

# 一种网店客户数据挖掘系统的设计与实现

文/李博

**摘要:** 本文讨论了一种以数据挖掘技术、XML技术以及数据库技术为基础的网店客户数据挖掘系统的设计和实现方法。通过该系统可以帮助网店经营企业准确的把握网络市场中客户购买行为的规律和需求变化的趋势,提高网络营销的效率。

**关键字:** 数据挖掘;XML;电子商务

DOI:10.14011/j.cnki.dzsw.2015.02.026

数据挖掘技术是数据库应用研究较为活跃的领域之一,因为其现实意义,使得数据挖掘的技术研究和应用有了很大的发展,在国内外科研领域都备受关注。特别是在电子商务领域数据挖掘技术有着巨大的使用价值。随着电子商务市场的日臻成熟,第三方电子商务平台上众多的网店之间的竞争正日益激烈,网店开展网络营销的成本日益增加。每个网店企业通过一定时间的经营都会收集到众多的客户数据,如何通过庞大的客户访问数据挖掘出潜在的客户需求的金矿,使网店经营企业快速有效的通过现有客户数据掌握市场的变化的规律和趋势,已经成为网店经营管理者急需解决的问题。本文将探讨一种网店客户数据挖掘系统的设计与实现的方法。

## 1、网店客户购买数据挖掘系统的设计

### 1.1 客户购买数据挖掘系统的实现原理

对客户数据进行挖掘的关键要点就是把现有的网店访问日志经过数据预处理转换成满足XML文件格式要求的结构化数据,然后使用DBMS工具将结构化的XML文件转换成关系型数据库,最后运用数据挖掘的方法从数据库抽取能够反映客户消费和购买行为的数据。具体工作过程如下:

#### 1) 确定数据来源(主要是顾客购买记录)

数据挖掘离不开海量的数据,每一个网店在经营一段时间以后都会记录大量的客户购买记录,这些客户购买记录当中就蕴藏着大量的信息,这些信息我们进行数据挖掘的数据来源。

#### 2) 进行数据筛选

客户购买记录中蕴藏大量的信息不可能全部为数据挖掘的对象,从数据源中提取需要的数据,将无用的数据和干扰信息从数据源中剥离,为进一步的工作做准备。

#### 3) 数据格式转换

将进一步筛选过的有挖掘价值的非格式化数据数据转换为XML格式的文件并存储。

#### 4) 导入数据库

将表示客户访问数据的XML格式文件使用数据库管理工具进一步转换成可以存储于数据库的关系型数据库文件。<sup>[1]</sup>

#### 5) 进行数据挖掘

应用数据挖掘的相关算法,对存储在关系型数据库中的数据进行分析,发现海量数据中隐藏的客户行为模式和消费趋势。

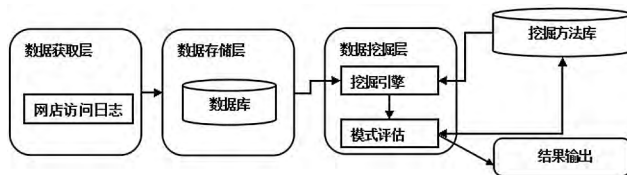
#### 6) 对数据挖掘的结果进行甄别

通过数据挖掘总结出的一些模式有时可能脱离客户真实行为模式的实际,因此在最后一步需要为对数据挖掘总结出的规则进行甄别,这个工作目前还需要有经验的专家人工完成,今

后考虑结合人工智能的技术加以解决。

### 1.2 网店客户购买数据挖掘系统模型

在上述系统设计的基础之上,我们提出了网店客户购买数据挖掘系统的模型,该模型由三个层次组成,其逻辑架构如图一所示。



图一

## 2、系统设计与实现

### 2.1 系统开发与运行环境

硬件环境: CPU Intel I3380M/RAM 2G/硬盘 320G

软件配置:

- 操作系统: Windows 7 SP1
- 开发工具: Microsoft .NET 2005/Visual C#
- 数据库管理系统: MSSQL Server 2008
- 辅助软件: SPSS Clementine 11.1; SQL Server 2008 Analysis Services (SSAS)

### 2.2 主要技术与系统实现

通过对上述对客户购买数据挖掘系统模型的分析可知,该系统主要由用户接口模块、数据清洗模块、数据格式转换模块、数据库生成模块和数据挖掘引擎模块等组成。

#### 1) 用户接口模块

本系统最终目的还是为客户的决策提供支持,因此友好的界面设计是用户与系统交互的基础。简洁而易于理解的界面有利于提高用户对系统的使用效率。

#### 2) 数据预处理模块

客户访问数据进入数据预处理模块进行清洗,去除无关的信息,剥离出对数据挖掘有价值的信息。数据预处理模块对原始访问数据进行分析,将用户购买数据记录逐条的分割成十个字段,分别为: u\_id (访问者编号), u\_date (到访日期), u\_time (到访时间), u\_orderid (订单编号), u\_product (客户购买的商品), u\_browser (使用的浏览器类型), page (首次到访页面), place (客户所在地区), payment (支付方式), logistic (物流方式), 同时删除访问数据中与以上字段不相干的数据。然后将经过预处理的数据存入中间文件。

#### 3) XML转换模块

该模块程序使用.NET的相关的方法编写,主要功能将预处理过的客户购买数据转换成标准化XML格式的数据文件进行存储。该程序的主要实现原理是对经过预处理的中间文件中的数据记录逐个分割并存入数组,然后将数组的内容按照XML的格式写入文件,完成转换。

(下转第83页)

表1 电子商务专业核心能力培养课程

目标能力	课程	所属课程类别
规划、策划能力	管理学	学科基础课程
	经济学	学科基础课程
	电子商务导论	学科基础课程
	现代营销学	学科基础课程
	网络经济学	学科基础课程
执行能力	高级程序语言设计	专业课程
	网页制作	专业课程
	网站开发与设计	专业课程
	网络营销	专业课程
	数据统计与分析	专业课程
	搜索引擎营销	岗位课程
	美工及平面设计	岗位课程
	电子商务职业证书课程	岗位课程
	校内实训	岗位课程
	网商创业实训	岗位课程
创新能力	顶岗实习	岗位课程
	移动电子商务	专业提升课程
	金融电子商务	专业提升课程
	电子商务前沿专题	专业提升课程
	电子商务创新、创意、创业	专业提升课程

每年可选派年轻教师到企业挂职锻炼,鼓励教师考取与所教学科相对应或相接近的各种职业资格证书;另一方面外聘政府主要经济部门的领导、大中型企业高管作为兼职教授,他们长期从事电子商务方面的实践工作,具有丰富的实战经验,可在电子商务专业实践教学发挥巨大的作用。

## 2、加大经费投入,加快专业实验室的建设

电子商务专业是具有很强实践应用性特征的专业,作为应用型本科院校,人才培养方案中实践课程教学的比例不断提高。在当前企业难以全部接纳学生实习实践的情况下,结合电子商务的特点,让学生在校内根据人才培养目标及所需具备的能力进行相关的课程实训或综合的实训就显得非常必要。为此,学校要加大实验、实训经费的投入,加快电子商务专业实验室的建设,这样学生能够在虚拟的环境中进行演练,进而培养学生的相关能力。

## 3、加强与政府、企业的合作,建设好电子商务人才协同培养的实习、实践基地

目前,国家和各级政府都非常重视电子商务的发展,也有相应的经费支持电子商务的人员培训;同时,越来越多的企业在开展电子商务业务,企业愿意让电子商务的学生提前介入企业相关业务,而不是毕业后才接触企业电子商务业务。在此背景下,高校应该加强与政府、企业的合作,建设真正意义上电子商务人才协同培养的实习、实践基地,让学生在学、大、大二、大三就可以接触到企业的电子商务业务,提早了解社会,这样通过四年的理论学习和实践锻炼,就能够实现毕业生与企业的无缝对接。

## 4、鼓励学生参加电子商务学科竞赛,通过竞赛提升能力

目前电子商务专业性的学科竞赛有“全国大学生电子商务网络创新应用大赛”和“全国大学生电子商务创新、创业及创业挑战赛”。学校或二级学院可组建电子商务学科竞赛指导小组,指导小组成员由骨干教师组成,针对学科竞赛项目,提前组织发动,进行校内选拔,指导教师对参与竞赛的学生进行

定期培训,并根据竞赛项目要求,可聘请相关企业指导教师,使学生真正在学科竞赛中得到锻炼,提升学生的创新能力;同时,针对学科竞赛比较好的项目,鼓励学生进行自主网络创业,让竞赛项目真正发展壮大。

## 5、加大对电子商务创业的支持力度,让学生在创业实践中得到全面锻炼

目前,应用型本科院校非常重视学生的创业,不少学校或地方政府设有创业孵化基地。为此,高校应在学生创业方面加大支持力度,通过学生创业项目进驻,并给项目进行指导,提高学生电子商务的应用能力和创新创业能力。同时,积极联系企业,引入电子商务真实项目,以电子商务创新创业教育为核心教学实训内容,实现创新创业教育与实践教学的有机结合。<sup>[1]</sup>

## 参考文献

- [1] 温希波,邢志良,张策.应用型本科电子商务专业人才培养的关键问题[J].教育与职业,2013(3):181-182
- [2] 邓之宏,邵兵家,段建.中国企业电子商务岗位人才核心技能的需求研究[J].实验技术与管理,2012(11):1360140
- [3] 李逸平,邓柳莺,郑吕燕.独立学院电子商务专业实践教学体系的构建与实现[J].电子商务,2013(1):81-83
- [4] 朱彦杰.电子商务应用型人才培养的课程体系构建[J].电子商务,2013(10):85-86

作者简介:覃艳华,电子科技大学中山学院副教授。

(上接第55页)

## 4) 数据库导入模块

利用.NET的相关方法并结合数据库管理工具建立支持数据挖掘的客户购买数据库,编写相关程序将已经转换成XML格式的客户访问数据逐条的导入到数据库并形成日志数据表方便进行后续的数据挖掘。

## 5) 数据挖掘引擎模块

数据挖掘引擎是实现客户购买数据挖掘系统的实现关键。优秀的数据挖掘算法不仅可以使数据挖掘的结果更加准确,也可以提高数据挖掘的效率。本系统主要用到的算法是该模块利用数据挖掘算法对数据进行挖掘,主要包括算法的优化、日志数据表的删除操作以及挖掘结果集的保存与删除操作等。这里主要用到的算法是K-Means算法。主要是利用该算法发现最相似的客户聚类,通过对聚类的分析来得出网店众多的顾客一般的购买行为模式,从而可以适当地调整网站营销的策略中的来提高网络营销的效果,进而增加销售量。<sup>[2]</sup>

## 3、小结

本文主要讨论了一种基于XML技术的网店客户访问数据挖掘系统的设计与实现方法。该系统可以帮助网店经营企业大量的客户访问数据进行分析,从中发现市场的变化规律和客户的行为模式,使企业的网络营销工作更有针对性,同时也更富有成效。<sup>[1]</sup>

## 参考文献:

- [1] 张光辉.XML与Web数据挖掘分析[J].中原工学院学报,2002,13(4):61-64.
- [2] 汤艳艳.数据仓库中基于面向对象技术的多维数据模型的设计与实现.计算机应用技术[J].2003,5(10):57-6

作者简介:李博,广东理工学院 讲师,研究方向:网络营销。