验结果如图 3 所示,本文算法明显优于文献[13]提出的相似度算法。

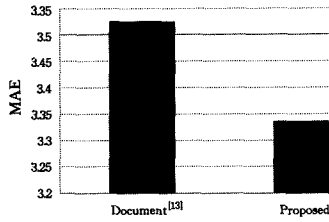


图 3 相似度对比

在都运用了最近邻居集预测算法(F)进行预测的环境下,将本文的相似度算法(算法(D))与文献[13]提出的相似度算法(算法(B))以及文献[18]提出的基于项目评分的协同过滤算法进行 MAE 值的对比,实验结果如图 4 所示。本实验的目的是体现本文改进的算法优于其他研究者改进的算法,同时也证明了运用 K 个最近邻居集的信息对目标用户进行推荐的方式优于运用其他所有用户的信息对目标用户进行推荐的方式。横坐标选为最近邻居个数 K,纵坐标为 MAE 值,邻居个数从 10 开始,间隔为 10,最终增至 50 个。

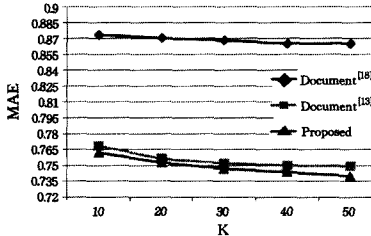


图 4 比较相似度算法的 MAE

在都运用了最近邻居集算法(F)进行预测的环境下,将本文的相似度算法(算法(D))与文献[13]提出的相似度算法(算法(B))及文献[19]提出的一种改进的协同过滤算法进行 RMSE 值的对比。实验结果如图 5 所示,本实验的目的是体现本文改进的算法优于其他研究者改进的算法。横坐标选为最近邻居个数 K,纵坐标为 RMSE 值,邻居个数从 10 开始,间隔为 10,最终增至 50 个。

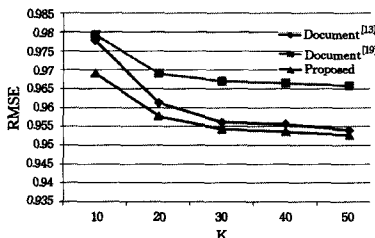


图 5 比较相似度算法的 RMSE

如图 4、图 5 所示,随着最近邻居的个数的增加,进行比较的几种算法的 RMSE 和 MAE 值都呈下降趋势,最终稳定在一个位置。本文算法从邻居个数为 10 开始,MAE 和 RMSE 值都小于另外两种算法,并且小于利用所有其他用户信息进行预测(算法(C))的算法。当 K 达到 50 时基本稳定,MAE 值达到 0.7396, RMSE 值达到 0.9536。实验证明利用正切函数来动态调整相似度是可以实现的。

### 5.5 hill 系数对推荐精度的影响

本文最终选定单调饱和函数 hill 函数作为评分权重:

$$\omega_{ij} = \frac{R_{ij}^e}{(1+R_{ij}^e)^\xi}$$
,  $\xi$  用于调节推荐准确性。在实验中  $\xi$  取 2 最为合适,即 MAE 值最小。在实验时,最近邻居个数 K 取为 50,实验结果如图 6 所示,横坐标为 hill 系数。

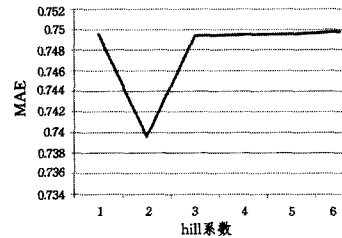


图 6 hill 系数对推荐准确性的影响

### 5.6 评分权重对推荐精度的影响

本文最终选择系数为 2 的 hill 函数作为推荐算法的评分权重。在运用最近邻居集思想进行预测的基础上,逐渐增加最近邻居个数,将本文提出的评分权重 hill 函数与文献[13]提出的评分权重以及以反正切函数(单调饱和函数)为权重的算法进行 MAE 值的比较。本实验的目的是体现本文改进的算法优于其他研究者改进的算法,实验结果如图 7 所示,横坐标为最近邻居个数,纵坐标为 MAE 值。

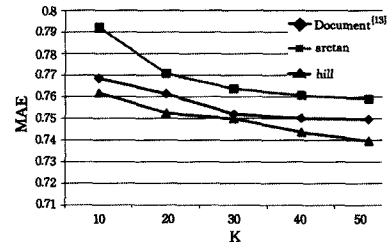


图 7 不同权重对推荐精度的影响

从图 7 的实验结果可以看出,本文采用的 hill 函数作为权重,在推荐精度上优于文献[13]的权重算法。在单调饱和函数中,与反正切函数进行了对比,发现其在推荐精度上以微弱的优势领先于反正切函数。

**结束语** 本文构建了切实可行的一种基于增加相似度系数的加权二部图推荐算法。改进了评分权重的划分不够细致的问题,完善了二部图推荐算法中的相似度算法,使得推荐精度得到了提高。在预测目标用户对目标项目的评分中,融合了最近邻居集的思想,根据相似度寻找目标用户的最近邻居集,降低了算法的复杂度。今后的研究方向是从评分矩阵的可信度入手,利用攻击检测算法筛选出评分不可信的用户,以提高推荐精度;同时用户以及项目冷启动也是未来的一个研究方向。在实验中发现程序运行时间较长,这也是将要改进的一个方面。

### 参考文献

- [1] Xu H L, Wu X, Li X D, et al. Comparison study of Internet recommendation system[J]. Journal of Software, 2009, 20(2): 1-10(in Chinese)  
许海玲,吴潇,李晓东,等. 互联网推荐系统比较研究[J]. 软件学报, 2009, 20(2): 1-10
- [2] Schafer J B, Konstan J, Riedl J. Recommender systems in E-commerce[C]//Proc of E-COMMERCE. 1999:158-166
- [3] Liu J G, Zhou T, Wang B H. Research progress of personalized

- recommendation system[J]. Progress in Natural Science, 2009, 19(1):1-12(in Chinese)
- 刘建国,周涛,汪秉宏. 个性化推荐系统的研究进展[J]. 自然科学进展, 2009, 19(1):1-12
- [4] Wang F H, Jian S Y. An effective content-based recommendation method for Web browsing based on keyword context matching[J]. Journal of Informatics and Electronics, 2006, 1(2):49-59
- [5] Wartena C, Slakhorst W, Wibbels M, et al. Selecting keywords for content based recommendation[C]//Proceedings of the 19th ACM International Conference on Information and Knowledge Management(CIKM 10). New York: ACM Press, 2010: 1533-1536
- [6] Herlocker J L, Konstan J A, Terveen L G, et al. Evaluating collaborative filtering recommender systems[J]. ACM Transactions on Information Systems, 2004, 22(1):5-33
- [7] Chen Y L, Cheng L C. A novel collaborative filtering approach for recommending ranked items[J]. Expert Systems with Applications, 2008, 34(4):2396-2405
- [8] Hong W X, Weng Y, Zhu S Z. Hybrid recommender system for vertical e-commerce website[J]. System Engineering Theory & Practice, 2010, 30(5):928-935(in Chinese)
- 洪文兴,翁洋,朱顺痣. 垂直电子商务网站的混合型推荐系统[J]. 系统工程理论与实践, 2010, 30(5):928-935
- [9] Zhou Tao, Ren Jie, Medo M, et al. Bipartite network projection and personal recommendation[J]. Physical Review E, 2007, 76(4):04115
- [10] Liu Jian-guo, Wang Bing-hong, Guo Qiang. Improved collaborative filtering algorithm via information transformation[J]. International Journal of Modern Physics C, 2009, 20(2):285-293
- [11] Huang Z, Chen H, Zeng D. Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering[J]. ACM Transactions on Information Systems, 2004, 22(1):116-142
- [12] Zhang X M, Jiang S Y. Personalized recommendation algorithm based on weighted bipartite network[J]. Journal of Computer Applications, 2012, 32(3):654-658(in Chinese)
- 张新猛,蒋盛益. 基于加权二部图的个性化推荐算法[J]. 计算机应用, 2012, 32(3):654-658
- [13] Wang Q, Duan S Y. Improved recommendation algorithm based on bipartite networks[J]. Application Research of Computers, 2013, 30(3):771-774 (in Chinese)
- 王茜,段双艳. 一种改进的基于二部图网络结构的推荐算法[J]. 计算机应用研究, 2013, 30(3):771-774
- [14] Tao W A, Fan H L. Collaborative filtering recommendation algorithm based on nearest-neighborhood and rating support[J]. Application Research of Computers, 2012, 29(5):1723-1725, 1728(in Chinese)
- 陶维安,范会联. 基于评分支持度的最近邻协同过滤推荐算法[J]. 计算机应用研究, 2012, 29(5):1723-1725, 1728
- [15] Ansari, Essegaier A S, Kohli R. Internet recommendation systems[J]. Journal of Marketing Research, 2000, 37(3):363-375
- [16] Herlocker J L. Evaluating Collaborative Filtering Recommender Systems[J]. ACM Transactions on Information Systems, 2004, 22(1):5-53
- [17] Xiang L. Practice of Recommendation system[M]. Beijing: Posts & Telecom Press, 2012(in Chinese)
- 项亮. 推荐系统实践[M]. 北京:人民邮电出版社, 2012
- [18] Deng A L, Zhu Y Y, Shi B L. A collaborative filter recommendation algorithm based on item rating prediction[J]. Journal of Software, 2003, 14(9):1621-1628(in Chinese)
- 邓爱林,朱扬勇,施伯乐. 基于项目评分预测的协同过滤推荐算法[J]. 软件学报, 2003, 14(9):1621-1628
- [19] Luo Q, Miao X J, Wei Q. Further research on collaborative filtering algorithm for sparse data[J]. Computer Science, 2014, 41(6):264-268(in Chinese)
- 罗琦,缪昕杰,魏倩. 稀疏数据集协同过滤算法的进一步研究[J]. 计算机科学, 2014, 41(6):264-268
- [20] Sun Z, Luo N, Kuang W. One real-time personalized recommendation systems based on Slope One algorithm[C]// Eighth International Conference on Fuzzy Systems and Knowledge Discovery( FSKD 2011). Shanghai, China. 2011:1826-1830
- [21] Liu Z K, Zhang C, Zhang Y C, et al. Solving the Cold-Start Problem in Recommender Systems with Social Tags[J]. Epl, 2010, 92(2):28002-28007
- [22] Luo X, Ouyang Y X, Xiong Z, et al. The effect of similarity support in K-nearest-neighborhood based collaborative filtering[J]. Chinese Journal of Computers, 2010, 33(8):1437-1445 (in Chinese)
- 罗辛,欧阳元新,熊璋,等. 通过相似度支持度优化基于 K 近邻的协同过滤算法[J]. 计算机学报, 2010, 33(8):1437-1445
- [23] Wang Pu, Ye Hong-wu. A Personalized Recommendation Algorithm Combining Slope One Scheme and User Based Collaborative Filtering [C]// International Conference on Industrial and Information Systems. 2009:152-154
- [24] Shani G, Brafman R, Heckerman D. An MDP-based recommender system[J]. Journal of Machine Learning Research, 2005, 6(1):1265-1295
- [25] Esslimani I, Brun A, Boyer A. Densifying a Behavioral Recommender System by Social Network Link Prediction Methods[J]. Social Network Analysis and Mining, 2011, 1(3):159-172
- [26] Herlocker J L, Konstan J A, Terveen L G, et al. Evaluating Collaborative Filtering Recommender Systems[J]. ACM Transactions on Information Systems, 2004, 22(1):5-53
- [27] Zhou Tao, KUSCSIK Z, Liu Jian-guo, et al. Solving the apparent diversity-accuracy dilemma of recommender systems[J]. PNAS, 2010, 107(10):4511-4515
- [28] Guo G, Zhang J, Yorke-Smith N. A novel Bayesian similarity measure for recommender systems[C]// International Joint Conference on Artificial Intelligence. 2013:2619-2625
- [29] Guo G, Zhang J, Thalmann D. A Simple But Effective Method to Incorporate Trusted Neighbors in Recommender Systems[C]// International Conference on User Modeling, Adaptation and Personalization. 2012:114-125