

<<基于统计视角的数据挖掘研究>>

图书基本信息

书名：<<基于统计视角的数据挖掘研究>>

13位ISBN编号：9787811402681

10位ISBN编号：7811402688

出版时间：2010-12

出版时间：浙江工商大学出版社

作者：徐雪琪

页数：186

字数：170000

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

## <<基于统计视角的数据挖掘研究>>

### 内容概要

笔者认为数据挖掘的出现，正是统计学适应这一变化的新的发展方向，数据挖掘并不是为了替代传统的统计分析技术，而是统计分析方法的延伸和扩展。

本书从统计学视角研究数据挖掘，以期从统计学角度对数据挖掘理论的研究有所突破和创新，同时对统计学理论在数据挖掘方向的发展做出探索。

## <<基于统计视角的数据挖掘研究>>

### 作者简介

徐雪琪，女，1974年10月生，浙江慈溪人。

经济学博士，现为浙江工商大学统计与数学学院讲师。

主要学习与研究领域为社会经济统计方法与应用、数据挖掘。

先后主持厅级科研项目1项，作为主要成员参与多项国家级科研项目和省(部)级科研项目的研究工作。

在《统计研究》、《科研管理》、《数据分析》等学术期刊上公开发表论文多篇。

# <<基于统计视角的数据挖掘研究>>

## 书籍目录

前言

第一章 绪论

第一节 选题意义

第二节 研究动态与文献综述

第三节 论文结构与研究方法

第四节 难点和创新

第二章 基于统计视角的数据挖掘理论体系

第一节 数据挖掘与统计学

第二节 基于统计视角的数据挖掘理论体系

第三节 本章小结

第三章 数据挖掘统计方法综述

第一节 数据挖掘基本问题界定

第二节 关联挖掘统计方法综述

第三节 聚类挖掘统计方法综述

第四节 分类挖掘统计方法综述

第五节 回归挖掘统计方法综述

第六节 本章小结

第四章 数据挖掘统计方法进一步研究

第一节 特征描述统计方法研究

第二节 聚类挖掘距离函数和相似系数研究

第三节 本章小结

第五章 数据挖掘质量问题研究

第一节 数据挖掘质量问题分类

第二节 源数据质量问题的处理方法

第三节 数据集成时质量问题的处理方法

第四节 数据分析时质量问题的处理方法

第五节 本章小结

第六章 可视化数据挖掘原型实现

第一节 数据挖掘原型设计基础

第二节 可视化数据挖掘框架系统设计

第三节 本章小结

第七章 数据挖掘在实际数据库上的应用示例

第一节 实验数据库介绍

第二节 客户特征描述——谁是最有价值的客户

第三节 铃声关联挖掘

第四节 本章小结

第八章 总结及研究展望

第一节 总结

第二节 研究展望

参考文献

后记

## &lt;&lt;基于统计视角的数据挖掘研究&gt;&gt;

## 章节摘录

(二) 转换性问题 聚类方法、神经网络方法、遗传方法、决策树方法等都可作为数据挖掘方法,不同的方法对数据有不同的要求,如神经网络方法要求所有的输入变量都必须是(0,1)或(-1,+1)之间的实数,而决策树方法擅长处理非数值型数据,甚至有些决策树算法专为处理非数值型数据而设计,所以数据分析时首先必须把数据转换成符合算法具体要求的形式。

如果数据未转换成所需的形式或转换不当都会产生转换性问题。

(三) 生成性问题 为了数据分析的需要,有时要对已给定的属性进行运算以生成新的属性,用以发现数据属性间相互关联的信息,并能提高对高维数据结构的理解和分析精度。

如,可运用统计方法对数据进行汇总、求平均数、求百分位数等生成新的属性。

在运用统计方法时若对数据的统计含义不够理解,如在时间数列中数据为时点指标,但却作为时期指标来求平均数等便会产生生成性问题。

(四) 模式或模型选择性问题、数据挖掘中模式或模型选择性问题也可称为模式或模型评价问题。

数据挖掘中的模式(patterns)是指在一个数据挖掘库中出现频率足以揭示它们之间有关联的一系列事件。

对于每一次数据挖掘任务,不管是采用一种数据分析方法还是采用多种不同的数据分析方法,其结果都有可能产生大量的模式,若选择了不利于决策的模式而放弃了有利于决策的模式,便会产生模式选择性问题。

另外数据挖掘方法中许多是基于计算或统计方法的。

例如,线性模型,依据解释变量个数的不同而不同;图模型,依据约束条件(图的边数)个数的不同而不同;树模型,依据叶子个数的不同而不同;多层感知器,依据隐含层和节点个数的不同而不同。因此当一类模型被确立后,从中找到“最佳的”模型就成为急需解决的问题。

而且,一个特定的数据问题可以用许多方法来解决。

例如,在预测分类问题中,可以使用logistic回归、决策树模型和神经网络等方法,如何从各种方法产生的模型中找到“最佳的”模型也成为急需解决的问题。

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>