

doi:10.3969/j.issn.1007-7162.2012.04.002

客户价值可拓知识挖掘软件研究

朱伶俐, 李卫华, 李小妹

(广东工业大学 计算机学院, 广东 广州 510006)

摘要: 针对企业如何采取可靠的差异化营销策略以实现客户价值最大化问题, 提出了一种基于可拓学理论和可拓数据挖掘技术的客户价值可拓知识挖掘软件研究, 描述了软件的体系结构及主要功能模块的设计原理, 实现了客户价值的可拓分类知识获取. 实验结果表明本软件在数据处理、建立评价体系、计算客户价值方面具备一定的通用性和灵活性, 能够满足不同企业的需求.

关键词: 客户价值; 企业; 可拓知识; 可拓数据挖掘; 软件

中图分类号: TP31

文献标志码: A

文章编号: 1007-7162(2012)04-0007-07

Research on Extension Knowledge Mining Software for Customer Value

Zhu Ling-li, Li Wei-hua, Li Xiao-mei

(School of Computer Science, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: Aiming at the problems with enterprises' taking different reliable marketing strategies to maximize the customer value, it researched the extension knowledge mining software for customer value, based on extenics theory and extension data-mining technology, described the software architecture and the designing principles of the main function modules, thus acquiring extension classification knowledge. The experimental results show the software has general and flexible ability concerning such aspects as data processing, the establishment of the evaluation system and calculation of the customer value, so it can satisfy different enterprises' demands.

Key words: customer value; enterprise; extension knowledge; extension data mining; software

随着信息技术的快速发展, 企业的竞争环境逐渐从“以产品为中心”向“以客户为中心”转变, 企业之间的竞争也愈发激烈; 客户逐渐成为企业最重要的资源, 是企业成功至关重要的因素和利润来源, 谁把握了客户谁就能获得市场竞争优势. 趋于信息技术所呈现的变化及追求利益的最大化, 企业日益广泛地研发和应用客户关系管理 (Customer Relationship Management, CRM) 软件^[1]. 有关客户价值的理论研究和技术研究在电信、移动等领域的通讯企业, 保险、银行和证券等行业也逐渐受到关注. 如何发现潜在客户并识别重要客户, 如何度量客户价值等问题是 CRM 的首要步骤, 于是作为 CRM 核心内容之

一的客户价值分析和管理的的重要性就愈显突出^[2].

目前常见的 CRM 系统将客户价值管理归入销售管理的客户行为分析模块中, 并没有把它单独作为一个模块, 还没有对客户价值分析给以充分的重视. 同时, 现有的客户价值分析结果对于企业决策支持不够, 客户价值细分没有能够指导企业对于不同价值客户采取差异性营销措施^[3]. 从数据库技术、数据仓库和数据挖掘技术角度对海量的客户数据进行客户细分、客户流失预测等已成为企业客户价值管理和决策的科学化、智能化必然趋势. 然而, 利用现有的数据挖掘理论与方法实现的客户价值分析系统只能挖掘静态的数据、不变化条件下的知识. 如果企

收稿日期: 2012-04-11

基金项目: 广东省自然科学基金资助项目 (10151009001000044)

作者简介: 朱伶俐 (1987-), 女, 硕士研究生, 主要研究方向为智能软件.

业采取某种营销策略后会导致客户价值有什么样变化的规律知识是现有数据挖掘没有关注的,因此就不能对营销策略变化作出合理的评价.本文提出采用可拓学理论和可拓数据挖掘方法相结合的方案,从这个角度去设计与实现客户价值可拓知识挖掘软件,为企业掌握营销活动前后客户价值变化规律和采取正确的营销策略提供方便、可靠的软件工具.

1 客户价值和可拓数据挖掘

1.1 客户价值

目前对客户价值的研究主要沿着3个不同的视角展开:客户视角、企业视角和客户—企业视角^[4].站在不同的角度分析时,客户价值有不同含义,主要概括为两种:一是客户购买企业产品或服务时,客户从企业获得的价值;二是企业通过客户购买自己的产品或服务时,企业从客户获得的价值^[5].随着客户价值理论研究的深入,从客户全生命管理的角度来研究客户价值已经得到越来越多的认可,对客户价值的评价不仅要考虑客户当前价值,还要考虑其潜在价值^[6-7].在客户价值量化方面,更多的是从经济学的角度构造数学模型计算客户价值,数学模型多种多样,有些数学模型公式很复杂.

现有的客户价值分析软件,更多的是根据某企业的具体需求来设计,能够处理的客户数据格式有一定的限制性,评价指标在程序代码中固定,客户价值量化模型单一化,因此这样的软件仅适用于要求开发的企业,缺乏一定的通用性.因为不同的企业对客户价值衡量指标体系不一样,底层管理客户数据的平台不一样,客户价值度量模型不一样,如何做到客户价值软件对象的适用性,通用性是本文研究重要的出发点.

齐佳音博士认为客户价值体现在两个方面:一是客户的当前价值,二是客户潜在价值,通过当前价值和潜在价值的评价来间接评价客户价值^[6].本文在此基础上采用三层次树型结构的客户价值评价体系,根结点为客户价值,根结点下有当前价值和潜在价值两具结点;在当前价值和潜在价值结点下相应的叶结点指标.

1.2 可拓数据挖掘

客户所体现的价值和企业所期待的客户价值有时出现矛盾,可拓学研究的就是如何通过可拓变换,使矛盾问题变为不矛盾问题.可拓数据挖掘是指为了处理矛盾或不相容问题,在数据挖掘中引入可拓变换

方法,并在数据挖掘获得静态知识的基础上,通过可拓变换获取变化知识,即包含可拓变换的规则知识.因此,研究客户价值可拓知识挖掘软件,可以为企业解决各种各样客户价值矛盾问题提供策略^[8-10].

可拓数据挖掘是可拓学和数据挖掘相结合的产物,它研究用可拓学方法和数据挖掘技术,去挖掘数据库或知识库中与解决矛盾问题的可拓变换有关的知识,包括可拓分类知识、传导知识、可拓聚类知识以及其他变换知识^[9-11].它与传统数据挖掘的差异之一是:传统数据挖掘是知识的发现,而可拓数据挖掘不但可以挖掘知识,而且可以挖掘相关规则和可拓变换,从而挖掘变化的知识^[10-11].

2 客户价值可拓知识挖掘软件的设计

2.1 设计思想

站在企业的角度设计客户价值可拓知识挖掘软件时,通过将底层数据转换为统一格式数据实现屏蔽底层客户数据源异构,并对噪音、不完整等数据进行处理以确保数据的合理、规范和可靠;将源数据表中客户属性作为候选评价特征,企业可根据需要自由增删评价特征操作实现客户价值评价体系通用性;根据具体问题选择客户价值计算模型以准确、客观地评价客户价值,使软件具备一定的通用性和灵活性,可以适应不同企业的需求.

2.2 软件体系结构和总体功能模块

运用可拓学基元表示客户信息,关联函数作为客户价值计算模型,采用可拓数据挖掘方法挖掘客户价值可拓知识.客户价值可拓知识挖掘软件工作流程是:通过数据挖掘技术中的数据预处理技术去处理底层的客户数据源,得到合理、规范和可行的信息元集→动态建立客户价值评价指标体系→构建计算客户价值的关联函数模型→获得客户价值的可拓知识→可拓知识的可视化^[12-13].采用C/S结构设计,进行模块功能划分得到客户价值可拓知识挖掘软件体系结构(如图1所示)和总体功能模块(如图2所示).

Java数据库连接(Java Data Base Connectivity, JDBC)可通过在程序中动态加载数据库的驱动程序,修改相应数据库的URL、数据库名和密码连接不同的数据库;JDBC向任何关系数据库发送SQL语句,抽取一个或多个源数据表中的数据加以转换,再插入到目标数据库的数据表中,从而完成数据的各种转换功能.

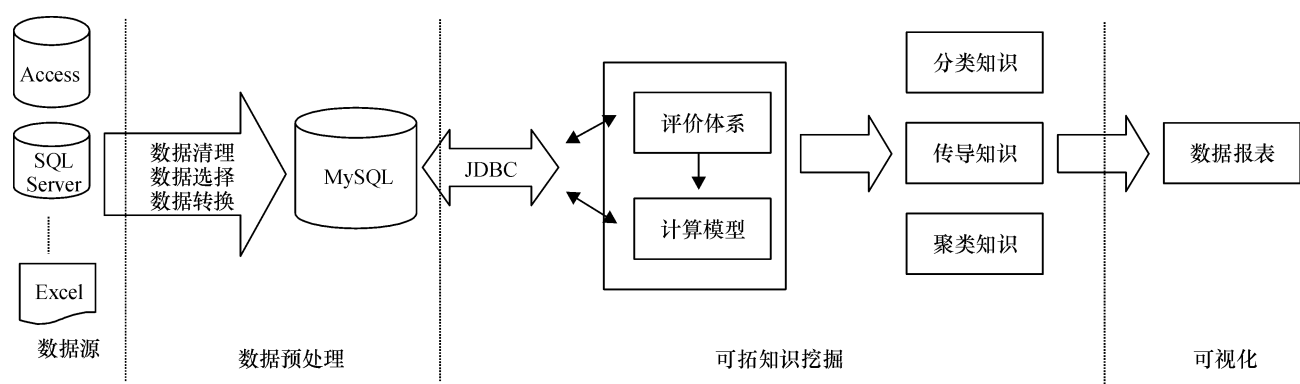


图1 客户价值可拓知识挖掘软件体系结构

Fig. 1 Extension knowledge mining software architecture for customer value

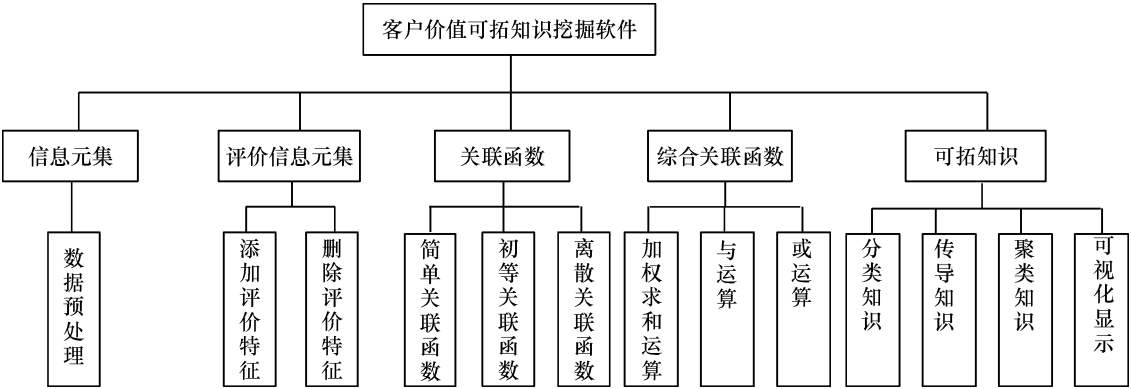


图2 客户价值可拓知识挖掘软件总体功能模块

Fig. 2 General function module figure of extension knowledge mining software for customer value

2.3 主要功能模块设计

2.3.1 获取信息元集

在分析客户价值前,需要对企业实施某营销策略变换前后的客户数据源按照一般数据挖掘的预处理过程对其进行数据清理、数据选择、数据转换等操作,将预处理后的客户数据由 MySQL 统一管理,以实现处理数据的通用性,屏蔽了企业之间的异构数据源,保证了获取的信息元集规范化、合理化和可操作性等。

根据企业选择的客户数据类型和数据库信息,通过 JDBC 技术动态地连接企业客户数据. 成功连接后,选择实施变换前后的客户数据信息表,在程序中分析表的结构、字段名和类型、数据类型的匹配,通过 SQL 语句实现数据转换并将转换后客户信息写入 MySQL 数据库表 origindata 和 change_origindata 里. 对 MySQL 中数据信息用可拓集中的基元表示为 $\{I\} = \{I_i | I_i = (O_i, C, V_i), i = 1, 2, \dots, n\}$, 其中 $C = \begin{matrix} c_1 & v_{i1} \\ c_2 & v_{i2} \\ \vdots & \vdots \\ c_m & v_{im} \end{matrix}$, $V_i = \begin{matrix} v_{i2} \\ \vdots \\ v_{im} \end{matrix}$ (I_i 对应数据库中的一条记录, c_m 为

相应记录属性, V_i 为对应属性值),至此实现了信息元集的获取^[14].

2.3.2 获取评价信息元集

由于不同企业的客户对象不同,客户价值评价指标各异,因此在软件中设计由企业根据需要选定的评价特征,以增加灵活性. 利用 JAVA 中 ResultSetMetaData 类获取 MySQL 数据库中表 origindata 的所有字段名,在界面 JList 列表控件中显示出来,作为候选评价特征. 点击列表中某项时,在弹出的窗口中选择属于哪个价值的指标,就成功地将该属性添加为评价特征;如果评价特征不在字段名集中,则设计成允许输入要添加的新评价特征名,并对其进行检验,新评价特征的量值是在某些字段属性量值基础上构造函数表达式来获取的. 新评价特征 d_p 与字段属性 c_{pi} 之间的量值函数关系表达式^[14]为: $d_p(O) = f_p(c_{p1}(O), c_{p2}(O), \dots, c_{ppp}(O))$. 因为函数关系表达式书写格式多样化,通过借鉴计算器模型实现原理使表达式形式规范化、统一化,便于计算机识别与处理函数表达式,计算新评价特征值过程简单化. 当客户价值的评价特征确定后,在后台据数据库 MySQL 中的表

evaluate 和 change_evaluate 保存了评价信息元集

$$\{D\} = \left\{ D_i \mid D_i = (I_i, C_D, U_i) = \begin{matrix} I_i & d_i, & u_{i1} \\ & d_2, & u_{i2} \\ & \vdots & \vdots \\ & d_q & u_{iq} \end{matrix}, \right.$$

$$i = 1, 2, \cdots, n \left. \right\},$$

其中, $V_i = \begin{matrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iq} \end{matrix}, C_D = \begin{matrix} d_1 \\ d_2 \\ \vdots \\ d_q \end{matrix}, D_i = \begin{matrix} D_{i1} \\ D_{i2} \\ \vdots \\ D_{iq} \end{matrix}.$

[14-16]

2.3.3 关联函数

采用可拓集中的关联函数 $k(x)$ 来定量地、客观地评价客户评价特征. 关联函数有简单关联函数、初等关联函数和离散关联函数^[9]. 不同的企业对评价特征的要求有所不同, 不同性质的评价指标适应不同的关联函数形式, 因此, 软件设计时需要体现企业主导地位, 由企业决定评价指标的关联函数模型.

在本文的关联函数模块中, 将评价特征以树型菜单形式显示在界面中, 只要选择相应菜单项, 就可建立相应评价特征的关联函数模型. 通过 getColumnTypeName 方法获取该特征的数据类型, 如果为非数值类型, 则将该评价特征定为离散关联函数; 如果为数值类型, 则通过 MIN、MAX 函数获取该评价特征的最小量值和最大量值, 并以人机交互界面询问企业是否对 [MIN, MAX] 区间满意, 满意时则该评价特征定为简单关联函数, 不满意时定为初等关联函数. 如果企业对默认的参数不满意, 则可由企业重新设定函数参数值, 使得客户价值分析更合理、更切实际. 具体如图 3 所示.

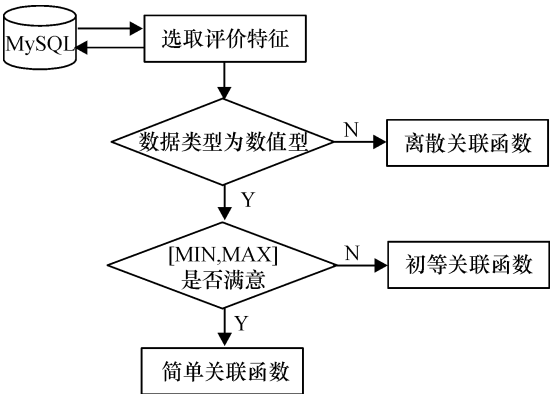


图 3 确定关联函数
Fig.3 Determine dependent function

2.3.4 综合关联函数

客户价值需要从多个评价特征角度去考察, 因此采用综合关联函数来计算客户价值及在变换下的客户价值的量变与质变过程. 客户价值由当前价值、潜在价值构成, 则需要构造客户价值的综合关联函数 K_{CV} , 当前价值的综合关联函数 K_{CCV} 和潜在价值的综合关联函数 K_{LCV} . 当认为客户 B 的每一评价特征都必须符合要求时, 综合关联函数公式为与运算:

$$K(B) = \bigwedge_{i=1}^m k_i(c_i(B)) = \bigwedge_{i=1}^m k_i(x_i),$$

其中 $\bigwedge_{i=1}^m$ 表示取 m 个值中的最小者; 当认为客户 B 至少一个评价特征符合要求时, 综合关联函数公式为或运算: $K(B) = \bigvee_{i=1}^m k_i(c_i(B)) = \bigvee_{i=1}^m k_i(x_i)$, 其中 $\bigvee_{i=1}^m$ 表示取 m 个值的最大者; 如果评价特征 $c_1, c_2, \cdots, c_m$ 权重系统为 $\lambda_1, \lambda_2, \cdots, \lambda_m$ 且满足 $\sum_{i=1}^m \lambda_i = 1$, 综合关联函数公式加权运

算: $K(B) = \sum_{i=1}^m \lambda_i k_i(c_i(B)) = \sum_{i=1}^m \lambda_i k_i(x_i)$ [9, 12, 14].

在本软件的界面设计中, 提供 3 个选项的单选按钮, 可根据企业的历史数据或领域知识来确定综合关联函数类型. 如图 4 所示.

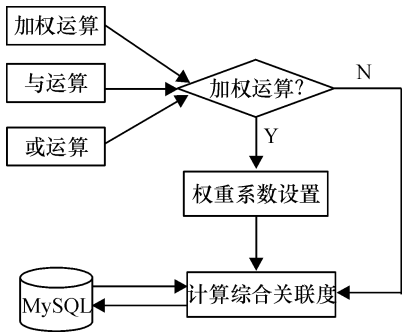


图 4 计算综合关联度
Fig.4 Calculate comprehensive dependent degree

2.3.5 获取可拓知识

根据企业的实际需要来构造当前价值、潜在价值、客户价值综合关联函数, 计算关联度, 分析客户价值的质变量变情况, 得到客户价值的某类可拓知识. 本文根据客户价值理论, 规定分类准则为: $K_{当前价值} \geq 0$ 且 $K_{潜在价值} \geq 0$ 时, 认为客户为企业的价值客户; $K_{当前价值} \geq 0$ 且 $K_{潜在价值} < 0$ 时, 认为客户为企业的次价值客户; $K_{当前价值} < 0$ 且 $K_{潜在价值} \geq 0$ 时, 认为客户为企业的潜价值客户; $K_{当前价值} < 0$ 且 $K_{潜在价值} < 0$, 则认为客户为企业的低价值客户. 通过计算变换前后关联度可以得到客户价值变化情况, 发现哪些客户发生了质变, 哪些客户发了量变, 哪些客户没有反应. 本软件可拓知识获取, 分为分类知识、传导知

识、聚类知识3种.软件分析客户价值的质变量变情况,用文字描述得到的可拓知识,用图形、报表等可视化技术呈现可拓知识,方便企业清晰、直观地了解和掌握客户的价值变化情况.

3 客户价值可拓知识挖掘软件的实现及实验结果

通过Java语言编码^[12,17-18],实现了客户价值可拓知识挖掘软件的基本功能.在本文中以某服装厂的服装销售数据为测试实例,对软件进行测试,进一步演示软件具有的功能特性,实现该服装厂的客户价值可拓分类知识获取.客户信息有客户名称、会员号、性别、年龄、职业、支付额度、购买数量、品牌忠诚度属性.

第1步:数据预处理.将服装厂变换前后服装销售数据进行数据预处理,保存到MySQL数据库中.客户信息经过数据预处理模块后,进入如图5所示的客户价值可拓知识挖掘软件主界面.



图5 客户价值可拓知识挖掘软件主界面

Fig.5 Main interface of extension knowledge mining software for customer value

第2步:构建评价体系.点击主界面“评价体系”子菜单“评价特征菜单”,在弹出图6界面中构建客户价值评价体系.

点击界面上“新特征”按钮后,在弹出界面图7中输入新评价特征名,及其与客户属性项的函数关系.通过设置弹出对话框捕获服装厂需求,确定服装厂客户价值评价体系.

第3步:确定关联函数模型,得到量化评价特征关联度值.在界面的文本框中输入简单或初等关联函数的参数值,离散关联函数等级排序及确定关联度为0的等级(见图8),根据等级排序来确定其它等级的关联度值.

$$k_1(x_{1i}) = \frac{x_{1i} - 300}{750 - 180} = \frac{x_{1i} - 300}{570} \text{ (单位产品支付 = 支付额度/购买数量)},$$

$$k_2(x_{2i}) = \frac{x_{2i} - 3}{10 - 1} = \frac{x_{2i} - 3}{9} \text{ (购买数量)},$$

1, x_{3i} = 高,
0.5, x_{3i} = 较高,
 $k_3(x_{3i}) = 0$, x_{3i} = 一般(品牌忠诚度),
-0.5, x_{3i} = 较差,
-1, x_{3i} = 差.

第4步:计算客户价值、当前价值、潜在价值综合关联度.选择图9相应的菜单项,输入评价特征的权重值,构造综合关联函数公式.



图6 构建评价体系界面

Fig.6 Interface of constructing evaluation system



图7 添加新评价特征界面

Fig.7 Interface of adding a new feature

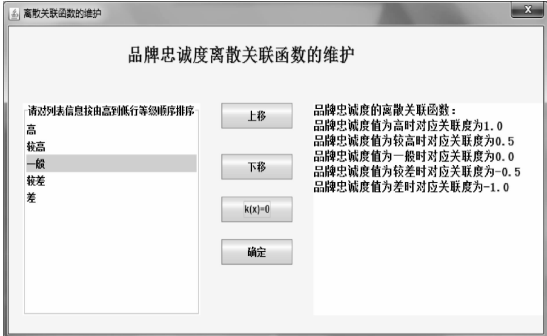


图8 离散关联函数界面

Fig.8 Interface of discrete dependent function

最后,获得可拓分类知识并可视化显示如图 10. 从输出结果可以发现次价值客户和潜在价值客户数量明显增加,说明这种变换策略对于提升次价值和潜在价值客户有良好的作用. 还可以清晰地获得每个客户的价值变换情况,变换前后客户价值比较. 通过发现的可拓分类知识,得到变换前后客户价值变化规律,可以衡量营销策略变换的作用,这是现有的一般数据挖掘软件没有研究或不关注的知识. 这样的可拓知识可为企业解决客户价值矛盾问题,提升客户价值,采取有效的营销策略提供好的参照依据.



图 9 计算综合关联度界面
Fig. 9 Interface of computing the comprehensive dependent degree



图 10 可拓知识可视化显示界面
Fig. 10 Interface of extension knowledge visualization

4 结论

本软件的特色是将可拓学理论和可拓数据挖掘方法相结合,分析客户价值,挖掘变换、变换的作用及该变换引起的客户价值变换规律知识,这是当前数据挖掘工具尚未研究的领域,也是传统的客户价值分析工具没有研究的. 本软件主要实现了可拓分类知识获取,从可拓学角度进行客户价值分类,有效地衡量营销策略优劣,为今后提升客户价值提供了策略依据. 软件在客户数据、评价体系、度量模型方面做到了一定的通用性、灵活性,适用于解决各种实际问题,在通用性、灵活性方面比文献[13]的工作做了更多的改进.

可拓数据挖掘需要庞大的数据支持,如何提升软件的数据处理速度及响应速度是需要继续研究的. 同时,如何去实现传导知识和聚类知识获取,并

将获取的可拓知识更加有效地利用起来等这些将是下一步的重点工作,为完善客户价值可拓知识挖掘软件和挖掘准确的、有价值的客户价值信息打下基础.

参考文献:

[1] 姚瑞娜. 客户价值评价指标体系研究[D]. 上海:上海财经大学国际工商管理学院,2007:1-7.
[2] 陈远高. 基于数据挖掘的客户价值管理研究[D]. 浙江:浙江大学管理学院,2002:2-10.
[3] 吴亮. 基于数据挖掘技术的客户价值管理[D]. 贵阳:贵州大学管理学院,2006:6-12.
[4] Wolfgang Ulaga. Customer value in business markets[J]. Industrial Marketing Management,2001,30:315-319.
[5] 杨永恒. 客户关系管理:价值导向及使能技术[M]. 大连:东北财经大学出版社,2002.
[6] 齐佳音,舒华英. 客户价值评价、建模、决策[M]. 北京:

- 北京邮电大学出版社,2005.
- [7] 陈静宇. 客户价值分析与客户价值细分模型研究[J]. 生产力研究,2007,(16):60-61.
- Chen Jing-yu. Research on Analytics of Customer Value and the model on class of customer value[J]. Productivity Research, 2007,(16):60-61.
- [8] 陈文伟. 挖掘变化知识的可拓数据挖掘研究[J]. 中国工程科学,2006,8(11):70-73.
- Chen Wen-wei. The research of mining the mutative knowledge with extension data mining [J]. Engineering Science, 2006, 8(11):70-73.
- [9] 蔡文,杨春燕,陈文伟,等. 可拓集与可拓数据挖掘[M]. 北京:科学出版社,2008.
- [10] 杨春燕,蔡文. 可拓数据挖掘研究进展[J]. 数学的实践与认识,2009,39(4):134-141.
- Yang Chun-yan, Cai Wen. Recent progress in extension data mining [J]. Mathematics in Practice and Theory, 2009, 39(4):134-141.
- [11] 杨春燕,蔡文. 可拓工程[M]. 北京:科学出版社,2007.
- [12] 杨春燕,李小妹,陈文伟,等. 可拓数据挖掘方法及其计算机实现[M]. 广州:高等教育出版社,2010.
- [13] 李小妹. CPI 指数变换对产品销售影响的可拓数据挖掘[J]. 数学的实践与认识,2009,39(4):179-183.
- Li Xiao-mei. Extension data mining based on products sales according to change of CPI [J]. Mathematics in Practice and Theory, 2009, 39(4):179-183.
- [14] 杨春燕,蔡文. 基于可拓集的可拓分类知识获取研究[J]. 数学实践与认识,2008,38(16):184-191.
- Yang Chun-yan, Cai Wen. Acquiring extension classification knowledge based on extensible set[J]. Mathematics in Practice and Theory,2008,38(16):184-191.
- [15] 蔡文,杨春燕. 评价信息元及其原信息元的获取方法[J]. 智能系统学报,2009,4(3):234-238.
- Cai Wen,Yan Chun-yan. A method for evaluation of information-elements and acquirement of the original information element [J]. CAAI Transactions on Intelligent Systems, 2009,4(3):234-238.
- [16] 杨春燕. 多评价特征基元可拓集研究[J]. 数学的实践与认识,2005,35(9):203-208.
- Yan Chun-yan. Study on the basic-element extension set of multi evaluating characteristics [J]. Mathematics in Practice and Theory, 2005,35(9):203-208.
- [17] 曲培新,庞永庆,林树泽,等. Java 项目开发案例精粹[M]. 北京:电子工业出版社,2010.
- [18] 林智扬,范明翔,陈锦辉. 深入浅出 Java Swing 程序设计[M]. 北京:中国铁道出版社,2005.