

一种基于 BP 神经网络的即时在线推荐系统

王 宁

(四川大学 计算机学院, 四川 成都 610064)

摘 要:推荐系统是一种方便用户网上购物的辅助工具。通过向用户自动推荐其感兴趣的商品,推荐系统可以节省用户的商品搜索时间并增加网站销售额。提出一种基于 BP 神经网络的即时在线电子商务推荐系统,对用户近期购物信息进行隐式收集,并通过 BP 神经网络对用户近期的购物兴趣进行分类判断,最后根据兴趣判断结果进行相关商品推荐。实验平台上的测试表明基于神经网络的电子商务在线推荐系统是可行的。

关键词:电子商务;BP 神经网络;在线推荐系统

中图分类号:TP393

文献标识码:A

文章编号:1673-629X(2009)07-0230-04

An Online Recommendation System Based on BP Network

WANG Ning

(College of Computer Science, Sichuan University, Chengdu 610064, China)

Abstract: E-commerce recommendation system is one of the tools which are used to provide customers better service. It is used to automatically provide the customer with products which they might care and need. Thus the system could help increase sales and save customers' time of product searching. An online e-commerce recommendation system based on the BP network was presented in the thesis. According to collection of users' recent shopping information and judgments of BP network, users were labeled in the classification, so that the recommendation system recommended goods the customers might concern with. The running results show that the online recommendation system based on neural network is feasible.

Key words: internet e-commerce; BP network; online recommendation system

0 引言

相对于在各个行业的广泛应用,神经网络在电子商务中的应用还处于起步阶段。电子商务中针对不同客户进行分类服务,大多基于数据库挖掘技术,国外已出现的采用神经网络进行的分类服务也集中在招标模型、市场分类等应用上。

文中对现有电子商务推荐系统所使用的推荐算法进行了研究和介绍,提出一种基于 BP 神经网络的即时在线电子商务推荐系统。通过建立相应的 BP 神经网络并应用到网上购物平台中,实现对顾客近期购物信息自动进行收集,并根据收集到的购物信息对顾客兴趣分类判断,从而向顾客推荐其感兴趣的商品。

文中所采用的推荐系统方法的特点在于:能跟踪顾客购买兴趣的阶段特征,采用隐性信息搜集方法,在线即时做出推荐,不需要用户为注册用户,不需要大

规模数据挖掘。

1 电子商务推荐系统综述

Resnick Varian 在 1997 年给出了电子商务推荐系统(Recommendation Systems)正式的定义,电子商务推荐系统是指利用电子商务网站向客户提供商品信息和建议,帮助用户决定应该购买什么产品,模拟销售人员帮助客户完成购买过程^[1]。其作用主要表现在三个方面:将电子商务网站的浏览者转变为购买者(Converting Browsers into Buyers);提高电子商务网站的交叉销售能力(Cross-Sell);提高客户对电子商务网站的忠诚度(Building Loyalty)^[2]。

在大型推荐系统中,为了产生高质量的推荐,推荐系统可能需要多种输入信息,主要包括^[3,4]:

(1)隐式浏览输入:将当前用户访问 web 站点的浏览行为作为推荐系统的输入,用户的浏览行为与访问普通 web 站点没有区别,并不知道推荐系统的存在。用户当前正在浏览的网页、用户选购的商品、用户的浏览路径等都可以作为隐式浏览输入的信息。

收稿日期:2008-10-07;修回日期:2009-02-19

作者简介:王 宁(1984-),男,四川成都人,硕士研究生,研究方向为计算机应用及网络性能分析;导师:吕光宏,博士,教授,研究方向为宽带 IP 网络 QoS 和 MPLS、网络性能分析。

(2)显式浏览输入:也是将用户的浏览行为作为推荐系统的输入,但与隐式浏览输入不同,用户通过显式输入向推荐系统提供自己的兴趣爱好。

(3)关键字输入:用户在搜索引擎中输入关键字作为推荐系统的输入。这种类型的输入不同于用户的随意浏览行为,用户输入的目的就是在web网站上搜索自己需要的信息。

(4)用户评分输入:将用户对信息的数值评分数据作为推荐系统的输入。推荐商务系统列出一系列信息让用户评分,用户的评分可以是一个数值,数值大小表示用户对信息的感兴趣程度。用户提供的评分数据使得推荐系统可以向用户提供个性化的推荐服务。

(5)用户文本评价输入:用户对自己已经知道的信息以文本形式进行个人评价,推荐系统本身并不能判断这些评价的好坏。其他用户浏览该信息时,可以看到该用户对信息的文本评价信息。

(6)编辑推荐输入:将领域专家对特定信息的评价作为推荐系统的输入,用户通过专家的专业介绍,可以对自己并不熟悉的信息加深认识,从而决定是否选择该信息或产品。

(7)用户购买历史输入:推荐系统将用户的购买历史作为隐式评分数据。一旦用户购买特定商品,则认为用户喜欢该商品。推荐系统根据用户的购买历史产生相应的推荐。

2 基于BP网络的推荐系统方案设计

现有的电子商务系统大多存在一些不足之处,具体表现在:

(1)需要用户显式地输入信息。

显式信息输入需要用户花费时间输入信息,无形中增加了用户的购物浏览负担^[5]。例如,在基于效用的推荐系统中,一开始就需要用户显式地输入需求信息,从而向用户推荐商品信息;在基于内容的推荐系统中,采用类似于邮件过滤技术,对用户输入的商品评价内容进行分析,从而判断商品的受欢迎程度和顾客喜欢,并据此向用户推荐合适商品^[6];在基于人口统计推荐系统中,虽然不需要用户对商品做出评价,但是需要收集用户的个人信息,这些信息有些可能会涉及隐私^[7]。

(2)没有考虑到用户购物需求的阶段性、实时性不足。

电子商务购物的顾客可以分为三类:一类是长期关注和购买某类固定商品的用户。比如运动爱好者网上浏览时会专注于运动装备信息,而很少关注其他网上商品;第二类是阶段性购物的消费者,他们没有特别

关注的商品,而是根据个人需求阶段性的改变关注的商品;第三类客户既有长期关注某类商品的习惯,也有与第二类顾客相似的购买需求。

对于第二类和第三类顾客,依靠挖掘历史记录和以前的商品评价,无法在线的根据用户的兴趣转移而及时进行商品推荐^[8]。推荐系统要适时向顾客推荐他们所关注的商品,就必须具有较强的在线实时性。

因此文中提出根据客户近期商品的关注点,电子商务推荐系统在线进行判断,即时向顾客推荐商品。

文中所采用的推荐系统的工作流程图如图1所示。

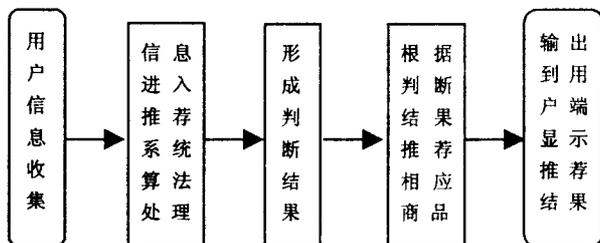


图1 电子商务推荐系统流程图

用户在网上购物浏览时,总是会浏览他们所感兴趣或需要的商品,这些点击浏览信息具有判断用户现在购物喜好和需求的价值。因此在用户信息收集时,选取用户最近网上点击浏览的商品记录作为推荐系统的输入。

推荐系统的算法处理部分,文中将所有商品进行归类划分,通过收集到的用户对不同类别商品的浏览次数信息对顾客进行在线分类判断,并根据判断结果自动向顾客推荐10项商品。为了便于测试和试验,文中将商品粗略划分为6大类:

- ①装饰饰品类商品;
- ②IT数码类商品;
- ③服装用品类商品;
- ④体育用品类商品;
- ⑤野营旅游用品类商品;
- ⑥美容用品类商品。

将用户划分为三个用户群:

(1)第一类顾客:追求时尚的年轻人群。

这类顾客的特点是关心商品①②,可能关注商品③④。当推荐系统判断用户近期喜好和该类用户的喜好相似时,推荐的10项商品为①②类商品和部分③④类商品。

(2)第二类顾客:喜爱运动和户外活动的人群。

此类顾客关心商品④⑤,较关心②③。推荐的10项商品为④⑤类商品和部分②③类商品。

(3)第三类顾客:注重仪表穿着的白领消费人群。

此类顾客关心③⑥美容用品类商品,可能关注⑤野营旅游用品类商品,偶尔会关注②IT 数码类商品。

所推荐的 10 项商品并不完全是推荐顾客点击最多的商品类的相关商品,而是将该顾客可能感兴趣的所有类别商品都进行适当的推荐,可以发掘出顾客潜在的购买需求。文中采用 BP 神经网络模型作为推荐系统的算法部分。

BP (back propagation)学习算法,是根据训练集自动进行反向学习的一种神经网络模型,分为输入层、一个或多个隐层和输出层三个神经网络层次。选择确定好合适的训练集后,训练集将对神经网络每一层的神经元参数进行修正,由于修正的顺序是从最后一层到第一层反方向传播进行的,故称为反向传播 (back propagation)学习算法^[9]。

当样本提供给网络后,实际的输出与要求的输出之间产生的差值用于对权值进行修正。对于具有单个输入值的输出节点,实际输出 y 与目标输出 t 间的误差由式(1)给出:

$$\delta = t - y \tag{1}$$

输出节点的信号是 x ,需要调整的为:

$$\Delta W = \eta \delta x \tag{2}$$

其中, η 是学习率。

新的权值为调整值与旧权值之和:

$$W_{\text{new}} = W + \Delta W \tag{3}$$

训练开始时,权值是很小的随机值。随着训练的进行,每个样本依次传送给网络并不断地修正权值,直至对任意的输入样本,差值都在允许的误差范围之内。

3 训练结果和网上购物推荐测试

3.1 BP 网络的建立

建立 BP 神经网络前首先要确定整个网络的结构:确定网络的节点层数和每层的节点数目。对此现在还没有成熟的理论,确定网络结构常常需要依靠经验和试验。

具有单个隐层的网络能表示形如

$$y = f(x) \tag{4}$$

的任意连续函数映射,而具有多个隐层的网络能表示更复杂的映射。

在建立网络时可以对不同的网络结构进行实验,最后选择一个最好的网络结构作为解决问题的神经网络模型。文中选择了 4 种神经网络结构作为候选实验网络:神经网络 1(6 节点输入层+4 节点第一隐层+3 节点输出层)、神经网络 2(6 节点输入层+7 节点第一隐层+3 节点输出层)、神经网络 3(6 节点输入层+4 节点第一隐层+4 节点第二隐层+3 节点输出层)和

神经网络 4(6 节点输入层+5 节点第一隐层+4 节点第二隐层+3 节点输出层)。

3.2 训练的结果

BP 网络的训练搭建环境为 Visual C++ ,共选用了 36 个训练样本组成训练集,前 3 个 BP 网络的训练周期为 100 个周期,第 4 个 BP 网络为 1000 个周期。通过训练和调整学习步长,4 个 BP 神经网络都收敛达到了训练要求。比较而言,选择网络结构最为精简的 BP 网络 1 作为在线电子商务推荐系统的判断处理模块。

神经网络 1 对训练集具体学习训练的结果,以及训练得到的各层神经网络的各个节点间的权值分别如图 2 所示。

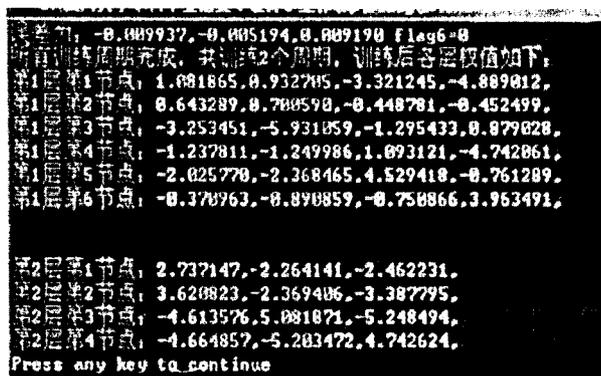


图 2 神经网络 1 训练结果

3.3 演示平台与推荐结果

测试平台操作系统为 Windows 2000 Advanced Server;动态网页编程工具采用 Visual Studio 2005。

推荐系统所需信息输入数据的选择有两个方案:一是采用用户最近 50 次购物浏览记录;另一种则是收集最近 10 次的购物浏览记录作为推荐系统的信息输入。前者可以更好判断用户的长期购物喜好,后者对用户的喜好变化响应更快。在实际试验时,考虑到推荐系统实时性是主要设计要求,采用了实时性更佳的第二种方案。图 3 是演示平台的部分截图。

4 结束语

文中提出一种基于 BP 神经网络的电子商务推荐系统,具有隐性收集信息,即时在线推荐,不需要用户注册,阶段性跟踪顾客兴趣等特点。在实验平台的测试结果表明使用该方案进行即时在线推荐是完全可行的。

进一步的工作可以考虑增加推荐模型的自适应性^[10]以及提高推荐效率。将模糊技术与 BP 神经网络相结合,对 BP 网络的输出不再是单纯的选取最大的结果,而是根据每个输出结果的隶属度大小分配相应

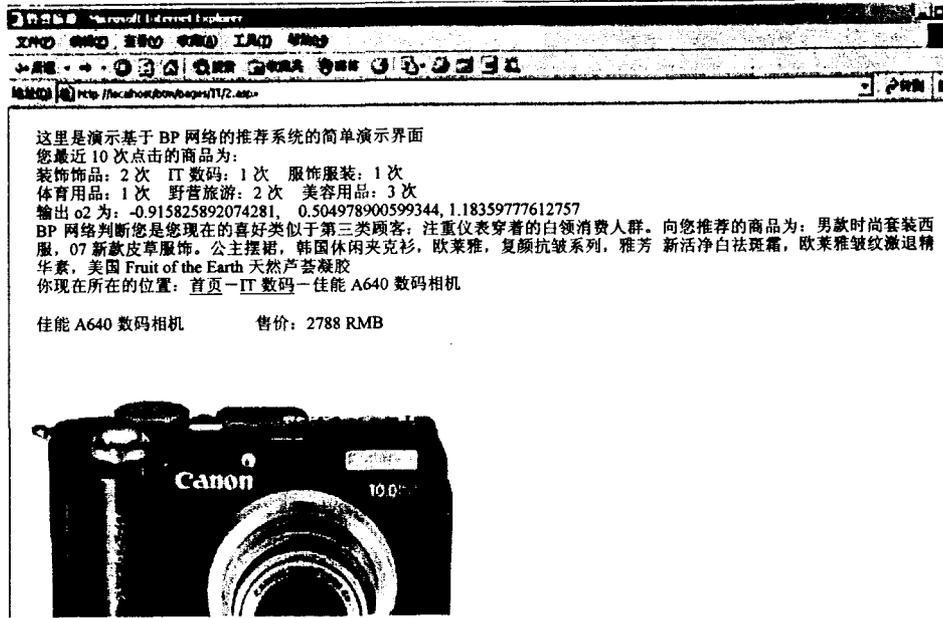


图 3 演示平台:具体商品页面——IT 数码商品

的推荐商品数目和品种,使推荐系统的命中率进一步的提高。

参考文献:

[1] Mobasher B, Dai H, Luo T. Discovery of Aggregate Usage Profiles for Web Personalization[C]//In Proceedings of the Web Mining for E-Commerce Workshop(WebKDD'2000). [s.l.]:[s.n.],2000:584-595.

[2] 王方华. 市场营销学[M]. 上海:复旦大学出版社,2001.

[3] Schafer J B, Konstan J, Riedl J. E-Commerce Recommendations Applications[J]. Data Mining and Knowledge Discovery,2001,5:115-153.

[4] Schafer J B, Konstan J, Riedl J. Recommender System in E-Commerce[C]//Proceedings of the first ACM conference on Electronic Commerce[s.l.]:[s.n.],1999:158-166.

[5] Resnick P, Varian H R. Recommendation Systems[J]. Communications of ACM,1997,40(3):56-58.

[6] Nasraoui O, Krishnapuram R. A New Evolutionary Approach to Web Usage and Context Sensitive Associations Mining[J]. International Journal on Computational Intelligence and Applications - Special Issue on Internet Intelligent Systems,2002,2(3):339-348.

[7] 张富国,徐升华. 推荐系统安全问题及技术研究综述[J]. 计算机应用研究,2008(3):656-659.

[8] 林霜梅,汪更生,陈弈秋. 个性化推荐系统中的用户建模及特征选择[J]. 计算机工程,2007(17):196-198.

[9] 高 隽. 人工神经网络原理及仿真实例[M]. 北京:机械工业出版社,2003.

[10] 刘平峰,聂规划,陈冬林. 电子商务推荐系统中推荐策略的自适应性[J]. 计算机工程与应用,2007(4):23-25.

(上接第 229 页)

[2] 杨胜文,史美林. 一种支持 QoS 约束的 Web 服务发现模型[J]. 计算机学报,2005,28(4):589-594.

[3] 刘 渡,张立臣. Web QoS 及其保证方法[J]. 计算机技术与发展,2006,16(7):32-34.

[4] 刘克非,王 红,许作萍. 一种基于服务质量预测的 Web 服务选择方法[J]. 计算机技术与发展,2007,17(8):103-109.

[5] Alemeida J, Dabu M, Manikutty A, et al. Providing differentiated level of service in Web content hosting[C]// In: Proc 1998 SIGMETRICS Workshop on Internet Server Performance. Piscataway, NJ: IEEE Press,1988:243-254.

[6] Shan Zhiguang, Lin Chuang, Wei Yaya. Prototype Implementation and Performance Evaluation of a QoS-based Web

Server[C]//In Proceedings of the IEEE International Workshop on Service-oriented System Engineering 2005 (SOSE 2005). Beijing, China:[s.n.], 2005:191-196.

[7] HP Labs. Web2K: Bringing QoS to Web Servers. HPL-2000-61, HP Labs Technical Reports[EB/OL]. 2000. http://www.hpl.hp.com/techreports/2000/HPL-2000-61.pdf.

[8] WebSphere Service Registry and Repository Handbook. IBM Redbooks publication [EB/OL]. 2007. http://www.redbooks.ibm.com/redbooks/pdfs/sg247386.pdf.

[9] 赵海萍,赵东风. 基于 Web2K 服务器的 QoS 调度策略分析[J]. 计算机工程与应用,2005(13):125-128.

[10] Nagel C, Evjen B, Glynn J. C# 高级编程[M]. 李敏波译. 北京:清华大学出版社,2006.