

<<大话数据挖掘>>

图书基本信息

书名：<<大话数据挖掘>>

13位ISBN编号：9787302298137

10位ISBN编号：7302298130

出版时间：2013-1

出版时间：清华大学出版社

作者：西安美林电子有限责任公司

页数：273

字数：343000

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>

<<大话数据挖掘>>

前言

前言本书的萌发上世纪80年代末到90年代初，国内外广泛流传着一句耐人寻味的话语：我们沉浸在数据的海洋中，却渴望着知识的淡水。

这句话生动地描绘了当时人们面对海量数据的迷惘和无奈。

就在这时，世界商业巨头沃尔玛从其庞大的交易数据库中演绎了一场“啤酒和尿布的故事”，揭示了一条隐藏在海量数据中的、美国人的一种行为规律：年龄在25~35岁的年轻父亲下班后经常要到超市去给婴儿买尿布，而他们中有30%~40%的人顺手为自己买几瓶啤酒。

受这条简单的客户行为模式的启发，沃尔玛调整了商品布局，并策划了促销价格，结果销售量大增。

这一现象引起了科学界的注意，他们将“啤酒和尿布的故事”引申为“关联规则获取”，进而将“从大量的、不完全的、有噪声的、模糊的、随机的数据中，提取隐含在其中的、人们事先不知道的、但又潜在有用的信息和知识的过程”定义为“数据挖掘”。

需求是成功之源，于是西方发达国家刮起了一场数据挖掘的风暴。

商业界发现了沃尔玛迅猛发展的秘诀，纷纷效仿。

电信行业也沸腾了，各公司纷纷争先恐后地利用数据挖掘这一锐利武器解决他们面临的最紧迫的问题（如客户分群、客户会流失原因及预测、业务套餐及响应、关联消费等）。

工业界也行动了，他们从堆积如山的数据中，挖掘出指导生产和管理的决策规则。

上世纪90年代中期以后，基于数理统计、人工智能、机器学习、人工神经网络等多种技术的数据挖掘技术已经成为研究和应用的热点，数据挖掘在我国也开始推广应用。

然而，从这么多年的情况来看，我国数据挖掘的应用与发达国家还有很大差距。

我们仅在互联网、金融、电信和商业等领域有一些成功的应用，而在其他行业如制造、航空、医药、军工、化工、税务、反恐和刑侦等只有少量的尝试。

为什么会这样呢？

IT界、企业界和学术界的有识之士无不在思考着这样的问题。

进行数据挖掘，数据是基础，难道是我国的信息化建设还未达到一定的程度，数据积累不够？

进入21世纪前可以这么说，可现在，显然不是。

目前，我国的大中型企业，大多建立了先进的信息化系统，甚至相当多的企业构建了数据仓库，而且数据日复一日、爆炸式地增长，可谓堆积如山。

然而，很多企业数据挖掘的认识还不全面，甚至感觉其神秘不可信，这样的话，生产管理中遇到了不能解决的问题，自然不会用数据挖掘的思想思考，甚至基层部门提出使用这样的方法，管理层却因对此不甚了解而无力推动。

为此，我们期望从领导层和生产一线的工作人员普及数据挖掘知识开始，唤起人们对数据新的认识：数据是客观实际的反映，它体现了营销规律、生产规律、经营规律和产品质量控制规律。

更重要的是，使企业管理告别基于简单统计分析的“报表”决策时期，跨入数据挖掘的“知识”决策时代。

为了实现这一目标，迫切地需要一本使企业管理者和基层工作者喜闻乐见的读物。

然而，市面上的数据挖掘书籍几乎全是教科书形式，理论性太强，满篇数学公式，让人望而却步，而且应用实例甚少，让人难以理解。

在这种情况下，我们大胆地萌发出一种案例教学法编写思路，以课堂教学为线索，介绍数据挖掘的基本概念和应用过程，让读者轻松地走进数据挖掘，领略数据挖掘的神奇魅力。

本书的读者群如果您是一位企业或政府部门的领导，您可以利用乘飞机的闲暇，与本书中的徐教授和各行各业的精英们一起，走进数据挖掘的世界，相信当您下飞机的时候，一定会浮想联翩，产生许多新的思路；如果您是一位企业管理、生产一线的技术人员，利用一个周末的休息时间，通过本书，您会对数据挖掘有初步而较为系统的了解和认识，您会自觉地尝试利用数据挖掘的方法解决实际问题；如果您是一位想系统学习数据挖掘知识的学生或科技工作者，亦可以通过本书的阅读，为以后的学习奠定好基础。

本书的内容全书共9章。

<<大话数据挖掘>>

第1章，揭开数据挖掘的面纱，从三个真实而有趣的故事开始，让读者了解数据挖掘的概念、数据挖掘产生与发展、数据挖掘的功能和数据挖掘技术，本章深入浅出地介绍了关联规则、聚类分析、预测（分类和回归）、时间序列等数据挖掘方法及常用算法；第2章简述数据挖掘流程，以某冶金企业生产中遇到的质量控制技术攻关难题的解决过程为线索，活灵活现地展现了一个数据挖掘问题的项目立项及其实施过程；第3章到第9章以典型案例的形式分别介绍了数据挖掘技术在电力行业、航空航天领域、冶金行业、税务与金融行业、故障诊断、电信行业、互联网行业方面的应用。

本书的特色形式新颖本书以EMBA班的“数据挖掘技术及其应用”教学为场景，通过教师与学员互动共鸣的形式，带领读者步入数据挖掘的神秘殿堂，领略数据挖掘的神奇魅力。

这种写作方式，避免了传统教科书理论性太强，数学公式繁多，让非专业数据挖掘者望而却步的缺陷。

案例导读本书通过数据挖掘的典型案例，引导读者领略如何利用数据挖掘技术解决各行各业生产和管理中的实际问题。

摒弃了晦涩难懂的理论，在解决问题的过程中了解数据挖掘技术及其应用方法，学会“让数据说话，以数据辅助决策”的新理念。

创作团队本书由西安交大美林数据挖掘研究中心策划，靖稳峰、卢耀宗等编写，程宏亮为本书审定了章节划分并精选了案例素材，王璐为本书审定了故事构思和语言风格，程宏斌、李炜、强劲和黄蓉等对本书提出了大量的建设性构想和修改意见，并参与了部分章节的编写。

陈浩铭和王羽为本书制作了精美插图。

致谢西安交通大学徐宗本院士在百忙中对本书的构思、写作给予了悉心指导，清华大学出版社栾大成编辑对本书原稿字斟句酌，使得本书增色不少，这里一并表示衷心感谢。

西安交大美林数据挖掘研究中心还有许多同事为本书的出版付出了大量心血，在此表示诚挚的谢意。

编者

<<大话数据挖掘>>

内容概要

本书以EMBA班的“数据挖掘技术及其应用”教学为场景，带领读者步入数据挖掘的神秘殿堂，领略数据挖掘的神奇魅力。

全书分为9章：第1章从三个真实故事开始数据挖掘之旅；第2章以某企业生产中遇到的质量控制难题的解决过程为线索，展现数据挖掘的实施过程；第3章到第9章以典型案例的形式分别介绍了数据挖掘技术在电力行业、交通航空领域、冶金行业、税务与金融行业、电信行业、故障诊断以及互联网行业的应用。

数据挖掘是一种专业性极强的技术，本书避开大量晦涩的概念和令人生畏的数学公式，以师生互动讨论的形式让读者走进数据挖掘殿堂，进而深入浅出、循序渐进地感知数据挖掘。

随着阅读，读者会自然而然地身临课堂，“让数据说话，从数据中发现规律，科学决策”等新的理念会使读者对实际工作中面临的复杂问题浮想联翩、另辟新径。

本书适合企事业部门的领导、管理人员、生产一线的技术人员，另外，学生或者行业工作者，可以通过本书的阅读，为以后的学习奠定好基础。

<<大话数据挖掘>>

作者简介

西安交大美林数据挖掘研究中心

中心集西安美林电子有限公司十五年信息化建设经验和西安交通大学二十多年数据挖掘基础与应用研究精华，致力于将先进的数据挖掘与分析处理技术与中国信息化建设紧密融合，帮助中国企业在物联网、电子商务、智慧制造、智能交通等领域创造价值。

中心数据挖掘研发团队处于国内第一梯队，拥有世界领先的视觉聚类和基于L1/2正则化的稀疏分类、回归算法及国内多行业数据挖掘实践经验。

西安交大美林数据挖掘研究中心努力以数据挖掘技术推动中国企业实现“中国创造”！

<<大话数据挖掘>>

书籍目录

第1章 揭开数据挖掘的面纱

1.1 历史的使命

1.2 数据挖掘的故事

1.2.1 震撼业界的发现

1.2.2 降低成本的绝活

1.2.3 出奇制胜的小纸条

1.3 什么是数据挖掘？

1.4 历史的必然

1.5 数据挖掘能干什么？

1.5.1 关联（ASSOCIATION）规则挖掘

1.5.2 聚类

1.5.3 预测

1.5.4 序列和时间序列

1.6 数据挖掘工具

第2章 数据挖掘流程

2.1 李部长其人

2.2 老革命遇见了新问题

2.3 钓鱼钓来了数据挖掘思路

2.4 数据挖掘项目立项

2.5 数据挖掘项目实施

2.5.1 业务理解阶段（BUSINESS UNDERSTANDING）

2.5.2 数据理解阶段（DATA UNDERSTANDING）

2.5.3 数据准备阶段（DATA PREPARATION）

2.5.4 建模阶段（MODELING）

2.5.5 模型评估阶段（EVALUATION）

2.5.6 部署阶段（DEPLOYMENT）

2.6 李部长的展望

第3章 数据挖掘在电力行业的应用

3.1 应用前景

3.2 电力设备状态检修

3.3 电力系统暂态稳定性评估

3.4 负荷预测

3.5 盗电检测

3.6 电力数据挖掘系统的构建

第4章 数据挖掘在交通航空领域的应用

4.1 铁路票价制定

4.2 高铁轨道检修

4.3 交通流量预测

第5章 数据挖掘在冶金行业的应用

5.1 流程工业这点儿事

5.2 产品质量控制

5.3 高炉炉温预测

5.4 磨矿粒度预测

<<大话数据挖掘>>

5.5 炼焦配煤优化

第6章 数据挖掘在税务、金融行业的应用

6.1 税务稽查

6.2 反洗钱

6.3 股票指数追踪

第7章 数据挖掘在故障诊断中的应用

7.1 火箭发动机故障诊断

7.2 机械设备故障诊断

7.3 核动力设备故障诊断

7.4 船舶动力故障诊断

第8章 数据挖掘在电信业中的应用

8.1 市场细分

8.1 市场细分

8.2 精确营销

8.3 业务响应

8.4 客户流失分析

第9章 Web数据挖掘

9.1 Web数据挖掘概述

9.1 Web数据挖掘概述

9.2 垂直搜索引擎中的数据挖掘

9.3 面向电子商务的数据挖掘

9.4 社交网络中的数据挖掘

参考文献

<<大话数据挖掘>>

章节摘录

版权页：插图：徐教授：“刘总，问你一个可能涉及到隐私的问题，你负责你们公司产品的营销活动策划吧？”

”刘总说：“是的，徐老师，你比外边那些算命的能掐会算多了。”

”姚局长说：“徐教授，莫非您也精读了周易？”

”大家都笑翻了，开始更加好奇徐老师是怎么知道的呢。

徐教授：“因为Unica Model 1这个软件是一款典型的、针对市场营销和策划行业而研发的软件。”

”“原来是这样”，学员们恍然大悟。

徐教授接着说：“Unica Model 1这个软件很经典，非常畅销。

它涵盖了响应模型、交叉销售模型、客户价值评估模型、市场细分模型等，这四部分简直就是这个软件的四大金刚。

还有那个同学愿意自告奋勇地给大家讲讲其他数据挖掘工具？”

”工行的张行长说：“我对SAS软件了解一些，该系统全称为Statistics Analysis System，最早由北卡罗来纳大学的两位生物统计学研究生编制，并于1976年成立了SAS软件研究所，正式推出了SAS软件。经过多年的发展，SAS已被全世界120多个国家和地区的近3万家机构所采用，直接用户则超过300万人，遍及金融、医药卫生、生产、运输、通讯、政府和教育科研等领域。

”上海一家钢铁公司的贾总站起来了，补充说道：“我们公司使用的就是SAS软件。

由于SAS系统是从大型机系统发展而来，在设计上也完全针对专业用户，因此其操作至今仍以编程为主，人机对话界面不太友好，并且在编程操作时需要用户最好对其使用的统计方法有较清楚地了解，非统计专业人员掌握起来较为困难。

而且SAS极为高昂的价格和只租不卖的销售策略使得实力不足的个人和机构只能望而却步。

不过，由于其功能强大，我公司专业人员较多，这几年我们不惜巨资每年都在租用该软件。

”徐教授感到很惊讶：“咱们这个班果真卧虎藏龙。”

张行长和贾总回答地非常专业。

不知道的人还以为你俩是SAS公司的‘山寨’销售专家呢。

”贾总笑了笑，不好意思地说：“我大学同宿舍的一位同学在SAS北京办事处工作，经常来上海推销他们的产品，每次顺便来我这儿蹭酒喝，免不了给我叨叨他们的SAS，时间长了我就耳熟能详了。”

”徐教授也乐了：“原来如此！”

”徐教授的话音刚落，市统计局程副局长立即站了起来：“SAS太专业了，我们统计分析用SPSS。”

<<大话数据挖掘>>

编辑推荐

大话经典系列新品重拳出击！

掌握数据爆炸时代先人一步的新思维EMBA班教学实景展现解析电力、交通、税务、故障诊断、互联网等行业的真实案例谈笑间深入理解数据挖掘的来龙去脉，入门必读！

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>