自适应推荐算法在电子超市个性化服务系统中的应用研究

罗奇[1]，余英[2]，赵呈领[2]，曹艳[2]

(1. 武汉科技大学中南分校 信息工程学院，湖北 武汉 430223；2. 华中师范大学 信息技术系，湖北 武汉 430079)

摘 要：为了满足电子超市中用户的个性化的服务需求，提出并实现了一种基于支持向量机的自适应推荐算法。首先，将用户模型按照层次化方式组织成领域信息和原子需求信息，考虑多用户同类信息需求。采用支持向量机对领域信息节点中的原子需求信息进行分类协同推荐，然后在针对每一领域信息节点中的原子信息需求进行基于内容的过滤。该算法克服了分别采用协同推荐和基于内容的推荐单一方法的缺点，大大提高了信息的查准率和查全率，尤其适合大规模用户群的信息推荐。该算法用于基于电子超市的个性化推荐服务系统(PRSSES)中，结果表明是有效的。

关键词：电子超市；个性化推荐；支持向量机；查全率；查准率

中图分类号：TP302 **文献标识码**：A **文章编号**：1000-436X(2006)11-0183-04

Research on personalized service system in E-supermarket by using adaptive recommendation algorithm

LUO Qi[1], YU Ying[2], ZHAO Cheng-ling[2], CAO Yan[2]

(1. Dept. of Information Engineering, Wuhan University of Science and Technology Zhongnan Branch, Wuhan 430223, China;
2. Department of Information & Technology, Central China Normal University, Wuhan 430079, China)

Abstract: To meet the personalized needs of customers in E-supermarket, a new adaptive recommendation algorithm based on support vector machine was proposed. Firstly, user profile was organized hierarchically into field information and atomic information needs, considering similar information needs in the group users. Support vector machine (SVM) was adopted for collaborative recommendation in classification mode, and then vector space model (VSM) was used for content-based recommendation according to atomic information needs. The algorithm had overcome the demerit of using collaborative or content-based recommendation solely, which improved the precision and recall in a large degree. It also fit for large-scale group recommendation. The algorithm was used in personalized recommendation service system based on E-supermarket (PRSSES). The system could support E-commence better. The results manifested that the algorithm was effective.

Key words: E- supermarket; personalized recommendation; support vector machine; precision; recall

1 引言

随着电子商务和网络的不断发展，越来越多的商业企业转向电子商务的经营模式[1]。电子商务经营模式可大大节省物理环境下的所需要的成本，为顾客带来了方便，日益受到人们的关注。因此，不少商业企业都建立了电子超市网站，为顾客提供商品和信息的服务。但是在实际应用中，却只有

收稿日期：2006-07-12；**修回日期**：2006-10-25

基金项目：湖北省高等学校教学研究项目（20050185）

Foundation Item: Hubei Province College Teaching Research Project (20050185)

2%~4%的访问者购买商品，很难吸引顾客的主动参与[2]。经调查研究表明，这主要是商品选购的个性化推荐系统还不完善，提供的信息和商品都不能准确的满足用户的需求。如果电子超市网站想吸引更多的访问者成为自己的客户、提高已有客户的忠诚度、增强自己的交叉销售能力，就必须具有个性化的设计。也就是说提供的商品和信息必须根据用户的需求，而个性化的设计的关键在于如何根据用户的兴趣进行个性化的推荐。

目前，国内外也有不少学者对个性化推荐算法进行了大量的研究例如传统的协同推荐算法和基于内容的推荐算法[3]。虽然协同推荐算法能够挖掘出被推荐用户潜在的新兴趣，但是它存在着稀疏、冷开始、特殊用户等问题。同样，基于内容的推荐算法也存在着信息挖掘不全面、推荐内容有限、缺乏用户反馈等不足之处[4]。

基于此，本文在改进传统算法基础上，提出一种基于支持向量机的自适应推荐算法。首先，将用户模型按照层次化方式组织成领域信息和原子需求信息，考虑多用户同类信息需求。然后采用支持向量机对领域信息节点中的原子需求信息进行分类协同推荐，在针对每一领域信息节点中的原子信息需求采用向量空间模型进行基于内容的过滤，保证查准率。将算法应用到电子超市系统（PRSSES）中,能更好的支持电子商务的发展。

2 用户模型

一般地，考虑由若干领域信息空间 S_d 构成的全局信息空间 $S_{global}=S_{d1}\cup S_{d2}\cup\cdots S_{dn}$，个人信息需求空间 $SPIN$ 是若干领域信息 $DIS=DIS_{i1}\cup DIS_{i2}\cup\cdots DIS_{in}$ 的一个子集，而 DIS 又是 GIS 的子集,它们之间的关系如图1所示[5]。

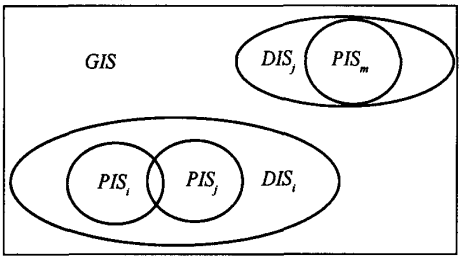

图1 用户信息结构关系

$$PIS\subseteq DIS\subseteq GIS$$

事实上，在大规模用户群中，用户信息需求存在很大的重叠，即：

$$PIS_{i1}\cap PIS_{i2}\cap\cdots PIS_{in}\neq\varnothing$$

在大规模用户群中以较高的概率成立。

通常，一个用户可能对若干领域信息感兴趣，因而层次化组织用户信息需求既符合用户的真实需求，也便于系统准确的提供推荐服务，本文按这种方式组织用户信息需求，如图2所示。

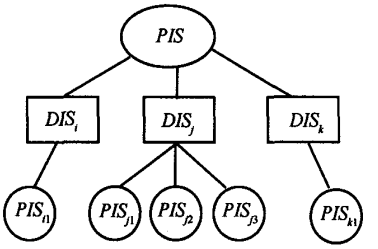

图2 层次化用户信息需求树

3 基于支持向量机的自适应推荐算法

3.1 推荐策略

将用户自定义的领域需求信息先进行分类推荐，然后再根据用户的实际原子信息需求进行基于内容的推荐。推荐分2阶段：

1) 一级过滤。通过支持向量机对领域信息需求进行分类，保证查全率。

2) 二级过滤。在一级推荐的基础上，在针对每一领域信息节点中的原子信息需求采用向量空间模型进行基于内容的过滤，保证查准率。

3.2 领域信息需求分类

支持向量机(SVM) 应用于分类，可以看作是感知器的推广[6]。在线性可分的情况下，建立一个超平面使得可分的两类数据到该平面的距离最小(通常称该平面为最优分离超平面)。对于非线性问题，它通过一个非线性映射（核函数）把原始数据从一个低维空间映射一个更高维空间（特征空间）的新数据集上，使得新数据集在该特征空间上是线性可分的，从而完成在高维空间上的分类。

通过以上分析，可以把线性可分和非线性可分问题用一个统一的式子将二者统一起来。对于样本空间 $\Omega=\{(x_i,y_i)i=1,2,\cdots,N\}\subset R^n\times\{-1,1\}$ 以及函数 $\{\Theta(x_i),\Theta(x_j)\}=K(x_i,x_j)$ 标准的支持向量机可以表示为

$$\min Q(w,\varepsilon)=\frac{1}{2}|w|^2+C\sum_{i=2}^{l}\varepsilon_i \qquad (1)$$

$$\text{s.t.}\ y_i[\{w\Theta(x_i)+b\}-1+\varepsilon_i \geqslant 0,$$

$$\varepsilon_i \geqslant 0, i=1,\cdots,l \tag{2}$$

1) 当 $K(x_i,x_j)$ 为线性变换，特别是当 $K(x_i,x_j)$ 为线性不变映射，且 C=0，$\forall\varepsilon_i=0$，式（1）、式（2）对应线性可分的情况。

2) 当 $K(x_i,x_j)$ 为将 Ω 变换到更高维 H 空间的非线性映射，且满足 Mercer 定理，即当 $K(x_i,x_j)$ 为核函数时，式（1）、式（2）对应非线性可分的情况，其中，C>0，为一常数，它控制对错分样本的惩罚程度，损失函数 $\sum_{i=1}^{l}\varepsilon_i$ 则是错分样本的一个上界。事实上 $\sum_{i=1}^{l}\varepsilon_i$ 可以表示为

$$F_\sigma(\varepsilon)\sum_{i=1}^{l}\varepsilon_i^\sigma \tag{3}$$

当 σ=1，损失函数一次，当 σ=2，其对应的是二次损失函数支持向量机。式（1）～式（3）可以归结为求解下列二次规划问题

$$W(a)=-\sum_{i=1}^{l}a_i+\frac{1}{2}\sum_{i,j=1}^{l}a_ia_jy_iy_jK(x_ix_j) \tag{4}$$

$$\text{s.t}: 0\leqslant a_i\leqslant C, i=1,\cdots,l;\ \sum_{i=1}^{l}a_iy_i=0 \tag{5}$$

其对应的最广最优分类面决策函数为

$$f(x)=\text{sgn}(\sum_{S.V.}y_ia_iK(x_ix)+b) \tag{6}$$

一个分类器本质上是一个判决函数，它将定义域空间 D 的变量 x_0 按照一定的原则划分到不相容的值域子空间 $C=\{C_1,C_2,C_3,\cdots,C_n\}$，$C_i\cap C_j=\varnothing$ 使输入的变量按照某种测试度 σ 在相似性方面与另一个确定的类最接近，即

$$f_\sigma=(x_0)=\arg\max_{c_i\in C}(c_i,x_0) \tag{7}$$

3.3　原子信息需求分类

本文采用 Salton 方法计算中的文档相似度，在 Salton 方法中，每一个文本表示成一个 n 维向量 $W=(w_1,w_2,w_3,\cdots,w_n)$，分量 w_i 表示对应特征在这篇文本中的权值，w_i 的度量如下表示[7]

$$w_i=\frac{tf_i\times\log(N/n_i)}{\sqrt{\sum_j(tf_j\times\log(N/n))^2}} \tag{8}$$

其中，tf_i 表示该特征在给定文本中出现的次数，N 是考查文本集合中文本的总数，n_i 是出现该特征的文本数。

相似度 s 则根据用户信息需求向量和每一篇要考察的文档的余弦值来计算

$$w_i=\frac{tf_i\times\log(N/n_i)}{\sqrt{\sum_j(tf_j\times\log(N/n))^2}} \tag{9}$$

对于推送结果，考虑用户的真正信息需求和潜在信息需求，将推送结果 R 分成正域结果集 R^+ 和正域结果集 R^-，分别对应实际用户需求和潜在用户需求，因而系统推荐决策如下

$$R=\begin{cases}R^+, s\geqslant s_0^+\\ R^-, s_0^-\leqslant s\leqslant s_0^+\\ \text{discard}, s<s_0^-\end{cases} \tag{10}$$

其中，s_0^+,s_0^- 表示正域和负域的推荐。

4　基于电子超市的个性化推荐系统模型（PRSSES）

电子超市的个性化推荐系统是一个游览器、Web 服务器、数据库服务器的 3 层结构，如图 3 所示。Web 服务器包括 WWW 服务器和应用服务器。

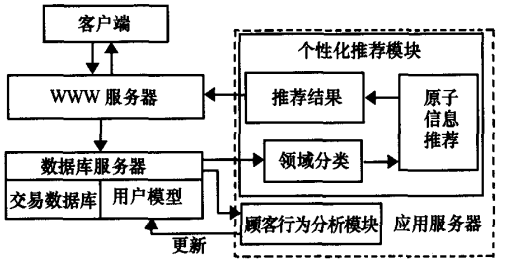

图 3　基于电子超市的个性化推荐系统模型

客户端:基于游览器的用户端,用户可输入用户名和密码后可登入公司的电子超市网站,游览网页,或者购买商品,所有信息通过 Web 服务器收集并保存在后台数据库服务器。

Web 服务器:收集用户的个人信息资料,并存储在用户模型中。同时把推荐结果生成动态网页呈现给用户。

数据库服务器:包括用户交易数据库和用户模型,其中用户交易数据库存储用户每次购买的详细记录。用户模型存储用户的个性资料,如姓名、年龄、

职业、兴趣、喜好等。

个性化推荐模块:电子超市个性推荐系统的核心部分。通过提取用户交易数据库和用户模型中的数据,为用户做出个性化的推荐。包括基于SVM的领域信息需求分类和基于VSM的原子信息需求推荐。

顾客行为分析模块:根据用户游览、购买历史、问卷调查和一些售后服务反馈意见收集用户的行为模式信息。收集到的数据经过分析模块后,然后存储在用户的信息模块,以便以后构建更新用户模型。

5 实验

在上述研究的基础上，结合与某社区的合作课题“个性化知识服务系统”的研究，笔者为某社区建设了一个提供电子超市个性化推荐系统网站，为了得到实验的对比结果，本文在个性化推荐模块中分别采用基于支持向量机的自适应推荐算法、单独的SVM分类算法和单独基于内容的VSM算法。实验数据来自电子超市中的商品，其中9603个商品作为测试集，3299个商品作为反馈评价集。选择商品最多的10类进行实验。在分类推荐阶段,采用基于LiBSVW的BSVM多分类器进行分类[8]，而在基于内容的推荐阶段相似度域值满足关系：$s_0^+ = 2s_0^-$，取s_0^+=0.3,推荐商品的种类为3，每种商品推荐的数量为5，实验结果如图4所示。

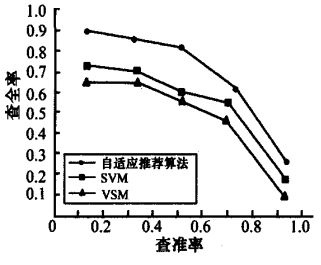

图4 自适应推荐算法、SVM和VSM算法性能比较

还将自适应推荐算法与k-NN、Rocchio进行比较，实验结果如图5所示。

从图4看出，基于支持向量机的自适应推荐算法比单独采用分类方法在信息查准率方面有所改善，而在查全率方面比单独采用基于内容的方法更加有效果。

从图5看出，基于支持向量机的自适应推荐算法比其他算法有效果，这主要由于SVM更适合处理高维复杂数据,直接从数据中自适应学习分类器所需要的参数，相比其他算法，具有更好的分类能力。

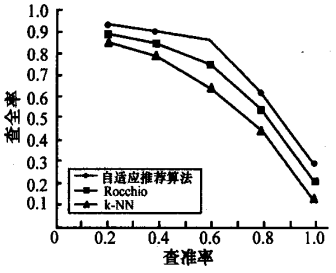

图5 自适应推荐算法、k-NN和Rocchio算法性能比较

6 结束语

本文提出了一种基于支持向量机的自适应推荐算法。将该算法应用于基于电子超市个性化推荐系统中，能更好地支持电子商务的开展。该算法已在实验中得到验证，结果表明是有效的。希望本文的工作能给相关人员有所参考。

参考文献：

[1] WU Y W. Commercial flexibility service of community based on SOA[A]. Proceedings of the Fourth Wuhan International Conference on E-Business[C]. Wuhan, 2005.467-471.

[2] YU L, LIU L.Comparison and analysis on E-commence recommendation method in China[J]. System Engineering Theory and Application, 2004, (8): 96-98.

[3] ZHANG B Q.A collaborative filtering recommendation algorithm based on domain knowledge[J]. Computer Engineering, 2005, 31(11): 29-33.

[4] DENG A L. Collaborative filtering recommendation algorithm based on item clustering[J]. Mini-Micro System,2005,25(9): 1665-1668.

[5] MA Z F. Support vector for adaptive information recommendation[J]. Mini-Micro System, 2004,25(3):385-387.

[6] VAPNIK V. The Nature of Statistical Learning Theory[M]. Beijing: Tsinghua University Press, 2000. 162-163.

[7] LI D,CAO Y D. A new weighted text filtering method[A]. International Conference on Natural Language Processing and Knowledge Engineering[C]. Wuhan, 2005.695-698.

[8] CHANG C, LIN C J. A library for support vector machines[EB/OL]. http://www.csje.nt-u.edu.tw/~cjlin/papers/2001.

（下转第192页）

计学报, 2002,9(5): 241-247。
DING Y, ZHAN H E, ZHANG T, *et al*.Networked collaborative process management system[J] .Journal of Engneering Design, 2002,9(5): 241-247.

[2] KLEIN M. Conflict resolution in cooperative design systems[J]. Computer Supported Work, 2000, 9(3-4):399-412.

[3] 史美林，向勇，杨光信. 计算机支持的协同工作理论与应用[M]. 北京: 电子工业出版社, 2000.
SHI M L, XIANG Y, YANG G X. Theory and Application of Computer Supported Cooperative Work[M]. Beijing: Publishing House of Electronics Industry. 2000.

[4] 李祥，王东哲，周雄辉等. 协同设计过程中的冲突消解系统[J]. 航空制造技术，2001, (1): 32-35.
LI X, WANG D Z, ZHOU X F, *et al*. Research on conflict resolution system for collaborative design[J]. Aeronautical Manufacturing Technology, 2001,(1):32-35.

[5] GU X, CHEN J, YANG Z.Research of network based coorperative work [J]. China Mechanical Engneering,2002,(6):481-483.

[6] 余春艳，侯宏仑，吴明晖等. 协同设计中以实体为中心的并发操作控制机制[J]. 计算机辅助设计与图形学学报,2004,16(9): 1289-1294.
YU C Y, HOU H L, WU M H, *et al*. Entity-centric concurrent manipulation control scheme for cooperative design[J]. Journal of Computer Aided Design & Computer Graphics, 2004,16(9): 1289-1294.

作者简介：

刘一良（1981-），男，山东招远人，山东师范大学硕士生，主要研究方向为计算机支持的协同设计、并发控制、计算机网络通信。

刘弘（1955-），女，山东济南人，博士，山东师范大学教授、博士生导师，主要研究方向为 CSCW、多 agent 系统、进化计算。

（上接第 186 页）

作者简介：

罗奇（1982-），男，湖北松滋人，硕士，武汉科技大学中南分校教师，主要研究方向为智能计算、数据挖掘与知识发现。

赵呈领（1956-），男，湖北武汉人，华中师范大学信息技术系教授、硕士生导师，主要研究方向为教育信息处理、电子商务、分布式计算。

余英（1975-），女，湖南常德人，华中师范大学硕士生，主要研究方向为电子商务、分布式计算等。

曹艳（1982-），女，湖南荆州人，华中师范大学硕士生，主要研究方向为计算机辅助教育、CSCL 等。