

人工免疫在电子商务推荐中的应用^①

任雪利, 代余彪

(曲靖师范学院 计算机科学与工程学院, 曲靖 655011)

摘要: Internet 的迅速发展使得电子商务中信息超载现象越发严重, 如何从商品信息的海洋中快速准确的挑选出需要的商品已成为电子商务网站急需解决的问题, 个性化推荐为解决该问题提供了可能。为了实现准确的推荐, 本文提出了使用人工免疫技术对用户进行聚类, 然后采用协同过滤技术进行推荐, 结果表明, 准确率可以达到 80% 以上。

关键词: 人工免疫; 协同过滤; 亲和力

Application of Artificial Immune to Recommendation of Electronic Commerce

REN Xue-Li, DAI Yu-Biao

(Department of Computer Science and Engineering, Qujing Normal University, Qujing 655011, China)

Abstract: The rapid development of Internet makes it more serious the overloading phenomenon of information in electronic commerce. How to find goods required quickly and accurately from the commodity information is a problem that needs to solve. Recommendation personalized offers a solution to the problem. In order to achieve an accurate recommendation, this paper presents the use of artificial immune clustering technology to the user, and then to recommend using collaborative filtering. It shows that the accuracy rate can reach 80% in the experiment.

Keywords: artificial immunity; collaborative filtering; affinity

Internet 的迅速发展为信息量的惊人膨胀提供了的土壤。大量有用信息虽然为人们提供了更多的价值, 然而信息的泛滥也意味着为了寻找合适的信息必须付出更大的成本。事实上, 调查发现仅仅通过浏览和简单的查询来寻找有用的信息变得相当困难, 在某种程度上过量的信息意味着信息缺乏, 因此就需要某种工具来迅速找到所需要的信息来辅助决策, 防止顾客迷失。与此同时, 顾客对于产品和服务的要求越来越高, 老福特的“黑色就是所有的颜色”早已不再适用, 顾客更加期望能够得到个性化和多样化的产品和服务。推荐系统利用信息技术带来的便利, 模拟销售人员帮助客户完成购买过程, 从而使客户避免信息“超载”所带来的麻烦, 因此, 日益受到企业的关注。

1 相关知识介绍

1.1 推荐技术介绍

根据推荐实施原则的不同, 常用的推荐技术可分

为: 基于内容过滤的推荐(Content-Based filtering)和基于协同过滤的推荐(Collaborative filtering)^[1]。基于内容过滤的推荐系统通过比较项(商品)之间的相似性而不是用户之间的相似性实现推荐功能, 忽略了用户的购买行为, 只考虑了商品之间的相似性关系。该方法简单, 有效, 建模和商品间的相似性度量可以脱机进行, 因而推荐响应时间快; 缺点是难以区分商品信息的品质和风格, 而且不能为用户发现新的感兴趣的物品, 只能发现和用户已有兴趣相似的商品。基于协同过滤的技术并不分析商品之间的相似性, 而是学习目标用户和历史用户之间购买行为的相似性, 不依赖商品的特征, 根据相似历史用户的购买行为生成推荐结果, 能够为用户发现新的感兴趣的物品, 不需要考虑商品的特征, 任何形式的商品都可以推荐, 但是这种方法推荐的准确性取决于相似用户的选择。对用户的正确分类会影响推荐系统的准确性, 为了有效的识别相似用户, 本文采用人工免疫的方法实现对用户的分类, 然

^① 基金项目: 曲靖师范学院基金(2008QN007); 云南省教育厅研究课题(09C0188)

收稿时间: 2010-06-21; 收到修改稿时间: 2010-08-06

后根据分类结果,采用协同过滤技术来实现商品的推荐。

1.2 人工免疫

人工免疫系统是一种模仿自然免疫系统功能的智能化方法,它实现了一种受生物免疫系统的启发,通过学习外界物质的自然防御机理的学习技术,提供噪声忍耐、无教师学习、自组织、记忆等进化学习机理,结合了分类器、神经网络和机器推理等系统的一些优点,因此具有提供新颖的解决问题方法的潜力^[2]。其研究成果涉及到控制、数据处理、优化学习和故障诊断等许多领域,已经成为继神经网络、模糊逻辑和进化计算后人工智能的又一研究热点。在人工免疫的发展过程中,提出了很多人工免疫算法模型,例如:免疫应答原理、免疫网络原理、克隆选择算法和免疫学习原理等。本文采用免疫网络学说实现用户的分类。

免疫网络学说^[3-5]认为:不仅在抗原的表位和抗体的对位间存在刺激和抑制关系,而且各抗体的对位和独特位之间也存在刺激和抑制关系,这就形成一种大规模网络系统。具体说,各抗体在人体内散乱浮游,它们之间通过相互通讯达到识别和排除抗原目的。免疫网络理论还提出,免疫细胞对识别信号(抗原或其它免疫细胞、分子)的应答有阳性反应和阴性反应两种。前者可以产生细胞增殖、细胞激活和抗体生成;而后者则导致免疫耐受或抑制。在该模型中,抗原和抗体以及抗体与抗体之间的相互作用控制着系统,并且维持系统的自适应性和多样性。模型中抗体浓度的计算方法如下式 1 所示:

$$\frac{dx_i}{dt} = k_1 m_i x_i y - \frac{k_2}{n} \sum_{j=1}^n m_{ij} x_i y_j - k_3 x_i \quad (1)$$

其中: k_1 : 刺激率, k_2 : 抑制率, k_3 : 死亡率, y : 抗原的浓度, x_i : 抗体 i 的浓度, x_j : 抗体 j 的浓度, m_{ij} : 抗体 i 和 j 的亲合力, m_i : 抗体 i 和抗原的亲合力, n : 抗体的数量

2 基于人工免疫的推荐系统

采用协同过滤进行推荐一般包括两个过程,首先需要确定该用户所属的近邻类,然后根据近邻的评分采用 TOP-N 原则进行推荐。由于采用 TOP-N 原则进行推荐很多文献都有详细地介绍^[5],因此,本文重点

说明使用人工免疫技术进行用户聚类的方法。

2.1 编码

将数据集分为训练数据集和测试数据集,本文以用户对多本书籍的评分作为实例来说明基于人工免疫的推荐系统,以使用该推荐系统的用户作为抗原 Ag ,将作为训练集的用户作为抗体 Ab ,对抗元和抗体的编码采用相同的策略,将用户对多本书的评分作为一个集合,集合中的每一个元素为一个二元组 $\langle id, rate \rangle$,表示用户对该书 (id 为书号) 的评分 ($rate$ 为评分等级)。

用户 1: $\{ \langle 5, 4 \rangle, \langle 12, 3 \rangle, \langle 51, 4 \rangle, \langle 120, 5 \rangle, \langle 145, 1 \rangle, \langle 160, 4 \rangle, \langle 188, 2 \rangle \}$

用户 2: $\{ \langle 12, 3 \rangle, \langle 51, 3 \rangle, \langle 120, 4 \rangle, \langle 145, 2 \rangle, \langle 160, 5 \rangle, \langle 188, 1 \rangle, \langle 190, 3 \rangle, \langle 211, 2 \rangle \}$

2.2 编码

1) 对用户的评分进行归一化。从数据集中抽取 2 条用户记录,筛选出他们共同评分的书目,用户对书籍的评分分为 5 个等级,即 1, 2, 3, 4, 5, 评分值归一化就是将评分 1~5 等级划归为 0~1 之间,即归一化后的值=用户的原始评分值/5,采用该方法对用户的评分进行归一化,得到下面的数值:

用户 1: $\{ \langle 12, 0.6 \rangle, \langle 51, 0.8 \rangle, \langle 120, 1 \rangle, \langle 145, 0.2 \rangle, \langle 160, 0.8 \rangle, \langle 188, 0.4 \rangle \}$

用户 2: $\{ \langle 12, 0.6 \rangle, \langle 51, 0.6 \rangle, \langle 120, 0.8 \rangle, \langle 145, 0.4 \rangle, \langle 160, 1 \rangle, \langle 188, 0.2 \rangle \}$

2) 计算用户各等级的频度值。频度值指的是用户对相同的书籍进行评分数值的统计表。由文献[1]中计算得出的用户评分归一化值构造用户频度表的方法是:以用户 1 的评分等级为行,以用户 2 的评分等级为列构造表格,表中的数值 f_{ij} 是用户在进行评分时,处在用户 1 的第 i 等级和用户 2 的第 j 等级书目的数量,最后一行和最后一列的值是对用户在该等级评分书目的数量,表格中左后一个单元格的数值是用户 1 和 2 共同评分的书目数量。

表 1 用户频度表

	0.2	0.4	0.6	0.8	1	
0.2	0 (f11)	1 (f12)	0 (f13)	0 (f14)	0 (f15)	1
0.4	1 (f21)	0 (f22)	0 (f23)	0 (f24)	0 (f25)	1
0.6	0 (f31)	0 (f32)	1 (f33)	0 (f34)	0 (f35)	1
0.8	0 (f41)	0 (f42)	1 (f43)	0 (f44)	1 (f45)	2
1	0 (f51)	0 (f52)	0 (f53)	1 (f54)	0 (f55)	1
	1	1	2	1	1	6

3) 计算亲和力。使用公式(2)计算亲和力

$$WK = \frac{Po(w) - Pe(w)}{1 - Pe(w)} \quad (2)$$

其中, WK: 亲和力; Po(w): 观测频度; Pe(w): 随机频度由于所有的数据均是根据真实数据进行, 所以没有随机频度, 因此, 本文取 Pe(w)为 0, 上式简化为 WK=Po(w), 可用公式(3)对 Po(w)进行计算

$$Po(w) = \frac{1}{n} \sum_{i=1}^g \sum_{j=1}^g w_{ij} f_{ij} \quad (3)$$

其中, g : 评分等级数; n: 共同评分书目数; fij : 频度, 即上表中的数值; wij: 权重(表中第 i 行 j 列数值的权值)

对于每个频度的权重, 使用表达式(4)进行计算:

$$w_{ij} = 1 - \frac{|i - j|}{g - 1} \quad (4)$$

在此, 可计算出 f21, f12, f33, f43, f54, f45 的权值分别为 0.75,0.75,1, 0.75,0.75,0.75,因此, 可得亲和力 WK(u1,u2)=1/6*(1*0.75+1*0.75+1*1+1*0.75+1*0.75+1*0.75)=0.792 即用户 1 和用户 2 的亲和力为 0.792, 说明用户 1 和用户 2 有较好的一致性。

2.3 算法主要流程

免疫网络的用户聚类步骤如下:

- ① 抗原编码
- ② 当 AIS 不稳定 & 有候选抗体 时;
- ③ 选择候选抗体;
- ④ 计算抗体与抗原及抗体与抗体的亲和力;
- ⑤ 删除亲和力小于阈值的抗体, 并产生新的抗体

进行补充;

⑥ 重复(3)和(4), 直到抗体的数量在多次迭代中保持不变为止;

⑦ 结束。

2.4 准确度

对系统的推荐结果使用准确度进行度量, 准确度的定义见式(5)。

$$\text{准确度} = 1 - \frac{\sum (\text{except} - \text{rate})}{n} \quad (5)$$

其中, except: 预测评分,rate: 实际评分,n 为预测的总次数。

3 实验结果

在本次实验中, 使用上面的方法对用户进行聚类, 通过收集整理卓越网、China-Pub 等多个电子商务网站上用户对书籍的评分情况, 然后选取 100 个用户的评分作为训练的实验数据, 对每一位用户留出一本书的评分最为测试, 在计算抗体浓度时, k1,k2,k3 分别取为 0.3,0.2 和 0.1, n 的值为 100, 阈值设为 0.7。实验基于 Windows XP 操作系统, AMD Sempron(tm) Processor 3400+ 1.80GHZ CPU,1GB 内存, 使用 Java 进行编码, 数据库使用 Microsoft Access 2003。以式 5 对预测结果进行度量, 实验结果表明, 推荐的准确度可以达到 80% 以上。该推荐系统是以推荐的对象为主体进行推荐的, 只考虑了对推荐的对象评分, 没有考虑影响用户评分的一些实际情况, 例如专业, 兴趣爱好等这些都可能影响用户的评分, 因此, 以用户为主体, 进行更准确的个性化推荐是下一步的研究方向。

参考文献

- 1 Pitsilis G, Marshall L. Trust as a key to improving Recommendation Systems. 2010,3-4.
- 2 Aickelin U, Cayzer S. The danger theory and its application to artificial immune systems. Artificial Immune Systems, 2002, 141-148.
- 3 Wang W, Gao SC, Li FJ, Tang Z. A Complex Artificial Immune System and Its Immunity. IJCSNS International Journal of Computer Science and Network Security, 2008,10-11.
- 4 Cayzer S, Aickelin U. A Recommender System based on the Immune Network, CEC2002, 807-813.
- 5 张光卫,李德毅,李鹏,康建初,陈桂生.基于云模型的协同过滤推荐算法.软件学报, 2007,18(10):2403-2411.