

# 基于人工免疫系统的推荐系统

高 静, 应吉康

(华东师范大学 信息学院 计算中心, 上海 200062)

**摘 要:** 简要介绍了免疫系统的基本概念, 以及人工免疫系统中的独特型网络模型, 将其算法思想引入推荐系统中, 提出了一种基于人工免疫系统的推荐算法。给出了算法的描述, 并对使用该算法前后的结果进行比较。数据实验结果表明, 该算法对推荐预测精确度的提高有明显的效果。

**关键词:** 人工免疫系统; 独特型网络; 免疫算法; 推荐系统

中图分类号: TP18

文献标识码: A

文章编号: 1673-629X(2007)05-0180-04

## A Recommendation System Based on Artificial Immune System

GAO Jing, YING Ji-kang

(Computing Center, Information Institute, East China Normal University, Shanghai 200062, China)

**Abstract:** Introduces the principle of immune system and idiotypic networks. Bring in the ideas to the recommendation system and propose a modified recommendation algorithm based on artificial immune system. Experimental results show that the proposed algorithm can achieve higher prediction accuracy.

**Key words:** artificial immune system; immune algorithm; idiotypic networks; recommendation system

### 0 引 言

免疫系统是高等脊椎动物体内能够识别和排除抗原性异物, 保护机体免受损害及维持内环境稳定的极为复杂的生物学系统<sup>[1]</sup>。自生物学家提出免疫理论以后, 人们发现, 免疫系统具有很多智能化的特征, 比如学习、识别、进化、自我组织能力等。而从信息处理的角度看, 免疫系统是与遗传系统、神经系统并列的人体三大信息系统之一。它具有大规模并行信息处理能力、强大的学习能力、记忆能力、识别能力、自适应性和鲁棒性、自组织能力和保持多样性的能力。之后, 人们利用这些特性针对具体工程问题进行数学建模, 来解决技术问题。在现在的日常生活中, 人们常常需要对未知事物进行选择, 这时就需要推荐系统的帮助。传统的推荐系统使用协作过滤技术<sup>[2]</sup>。有不少学者对协作过滤技术进行了进一步的探讨和研究, 但大多都是说明如何给出用户一系列好的推荐项目。而文中采用免疫系统的思想所提出算法, 是用来表明系统给出的推荐项目与用户实际上喜爱的是非常接近的。数据实验

结果表明了该算法的可行性和有效性。

### 1 人工免疫系统

免疫系统是一个动态系统。免疫细胞时刻处在旺盛的新陈代谢之中, 各种免疫效应分子在相互抑制和激励中维持着动态平衡。一般认为, 借鉴生物免疫系统的结构特征和工作机理, 为解决某方面工程技术问题所建立的系统或者算法结构, 统称为人工免疫系统(AIS)。

AIS网络模型研究开展很早, 迄今为止, 人们根据应用对象不同, 从不同角度模拟免疫系统, 提出了不少有价值的AIS网络模型。这里将用到1973年Jerne提出的独特型网络模型(Idiotypic Networks)<sup>[3]</sup>思想。他从抗体独特型出发, 认为抗体具有抗体和抗原两重属性, 抗体在识别抗原的同时, 也会被其它抗体识别而产生免疫反应, 因而整个免疫系统是一个通过抗体相互作用构成的网络结构。该理论对免疫记忆和免疫系统的进化等做出了较好的解释。之后, 一些免疫学者给出了数学框架<sup>[4]</sup>。1990年Farmer通过对免疫系统、神经系统以及遗传系统的特点及其异同进行分析比较后, 给出了基于联结主义的动态系统模型基本特点的描述。

下面的方程即为免疫系统的动态方程, 从

收稿日期: 2006-08-02

作者简介: 高 静(1981-), 女, 江苏南通人, 硕士研究生, 研究方向为信息系统研究与开发; 应吉康, 研究员, 硕士生导师, 研究方向为信息系统研究与开发。

Farmer<sup>[5]</sup>改写而来:

$$\frac{dx}{dt} = c \left[ \begin{pmatrix} \text{antibodies} \\ \text{recognised} \end{pmatrix} - \begin{pmatrix} I \text{ am} \\ \text{recognised} \end{pmatrix} + \begin{pmatrix} \text{antigens} \\ \text{recognised} \end{pmatrix} \right] - \begin{pmatrix} \text{death} \\ \text{rate} \end{pmatrix} = c \left[ \sum_{j=1}^N f_{ji} v_i v_j - k_1 \sum_{j=1}^N f_{ij} v_i v_j + \sum_{j=1}^n f_{ji} v_i u_j \right] - k_2 v_i \quad (1)$$

其中  $c$  是一个比例常量,由单位时间内抗体相遇的次数决定。 $N$  为抗体总数, $n$  为抗原总数。 $f_{ji}$  为抗体  $i$  与抗原  $j$  之间的匹配函数。第一项表示抗体  $i$  对抗体  $j$  的刺激度;第二项表示抗体  $i$  对抗体  $j$  的抑制度,常数  $k_1$  为抑制系数;第三项表示抗原  $j$  对抗体  $i$  的刺激度;最后一项表示细胞的死亡,常数  $k_2$  为自然死亡率。

由以上公式可以看出独特型影响可能是正也可能为负。更进一步,假设匹配函数是对称的,那么第一项与第二项的差值决定了独特型影响的正负性。现在来简化上述公式,假设在免疫系统中只允许一个抗原,简化公式如下:

$$\frac{dx}{dt} = k_1 f_i v_i u_j - \frac{k_2}{n} \sum_{j=1}^n f_{ij} v_i v_j - k_3 x_i \quad (2)$$

其中  $n$  为抗体总数, $f_i$  为抗体  $i$  与抗原(单个)之间的匹配函数, $f_{ji}$  为抗体  $i$  与抗体  $j$  之间的匹配函数。第一项是抗体  $i$  对抗原  $j$  的关系系数,常数  $k_1$  为刺激系数;第二项是抗体  $i$  对抗体  $j$  的关系系数,常数  $k_2$  为抑制系数;最后一项表示细胞的死亡,常数  $k_3$  为自然死亡率。

## 2 推荐系统

推荐系统是用于帮助用户确认对未知事物的兴趣度的。推荐系统通常使用协作过滤技术(CF)来进行预测推荐。协作过滤技术是从那些与之有相同喜好的人群里获得推荐信息来给指定用户提供推荐。协作过滤技术可以推荐给你的资源,有时你甚至对它的内容毫不知情,这是该技术的优势所在。有许多对协作过滤的改进算法<sup>[6]</sup>。推荐系统中的协同过滤技术利用客户的历史喜好信息计算客户之间的距离,目标客户对特定项目的喜好程度由其邻居对项目评价的加权平均值来计算,因此其核心是为一个需要推荐服务的当前用户寻找最相似的“最近邻居”集(Nearest-neighbor),即:对于一个目标用户,要产生一个根据相似度大小排列的“邻居”集合。常用算法有两种:一种是阈值法;二是 Top-N 法。阈值法是事先设定一相似性阈值,当用户与目标用户的相似性大于设定阈值,该用户就被认为是邻居;Top-N 法则是事先设定  $n$  个邻居用户数,在比较与目标用户的相似性后,取  $n$  个与目标用户最为相似的用户作为邻居。

统计上认为,具有最低相似性的用户对于推荐预测一般不起作用,所以在理论上应该取相似性高的用户,也就是与当前用户具有非常相似喜好的用户对推荐预测才能起作用,当然理论上相似性越高越好。但是由经验而知,如果用来推荐预测的“邻居”集用户数过少,预测也不会太准确,从这个角度应该是邻居用户越多越好,但是在某阶段系统中的用户有限,这是个客观问题。如何解决这一矛盾,在目前的现有算法中,没有一个统一的方法,都是凭经验解决。

在此,用免疫算法来解决这一问题,源于生物免疫系统当中存在的多样性自动调节,这种独特的浓度调节机制能保持免疫细胞群体的多样性,形成抗体的同时,还包括每个抗体的浓度,浓度高的抗体与抗原的相似性高,对推荐起主导作用,同时低浓度的抗体加入保证邻居用户数,对推荐预测起辅助作用。使用免疫算法很好的解决该矛盾,大大提高了推荐的质量。以影片推荐系统为例,其功能是:对于指定影片推荐给哪些观众;对于指定观众推荐哪些影片。

### 2.1 算法步骤

下面给出基于独特型人工免疫网络的推荐算法的核心部分。

算法步骤如下:

步骤 1:初始化人工免疫系统(AIS)。

步骤 2:对目标用户进行编码作为抗原(Ag)。

步骤 3:运行条件 while(系统不稳定)and(存在可用的数据),则:

(1) 将另一个用户作为抗体(Ab)加入到系统中;

(2) 计算抗体与抗原之间的匹配度;

(3) 计算抗体与其他抗体之间的匹配度;

(4) while(系统不稳定)and(系统数据容量饱和)

则:

迭代人工免疫系统

DO

步骤 4:while 循环后结束。

系统终止的情况:①不再存在可用的数据;②系统已稳定,例如系统数据饱和且再迭代十次也不会有什么改变。

每一次迭代改变一个抗体的浓度,这是通过:增加该抗体与抗原直接匹配的部分,减少那些与其他抗体匹配的部分。如果该抗体两者都无,则该抗体的浓度将不断降低。当该抗体的浓度低于每个阈值就从系统中除去该抗体,否则继续保存。此算法的目的是增加那些互相不同的抗体(数据库用户)的浓度,这些抗体与抗原(目标用户)有一定的相似度(喜好情况)。这个过程正好符合抗体之间的抑制度,即上面提到的 Jern

的独特型思想。

### 2.2 系统工作描述

现假设在有 5 个用户和 4 个项目。用户集  $U = \{u_1, u_2, u_3, u_4, u_5\}$ , 项目集  $M = \{m_1, m_2, m_3, m_4\}$ 。关系表为:

- $u_1 = \{(m_1, v_{11}), (m_2, v_{12}), (m_3, v_{13})\}$
- $u_2 = \{(m_1, v_{21}), (m_2, v_{22}), (m_3, v_{23}), (m_4, v_{24})\}$
- $u_3 = \{(m_1, v_{31}), (m_4, v_{34})\}$
- $u_4 = \{(m_1, v_{41}), (m_4, v_{44})\}$
- $u_5 = \{(m_1, v_{51}), (m_2, v_{52}), (m_3, v_{53}), (m_4, v_{54})\}$

其中  $v_{ij}$  表示用户  $u_i$  对项目  $m_j$  的评价值。

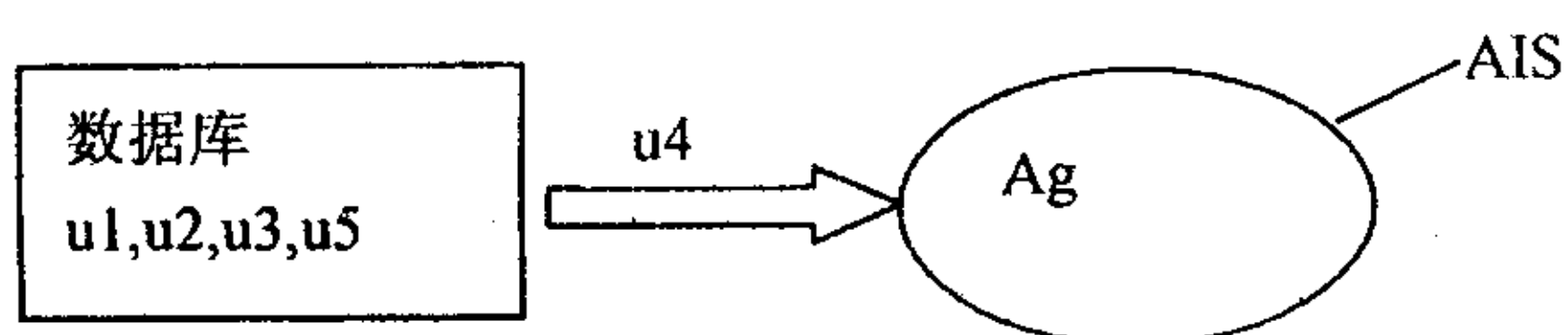
在现有数据信息中,并未拥有每个用户对每个项目的评价值。现在来预测用户  $u_4$  对项目  $m_3$  的评价。

根据上述系统及算法流程步骤如下:

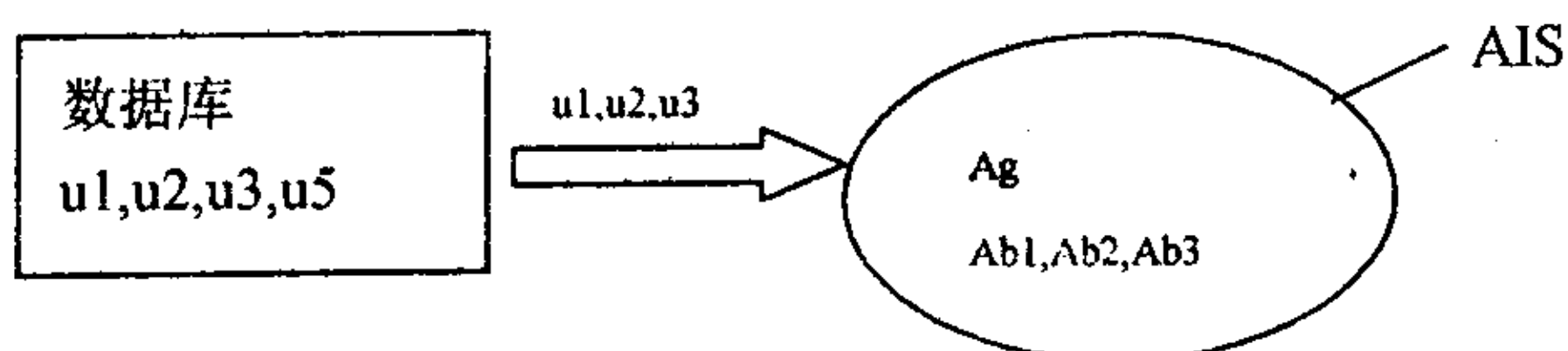
1)开始免疫系统为空,待从数据库中输入信息数据。



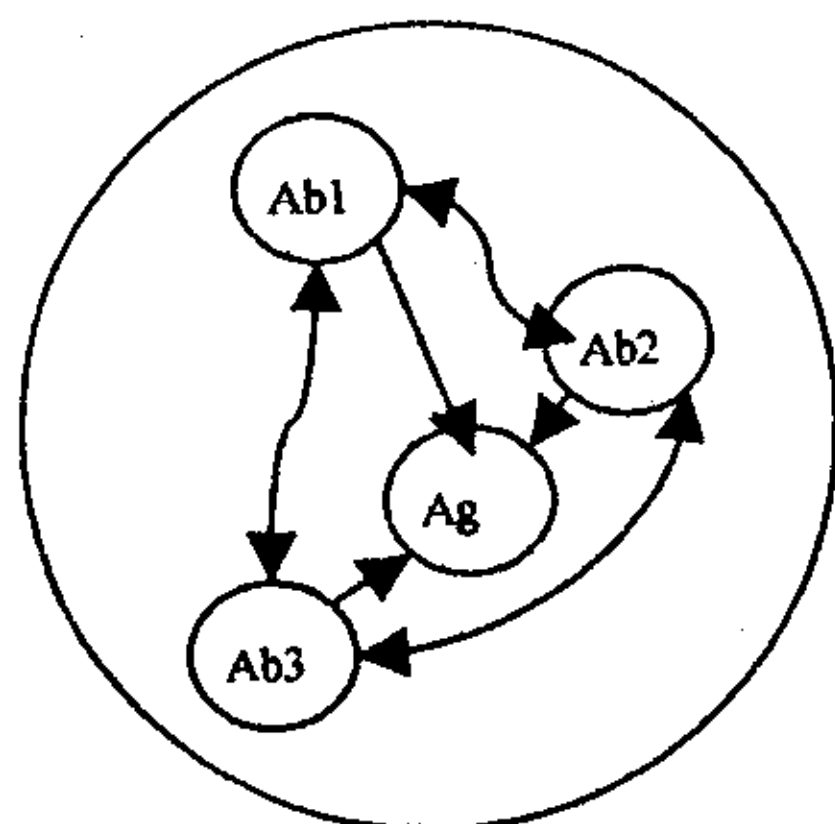
2)识别抗原(Ag):免疫系统确认抗原入侵。将目标用户  $u_4$  作为抗原添加入免疫系统中。



3)产生初始抗体群(Abn):激活记忆细胞产生抗体,随即选择出来一些抗体。

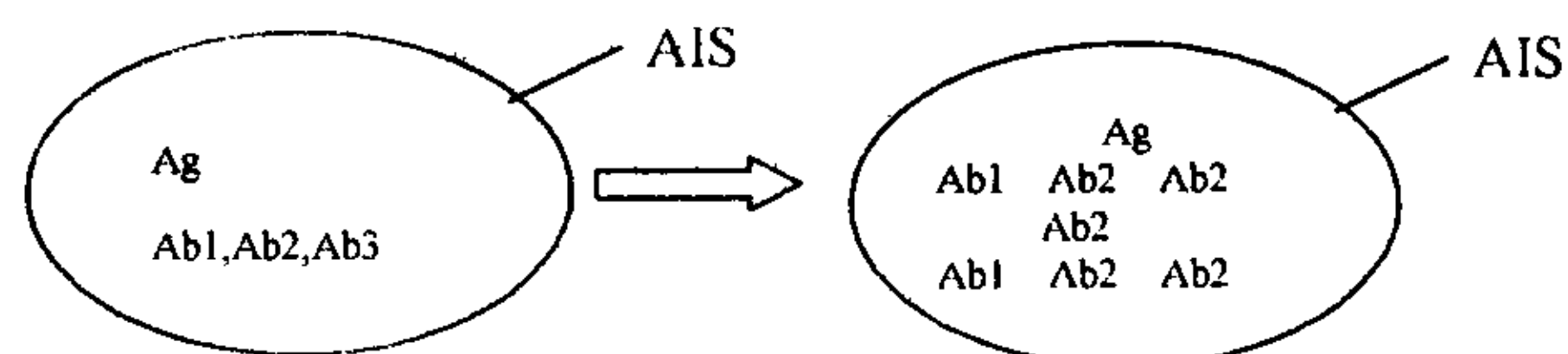


4)计算亲和力:计算抗原和抗体之间的亲和力。在此即为两者之间的相关性,抗原 Ag 与 Ab1, Ab2, Ab3 的相关性分别表示为 MS14, MS24, MS34。同时计算抗体和抗体之间的相关性,两两关系表示为: MS12 = CorrelCoef ( Ab1, Ab2 ); MS13 = CorrelCoef ( Ab1, Ab3 ); MS23 = CorrelCoef ( Ab2, Ab3 )。



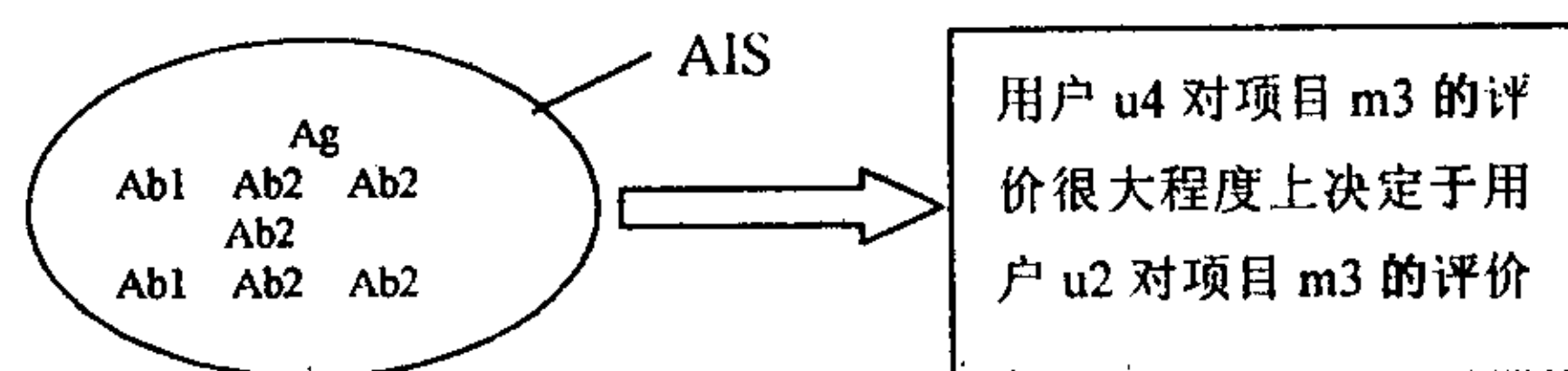
5)记忆细胞分化:将与抗原有最大亲和力的抗体加给记忆细胞。由于记忆细胞数目有限,新产生的与抗原具有更高亲和力的抗体替换较低亲和力的抗体。

6)抗体的促进和抑制:与抗原亲和力高的抗体受到促进,而与其他抗体相关性高的受到抑制。即生物免疫系统当中存在的一种特殊的浓度调节机制。简言之,在克隆扩增过程中,高浓度的抗体细胞会受到一定程度的抑制作用,而低浓度的抗体细胞则会受到促进。浓度调节机制有利于维持免疫系统群体的多样性,避免早熟现象,利于在全局范围内进行最优个体的搜索。



7)抗体产生:当系统稳定后,得到抗体以及抗体浓度。

8)在抗体以及抗体浓度的基础上,使用协作过滤技术给出推荐。



### 3 实验

从免疫系统中选择 100 人作为抗体,这些抗体与抗原(目标用户)有相似的好感,目标用户需要从推荐系统中获得对某影片的好感度(以影片系统为例),CF 使用公式(3)来计算预测推荐。

$$p = \frac{\sum_{i=1}^{100} (W_i * V_i)}{\sum_{i=1}^{100} (W_i * 1)} \tag{3}$$

其中  $p$  表示预测推荐值;  $W_i$  表示  $i$  抗体的权重,如果  $i$  抗体已对该影片投票则  $W_i$  为  $i$  抗体在系统中的浓度,如果  $i$  抗体未对该影片投票则  $W_i$  为零;  $V_i$  表示  $i$  抗体对该影片投票值(总  $\leq 1$ )。这里使用抗体浓度来作为抗体的权重,原因是既包含了抗体对抗原的相互关系又包含了抗体对其它抗体的相互关系。

实验使用的数据来自于公开数据库<sup>[7]</sup>,数据库中包含有 72 916 位用户对 1 628 部影片一共 2 811 983 张投票。

从数据库中随机抽出 100 个用户(是对 20 部以上的影片投过票的),隐藏掉每个用户其中一张投票,使用余下的信息来得到预测推荐。对每个用户重复 20 次,每次隐藏不同的投票,比较这 20 次预测推荐值与

实际投票值。使用公式(4) 来计算预测精确度。

$$PA = 1 - \frac{\sum_{i=1}^{20} |p - aV|}{20} \quad (4)$$

其中 PA 表示预测精确度; p 表示预测推荐值; aV 表示实际投票值。

图 1 是这 100 个用户使用传统的推荐系统获得的推荐预测精确度,平均预测精确度为 0.523 462。图 2 是这 100 个用户使用改进后的影片推荐系统得到的预测精确度,平均预测精确度为 0.807 613。结果可以看出预测精确度明显提高。

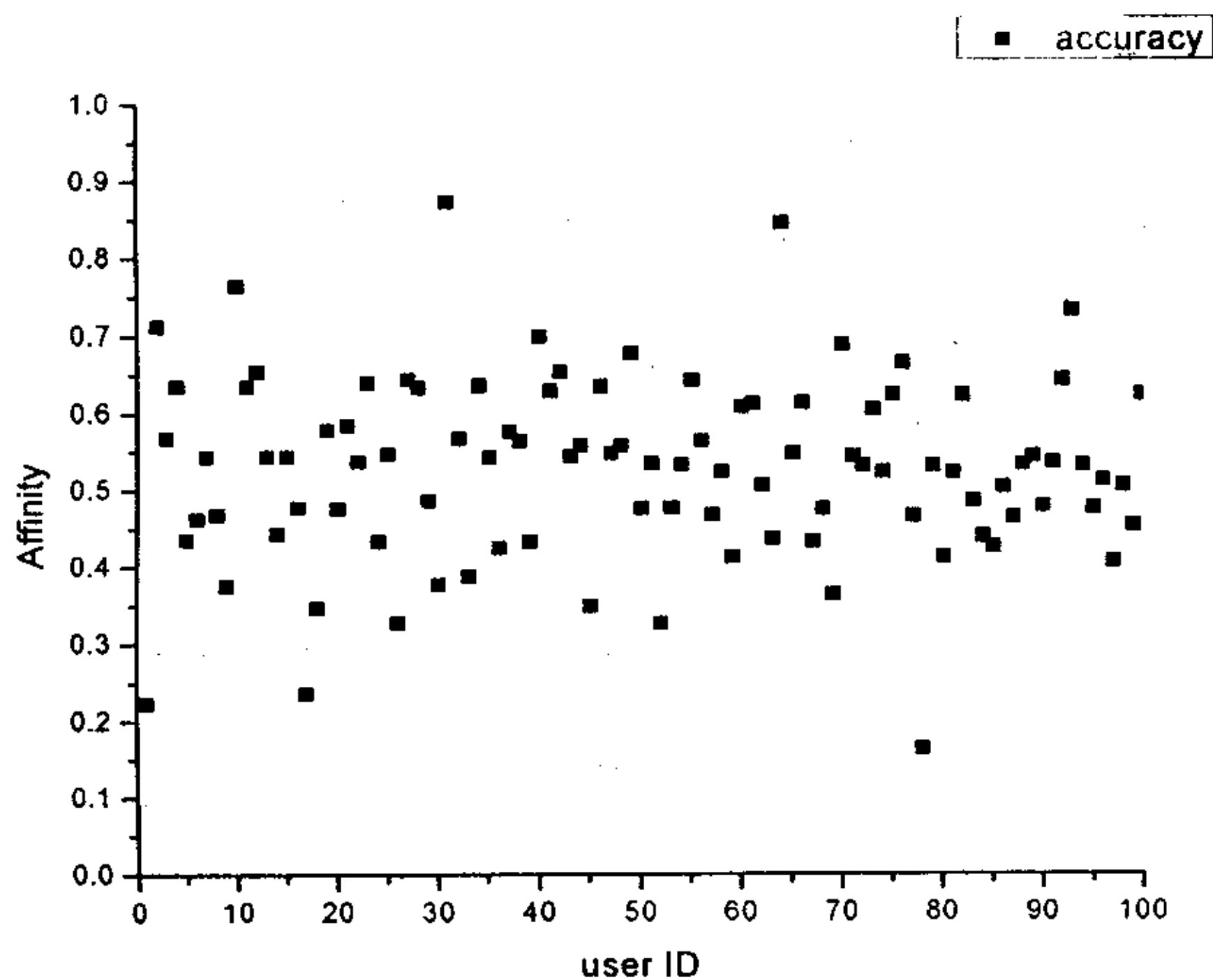


图 1 改进前预测精确度

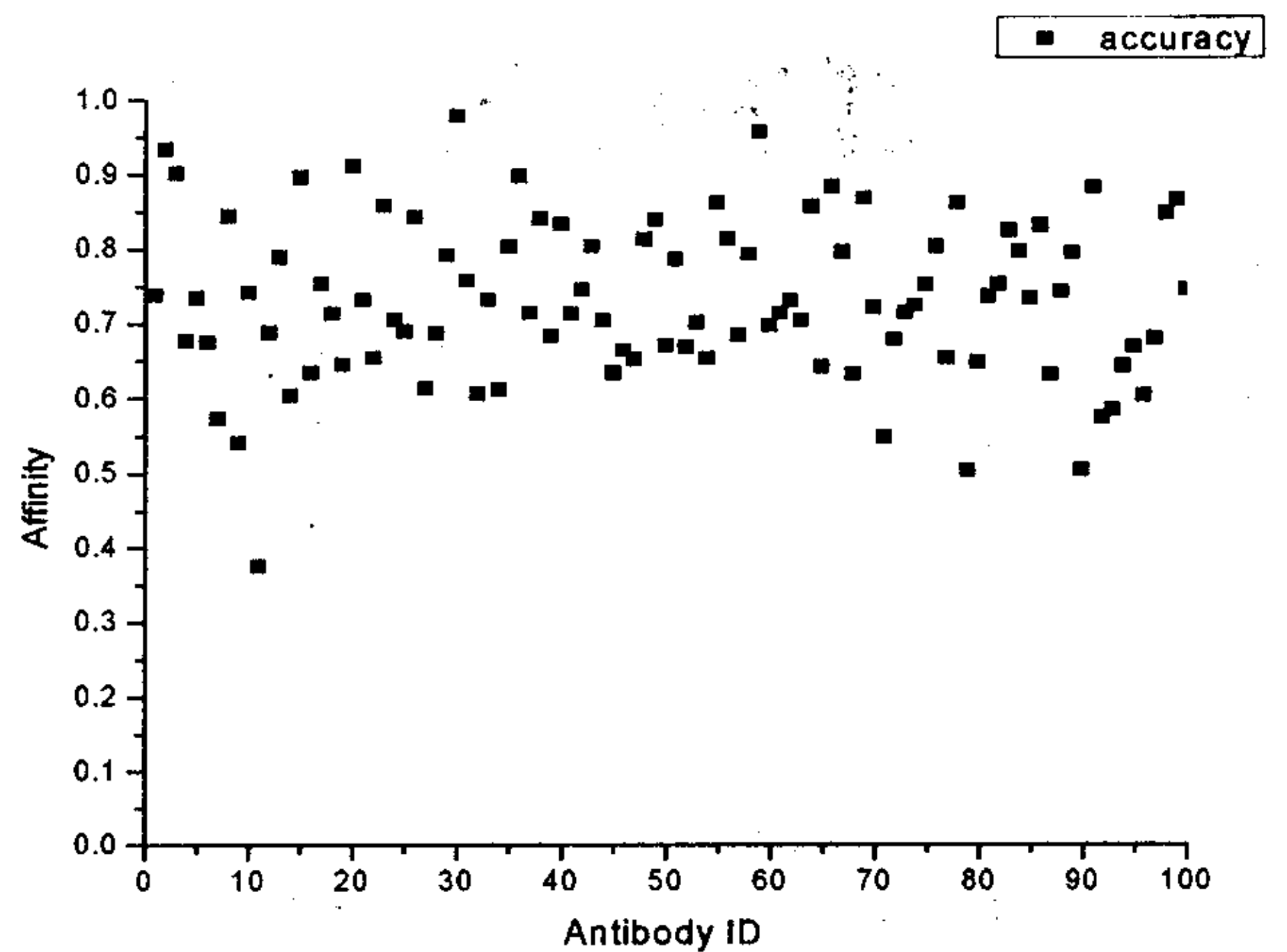


图 2 改进后预测精确度

#### 4 总 结

借鉴免疫系统的概念与理论,成功地将其思想引入传统的只使用协作过滤技术的推荐系统,提出了改进后的推荐算法。

数据实验结果表明,该算法对推荐预测精确度的提高有明显的效果。文中的实验数据是收集人们的真实数据,所以提出的算法对于其它非影片推荐系统具有普遍性意义。

(上接第 179 页)

#### 参考文献:

[1] Lemeshewsky G P. Multispectral Multisensor Image Fusion Using Wavelet Transforms[J]. Pro. SPIE, 1999, 3716:214 - 222.

[2] Toet A. Merging Thermal and Visual Images by a Contrast Pyramid Optical[J]. Engineering, 1989, 28(7): 789 - 792.

[3] 陆宏波,施惠昌. 基于小波变换的图像融合方法[J]. 微电子与基础产品,2001,27(5):54 - 57.

[4] Garzelli A. Possibilities and Limitations of the Use of Wavelets in Image Fusion:IGARSS 2002,Toronto, Canada,June 24 -

#### 参考文献:

[1] Dasgupta D. Artificial Immune Systems and Their Applications [M]. [s.l.]: Springer,1998.

[2] Billsus D,Pazzani M J. Learning collaborative information filters[C]//In Proceedings of the Fifteenth International Conference on Machine Learning. Madison, WI: Morgan Kaufman,1998:46 - 54.

[3] Jerne N K. Towards a network theory of the immune system [J]. Annals of Immunology,1973,125C:373 - 389.

[4] Perelson A S,Weisbuch G. Immunology for physicists[J]. Reviews of Modern Physics,1997,69:1219 - 1267.

[5] Farmer J D,Parkard N H,Perelson A S. The immune system, adaptation, and machine learning[J]. Physica,1986,22:187 - 204.

[6] Fisher D, Hildrum K, Hong J, et al. SWAMI: a framework for CF algorithm development and evaluation[EB/OL]. 1999. <http://guir.berkeley.edu/projects/swami/>.

[7] Compaq Systems Research Centre. EachMovie data set[EB/OL]. 1998 - 12. <http://research.compaq.com/SRC/each-movie/>.

26,2002[C]. Piscataway, NJ, USA: IEEE,2000:66 - 68.

[5] 彭玉华. 小波变换与工程应用[M]. 北京:科学出版社,2003.

[6] 李树涛,王耀南,张昌凡. 多传感器图像融合的客观评价与分析[J]. 仪器仪表学报,2002,23(6):651 - 654.

[7] Wu Jin, Tian Jinwen, Lin Jian. Multiscale Image Data Fusion Based on Local Deviation of Wavelet Transform, Intelligent Mechatronics and Automation: Proceedings 2004 International Conference Chengdu China, Chengdu, China, Aug. 26 - 31, 2004 [C]. Piscataway, NJ, USA: IEEE,2004:677 - 680.