第一图书网, tushu007.com

<<探索性数据分析及其在流程业的应>>

图书基本信息

书名：<<探索性数据分析及其在流程业的应用>>

13位ISBN编号：9787511420343

10位ISBN编号：7511420346

出版时间：2013-5

出版时间：中国石化出版社

作者：陆治荣

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

内容概要

《探索性数据分析及其在流程业的应用》全面介绍探索性数据分析的原理和应用，内容涉及数理统计、数据挖掘、模式识别和信息论等多门学科。
《探索性数据分析及其在流程业的应用》介绍了多种实用性的数据分析方法，并对几种重点算法进行了系统和深入的论述，书中还涉及数据分析中的方法论。

《探索性数据分析及其在流程业的应用》具有实用性、综合性和新颖性，既有理论又注重实践，并配以大量的图表和实例，可作为掌握大量数据、并希望从数据中获取信息和知识的工程技术人员以及企业管理人员等的参考书，也可供高等院校相关专业的师生参考。

作者简介

陆治荣，男，1966年毕业于清华大学工程数学力学系。
曾先后在航天二院，金陵石化炼油厂仪表车间、计算机站，霍尼韦尔公司（美）高技术执行部，思华数据技术有限公司，北京石油化工工程有限公司等单位工作。
近十年，专注于数据挖掘和数据分析技术的研究、开发和应用实践，期间，主持了两项数据分析软件的研发，并获得国家发明专利一项。

书籍目录

第1章绪论 1.1大数据时代 1.2数据、信息与知识 1.3探索性数据分析 1.4探索性数据分析与传统数据分析的差异 第2章数据预处理 2.1数据处理的必要性 2.2噪声的处理 2.3异常点的处理 2.4标准化处理 第3章描述性统计分析 3.1统计量 3.2统计图 第4章相关性分析 4.1简单相关 4.2复相关 4.3典型相关分析 4.4偏相关分析 4.5相关性分析的应用 第5章统计过程控制 5.1 SPC控制图 5.2多变量统计过程控制（MSPC） 5.3间歇式生产过程的统计控制 5.4 MSPC的应用 第6章基于信息论的数据分析 6.1信息熵 6.2联合熵与互信息 6.3最大关联最小冗余 第7章基于模式识别的数据分析 7.1模式识别的概念和方法 7.2特征提取方法 7.3模式识别应用于数据分析 第8章算法介绍 8.1主成分分析（PCA） 8.2偏最小二乘（PLS） 8.3支持向量机（SVM） 第9章DMoS软件介绍 9.1建模和数据分析的准备 9.2数据分析步骤 主要参考文献

章节摘录

版权页： 插图： 在特征选择时，首先，根据经验从全部特征中，选择部分特征构成一个特征子集，按照上述计算方法得到该特征子集的类间与类内离散比。
然后，逐一更换特征子集中的特征，构成新的特征子集，计算对应的类间与类内离散比，并与前面的结果比较，从中选取类间与类内离散比大的特征子集。
如此反复，直至选出具有最大离散比的特征子集，就是在距离意义上具有最优分类性能的特征子集。
（3）基于信息熵的评价准则 在信息论中，数据中的不确定性就是信息，如果数据中的不确定性越大，则它所含的信息就越多。
信息论的创始人香依将随机变量的平均不确定性定义为信息熵，并将它作为数据中信息量的度量。
如用X表示随机变量，则它的信息熵表示为：H（X），H值越大，包含的信息量越多。
两个随机变量X、Y之间在信息上的依赖程度用互信息来描述的，记作I（x，y），I值越大，表示这两个随机变量之间的依赖程度越高。
从信息论的观点进行特征选择的原则是：选择那些与目标变量在信息上有较强依赖关系的特征，而且，又要使选人特征之间的相互依赖性尽可能少。
上述特征选择原则称为“最大关联最小冗余”准则，简称mRMR准则（见本书第六章第三节）。
由于基于信息熵评价准则的特征选择对各个特征的数据分布没有做任何假设和限制，因此，既可应用于具有正态分布的特征，也可应用于非正态分布的特征。
6.特征提取 模式识别常常将描述模式的数据统称为特征。
为了实现机器识别，首先，将对象的特征转换成机器能够接收的信号，这是由仪表或其他自动化测量技术实现的，这些特征有时称之为“粗特征”。
所谓“粗”是指其中还有不少杂质，不够全面，不能直接用于模式识别，必须对它们进行提炼，得到“精特征”后才能用于模式识别。
因此，特征提取可以理解为从粗特征中提炼精特征的过程。
在分类问题中，人们总是根据那些能够识别对象的基本要素来进行分类的，这些基本要素就是对象的特征。
但是，特征不一定能观测到，观测变量不等于特征。
有些观测变量不包含任何有关分类的信息，这样的变量对分类问题毫无价值，应该予以删除。
有些观测变量从不同的角度反映对象特征的某个方面，尽管单个观测变量只包含一部分分类的信息，但是将这些观测变量综合在一起就能够形成对象的特征。
特征提取就是从这些观测变量中提取对象基本要素的过程。

编辑推荐

《探索性数据分析及其在流程业的应用》具有实用性、综合性和新颖性，既有理论又注重实践，并配以大量的图表和实例，可作为掌握大量数据、并希望从数据中获取信息和知识的工程技术人员以及企业管理人员等的参考书，也可供高等院校相关专业的师生参考。

第一图书网, tushu007.com

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