

数据可视化与 挖掘技术实践

SHUJU KESHIHUA YU WAJUE JISHU SHIJIAN

朱希安 王占刚◎编著



知识产权出版社

全国百佳图书出版单位

数据可视化与挖掘技术实践

朱希安 王占刚 编著



知识产权出版社

全国百佳图书出版单位

图书在版编目 (CIP) 数据

数据可视化与挖掘技术实践/朱希安, 王占刚编著. —北京: 知识产权出版社, 2017. 7

ISBN 978-7-5130-4990-0

I. ①数… II. ①朱… ②王… III. ①可视化软件—研究 ②数据采集—研究 IV. ①TP31 ②TP274

中国版本图书馆 CIP 数据核字 (2017) 第 150927 号

内容提要

本书将可视化与数据挖掘技术应用于煤层气田勘探与生产实践, 在介绍数据可视化与数据挖掘关键技术的理论、技术和方法的基础上, 以煤层气产业为例, 详细阐述了煤层气数据可视化与挖掘系统的总体架构和功能结构设计; 描述了系统数据库设计过程; 重点介绍了系统主要模块设计与核心算法实现; 最后阐述了煤层气数据可视化与挖掘系统三大子系统研发成果。

本书适合从事煤层气勘探、地质、矿山开采、测绘、城市地下规划等领域的研究开发人员及相关专业的教师、学生参考使用。

责任编辑: 栾晓航

责任校对: 谷 洋

封面设计: 刘 伟

责任出版: 刘译文

数据可视化与挖掘技术实践

朱希安 王占刚 编著

出版发行: 知识产权出版社有限责任公司	网 址: http://www.ipph.cn
社 址: 北京市海淀区气象路 50 号院	邮 编: 100081
责编电话: 010-82000860 转 8382	责编邮箱: luanxiaohang@cnipr.com
发行电话: 010-82000860 转 8101/8102	发行传真: 010-82000893/82005070/82000270
印 刷: 三河市国英印务有限公司	经 销: 各大网上书店、新华书店及相关专业书店
开 本: 720mm×1000mm 1/16	印 张: 14.75
版 次: 2017 年 7 月第 1 版	印 次: 2017 年 7 月第 1 次印刷
字 数: 170 千字	定 价: 50.00 元
ISBN 978-7-5130-4990-0	

出版版权专有 侵权必究
如有印装质量问题, 本社负责调换。

简 介

本书依据“关键技术—系统总体架构设计—数据库构建—系统详细设计—应用成果”的思路进行组织，将数据整合与管理、图形可视化、数据挖掘与优化有机融合为一体，在介绍数据可视化与数据挖掘关键理论、技术和方法的基础上，以煤层气产业为例，将可视化与数据挖掘技术应用于煤层气田勘探与生产实践。书中详细阐述了煤层气数据可视化与挖掘系统的总体架构和功能结构设计，系统数据库设计。重点研究了系统主要模块与核心算法实现，最后给出了煤层气数据可视化与挖掘系统三大子系统的研发成果。

本书适合从事可视化与数据挖掘技术研究的科研人员、教师、学生参考使用。亦可为从事煤层气勘探、矿山开采、测绘、城市规划等领域的研究开发人员及相关专业的教师、学生提供参考和帮助。

本书出版得到了国家科技重大专项——煤层气田地面集输信息集成及深度开发技术（项目编号：2011ZX05039-004-02）科研项目的资助，在此表示衷心的感谢。

前 言

面对“数据丰富而知识贫乏”的窘境，如何从大量数据中提取信息，并将其转化成有用的知识，已成为当前研究与应用的热点。可视化与数据挖掘技术正是在这样一种需求背景下得到广泛应用并迅速发展起来。

针对复杂数据集进行探索和表达的过程中，可视化技术是比较有效的途径之一。在计算机图形学、图像处理、计算机视觉及人机交互技术支撑下，数据可视化利用几何图形、色彩、纹理、透明度、对比度及动画技术等手段，以图形图像的形式直观、形象地表达抽象数据，并进行交互处理。数据可视化技术通过将数据变换为可识别的图形符号、图像、视频或动画，并以此呈现对用户有价值的信息。用户通过对可视化的感知，使用可视化交互工具进行数据分析，获取知识，并进一步提升为知识。

为了发现数据背后的潜在信息与知识，跨越数据和信息之间的鸿沟，数据挖掘（Data Mining）技术应运而生。数据挖掘是在海量数据的基础上，发现潜在的可以为人所用的甚至可能是违背常理逻辑的知识和信息。数据挖掘可以从实际数据中提取隐含在其中的、人们事先不知道的、但潜在有用的信息和知识。数据挖掘技术的应用十分广泛，可以用来进行商业智能应用和决策分

析等。

本书将可视化与数据挖掘技术应用于煤层气田勘探与生产实践，详细介绍了数据可视化和数据挖掘技术研究现状，并以煤层气产业为例，描述了煤层气田数据可视化与挖掘系统的应用与需求分析，详细阐述了系统总体架构方案、功能模块设计、主要流程算法，以及系统开发成果等内容。

本书共分为8章，由朱希安、王占刚撰写。在本书资料整理及校稿过程中，张朋、杨昊、邱中原、万韶、于芳源等同学参加了整理工作。

第1章作为入门概述，从整体上介绍了数据可视化与数据挖掘技术的概念、发展历史、技术特点等，并叙述两种技术在煤层气行业中的应用。

第2章介绍可视化技术的理论基础与技术方法，详细阐述了数据可视化概况、特点、作用、流程及相关概念，最后着重介绍了几种典型可视化方法。

第3章介绍数据挖掘技术的理论基础与技术方法，详细阐述了数据挖掘的基本概念、特点、应用及流程，着重介绍了目前几种典型的数据挖掘方法。

第4章介绍了系统设计、开发与运行的平台，详细阐述了 Visual .NET Framework 平台特性、优势与构造模块，介绍了 C# 开发语言、Visual Studio .NET 开发工具和数据库管理系统 SQL Server 2008，最后介绍了系统设计工具 Office Visio 和界面设计工具 DevExpress。

第5章详细介绍了煤层气可视化与数据挖掘系统的总体设计

情况，主要包括系统建设意义、建设目标和功能需求、总体架构设计、功能结构设计、系统主要工作流程设计和软件功能菜单和界面风格设计等。

第6章描述了系统数据库的具体设计过程，主要包括数据库设计需求分析、数据库概念结构设计、逻辑结构设计、物理结构设计和数据库软硬件环境设计等内容。

第7章选取系统中具有代表性的数据整合与管理模块、图形可视化模块、生产查询模块和生产预测模块作为典型案例，基于模块化设计思想详细描述了各个功能模块设计过程及核心算法实现。

第8章详细阐述了煤层气数据可视化与挖掘系统三大子系统研发成果。

本书依托国家科技重大专项——煤层气田地面集输信息集成及深度开发技术（项目编号：2011ZX05039-004-02），对煤层气数据可视化与挖掘系统进行了详细阐述，并得到了该科研项目的资助，在此表示衷心的感谢。

由于数据可视化和数据挖掘技术发展迅速，且作者水平有限，成书时间仓促，书中难免出现错误和不足之处，敬请广大读者批评指正。

作者

2017年3月28日

目 录

第 1 章 绪 论	1
1.1 引 言 / 3	
1.2 数据挖掘与数据可视化技术背景 / 6	
1.3 数据挖掘在煤层气行业中的应用 / 8	
1.4 数据可视化在煤层气行业中的应用 / 12	
1.5 数煤层气数据可视化与挖掘需求 / 15	
1.6 本章小结 / 17	
第 2 章 数据可视化技术	19
2.1 数据可视化技术概述 / 21	
2.2 数据可视化及相关概念 / 26	
2.3 数据可视化流程 / 31	
2.4 典型数据可视化方法 / 33	
2.5 本章小结 / 39	

第 3 章	数据挖掘技术	41
3.1	数据挖掘基本概念 /	43
3.2	数据挖掘流程 /	47
3.3	数据挖掘主要方法 /	50
3.4	本章小结 /	70
第 4 章	系统设计开发与运行平台	71
4.1	Visual. NET Framework 与 C# /	73
4.2	SQL Server 2008 数据库管理系统 /	81
4.3	系统与界面设计工具 /	84
4.4	本章小结 /	88
第 5 章	煤层气数据可视化与挖掘系统总体设计	89
5.1	煤层气数据可视化与挖掘系统建设意义 /	91
5.2	软件系统建设目标与功能需求 /	93
5.3	数据可视化与挖掘系统总体架构设计 /	95
5.4	数据可视化与挖掘系统功能结构设计 /	99
5.5	系统工作流程图设计 /	106
5.6	软件功能菜单和界面风格 /	108
5.7	本章小结 /	112
第 6 章	煤层气数据可视化与挖掘系统数据库设计	113
6.1	数据库设计需求分析 /	115
6.2	数据库概念结构设计 /	119

6.3	数据库逻辑结构设计 / 124	
6.4	数据库物理结构设计 / 127	
6.5	数据库软硬件环境设计 / 141	
6.6	本章小结 / 142	
第7章	煤层气数据可视化与挖掘系统功能模块 详细设计与实现	143
7.1	数据整合与管理功能模块详细设计与实现 / 145	
7.2	图形可视化模块详细设计与实现 / 158	
7.3	生产查询模块详细设计与实现 / 169	
7.4	生产预测模块详细设计与实现 / 181	
7.5	本章小结 / 192	
第8章	煤层气数据可视化与挖掘系统研发成果	193
8.1	系统建设总体成果 / 195	
8.2	煤层气田数据整合与管理子系统 / 196	
8.3	煤层气田异构数据图形可视化子系统 / 200	
8.4	煤层气田数据挖掘与优化子系统 / 208	
8.5	本章小结 / 216	
参考文献		217

绪 论

1.1 引 言

随着信息技术尤其是网络技术的飞速发展，人们利用信息技术生产和收集数据的能力大幅提高，计算机处理和存储的数据量急剧增长，数据库应用的规模、范围和深度也随之不断扩大；另一方面，随着社会经济的不断发展，商业竞争日趋白热化，人们迫切需要掌握隐藏在大量数据背后的具有决策意义的知识。面对“数据丰富而知识贫乏”的窘境，如何从大量数据中提取信息，并将其转化成有用的知识，已成为当前研究与应用的热点。数据挖掘与可视化技术正是在这样一种需求背景下得到广泛应用并迅速发展起来的。

数据挖掘技术是一个多学科交叉研究领域，兴起于 20 世纪 80 年代末，目前已经取得了重大研究进展。数据挖掘融合了当今信息处理技术中的很多研究热点，主要包括人工智能、机器学习、知识工程、面向对象方法、信息检索、高性能计算以及数据可视化等（朱明等，2012）。从本质上讲，数据挖掘过程是在海量数据中利用各种分析工具来发现所建立或假设的模型和现有数据间关系的过程。这些关系和模型的主要功能是对目标或事件进

行分析和预测，从而为用户提供决策支持。

数据可视化技术利用视觉化可视化的方式将纷繁复杂的数据集艺术地展示出来，使人们能够以更直观的方式看到数据及其结构的关系，不再局限于以往通过关系数据表来分析数据的形式（张浩等，2012）。数据可视化通常针对大型数据库或数据仓库中的数据，是可视化技术在非空间数据领域的应用。

借助于计算机强大的处理能力、计算机图形图像算法以及可视化算法等，数据可视化将海量数据转换为静态或者动态图形、图像直观地呈现在人们面前，同时可以通过交互手段控制数据的抽取和画面的显示过程，从而使得隐含于数据之中的不可见的现象变为可见。数据可视化技术为人们分析理解数据、形成概念、发现规律等行为提供了强有力的手段（陈为等，2013）。

数据挖掘与数据可视化是紧密联系的。数据挖掘的目的是找出“数据矿山”中真正需要的具有决策意义的信息，而数据可视化技术能够实现对数据信息的分析和提取，以图形、图像、虚拟现实等容易被人们所辨识的方式或手段来展现原始数据间的复杂关系、潜在信息以及发展趋势等，从而极大地丰富了科学发现的过程（施惠娟，2010）。可视化技术能够准确地表达以及直观地展示数据挖掘的过程和结果，通过使用户深入地理解问题，选择更适当的数据挖掘算法，从而达到深入剖析数据的目的。

数据可视化与信息图形、信息可视化、科学可视化以及统计图形学等学科的发展密切相关。在研究、教学和工程等领域，数据可视化仍是一个极为活跃而又关键的方向。“数据可视化”这条术语使得成熟的科学可视化领域与较年轻的信息可视化领域得

以实现统一。可视化发展与推广的主要原因是颜色和几何表示很容易被人类的大脑所识别和理解，数据可视化技术能够将提供的数据挖掘结果通过操作界面更为自然和直观地展示给用户。目前，数据挖掘用户经常使用传统的数据可视化工具绘制二维图和三维图，这也是两种技术结合的一个典型事例。理论和实践表明，这种方式在向用户展示数据时变得尤为适用。

1.2 数据挖掘与数据可视化技术背景

统计表明, 20 世纪 90 年代以来人类累计的数据量以每月高于 15% 的速度增加。在信息科学领域, 数据库与信息处理相关技术得到了快速发展, 然而信息的分析处理能力仍远远落后于信息飞速膨胀的速度, 从而导致了面对海量数据而无法有效利用的现象, 并且这一现象正随着数据量的暴增而逐渐恶化。导致这一现象的原因, 不仅是信息量的巨大、高维、多源、异构等因素, 更重要的是如何动态地获取数据、数据中噪声相互矛盾的因素以及数据关系异构与易质等深层次的问题存在。在此背景下, 数据挖掘和数据可视化技术应运而生, 并逐渐显示出强大的生命力, 这些技术的蓬勃发展预示着信息处理进入了一个新的更高级的发展阶段。

数据挖掘对数据的处理并非简单的类似于数据库中检索匹配的操作, 而是一种面向不同领域的复杂的应用技术。数据挖掘不仅可以对数据进行宏观的统计分析、推理等, 还可以进行微观以及中观级的分析处理操作, 从而发现隐藏其中的潜在信息和知识 (邵峰晶等, 2003)。数据挖掘的目的是在海量数据的基础上, 发

现潜在的可以为人所用的甚至可能是违背常理逻辑的知识和信息，例如，经典的超市中“啤酒和尿布”挖掘案例，极大地促进了超市的销售利润。

在针对大型数据集进行探索和理解的过程中，数据可视化技术是最有效的途径之一。把数字置于视觉空间中，人们的大脑会更容易发现其中潜藏的模式，这是因为人类对较为直观的图形的理解和接受能力远远超过对文字等其他手段的理解和接受能力，往往能很快从中发现一些利用常规统计方法较难挖掘到或者容易被忽略的信息。

虽然数据挖掘和数据可视化技术都是以从数据中获取知识为目的，但是两者的手段侧重不同。这两种技术已成为科学探索、工程实践与社会生活中不可缺少的数据处理和分析手段。数据挖掘通过计算机获取隐藏在数据背后的知识，并将得到的知识或规律直接给予用户；而数据可视化将数据或规律用直观的易于感知的图形符号呈现给用户，并且通过可视化的交互界面让用户交互地理解数据背后的本质。