第一图书网, tushu007.com

<<数据挖掘>>

图书基本信息

书名：<<数据挖掘>>

13位ISBN编号：9787111188285

10位ISBN编号：7111188284

出版时间：2006-4

出版时间：机械工业出版社

作者：Jiawei Han

页数：770

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

第一图书网, tushu007.com

前言

We are deluged by data——scientific data, medical data, demographic data, financial data, and marketing data. People have no time to look at this data. Human attention has become the precious resource. So, we must find ways to automatically analyze the data, to automatically classify it, to automatically summarize it, to automatically discover and characterize trends in it, and to automatically flag anomalies. This is one of the most active and exciting areas of the database research community. Researchers in areas includ- ing statistics, vis..

内容概要

我们产生和收集数据的能力正在快速增长。
除了大多数商业、科学和政府事务的日益计算机化会产生数据之外，数码相机、发布工具和条码的广泛应用也会产生数据。
在数据收集方面，扫描的文体和图像平台、卫星遥感系统和国际互联网已经使我们的生活被巨大的数据量所包围。
这种爆炸性的数据增长促使我们比以往更迫切地需要新技术和自动化工具来帮助我们将这些数据转换为有用的信息和知识。
　本书第1版曾被KDnuggets的读者评选为最受欢迎的数据挖掘专著，是一本可读性极佳的教材。
它从数据库角度全面系统地介绍了数据挖掘的基本概念、基本方法和基本技术以及数据挖掘的技术研究进展，重点关注其可行性、有用性、有效性和可伸缩性问题。
但是，自第1版出版之后，数据挖掘领域的研究又取得了很大的进展，开发出了新的数据挖掘方法、系统和应用。
第2版在这一方面进行了加强，增加了多个章节讲述最新的数据挖掘方法，以便能够挖掘出复杂类型的数据，包括流数据、序列数据、图结构数据、社群网络数据和多重关系数据。
　本书适合作为高等院校计算及相关专业高年级本科生的选修课教材，特别适合作为研究生的专业课教材，同时也可供从事数数据挖掘研究和应用开发工作的相关人员作为必备的参考书。
　本书主要特点是：全面实用地论述了从实际业务数据中抽取出的读者需要知道的概念和技术。
更新并结合了来自读者的反馈、数据挖掘领域的技术变化以及统计和机器学习方面的更多资料。
包含了许多算法和实际示例，全部以易于理解的伪代码编写，适用于实际的大规模数据挖掘项目。

第一图书网, tushu007.com

作者简介

韩家炜，伊利诺伊大学厄巴纳一尚佩恩分校计算机科学系教授。
由于在数据挖掘和数据库系统领域卓有成效的研究工作，他曾多次获得各种荣誉和奖励，其中包括2004年ACM SIGKDD颁发的创新奖。
同时，他还是《ACM Trarlsactiorls on Krlowledge Discovery fronl Data》杂志的主编，以

书籍目录

Foreword vii Preface ix Chapter1 Introduction 1.1 What Motivated Data Mining? Why Is It Important? 1.2 So, What Is Data Mining? 1.3 Data Mining-On What Kind of Data? 1.3.1 Relational Databases 1.3.2 Data Warehouses 1.3.3 TransactionalDatabases 1.3.4 Advanced Data and Information Systems and Advanced Applications 1.4 Data Mining Functionalities---What Kinds of Patterns Can Be Mined? 1.4.1 Concept/Class Description: Characterization and Discrimination 1.4.2 Mining Frequent Patterns, Associations, and Correlations 1.4.3 Classification and Prediction 24 1.4.4 Cluster Analysis 1.4.5 Outlier Analysis 26 1.4.6 Evolution Analysis 1.5 Are All of the Patterns Interesting? 1.6 Classification of Data Mining Systems 1.7 Data Mining Task Primitives 1.8 Integration of a Data Mining System with a Database or Data Warehouse System 1.9 Major Issues in Data Mining 1.10 Summary Exercises Bibliographic Notes Chapter2 Data Preprocessing 2.1 Why Preprocess the Data? 2.2 Descriptive Data Summarization 2.2.1 Measuring the Central Tendency 2.2.2 Measuring the Dispersion of Data 2.2.3 Graphic Displays of Basic Descriptive Data Summaries 2.3 Data Cleaning 2.3.1 Missing Values 2.3.2 Noisy Data 2.3.3 Data Cleaning as a Process 2.4 Data Integration and Transformation 2.4.1 Data Integration 2.4.2 Data Transformation 2.5 Data Reduction 2.5.1 Data Cube Aggregation 2.5.2 Attribute Subset Selection 2.5.3 DimensionalityReduction 2.5.4 Numerosity Reduction 2.6 Data Discretization and Concept Hierarchy Generation 2.6.1 Discretization and Concept Hierarchy Generation for Numerical Data 2.6.2 Concept Hierarchy Generation for Categorical Data 2.7 Summary 97 Exercises 97 Bibliographic Notes Chapter3 Data Warehouse and OLAP Technology: An Overview 3.1 What Is a Data Warehouse? 3.1.1 Differences between Operational Database Systems and Data Warehouses 3.1.2 But, Why Have a Separate Data Warehouse? 3.2 A Multidimensional Data Model 3.2.1 From Tables and Spreadsheets to Data Cubes 3.2.2 Stars, Snowflakes, and Fact Constellations: Schemas for Multidimensional Databases 3.2.3 Examples for Defining Star, Snowflake, and Fact Constellation Schemas……Chapter4 Data Cube Computation and Data GeneralizationChapter5 Mining Frequent Patterns, Associations, and CorrelationsChapter6 Classification adn PredidctionChapter7 Cluster AnalysisChapter8 Mining Stream, Time-Series, and Sepuence DataChapter9 Graph Mining, Social Network Analysis, and MultirelationalChapter10 Mining Object, Spatial, Multimedia, Test, and Wed DataChapter11 Applications and Trends in Data MiningAn Introduction to Microsoft's OLE DB forBibliographyIndex

媒体关注与评论

一本面向中高级读者的数据挖掘好书－ 评 《数据挖掘：概念与技术》左荣国rongguo.zuo@gmail.com近几年来，随着信息技术的飞速发展，计算机领域积累了大量的数据。
从元数据的角度考虑，这些数据仅仅是数据，数据产生后，后续对元数据的再次利用却很少。
长此下去，这些数据慢慢就会变成没有用的垃圾数据。
人们越来越认识到这个问题的严重性，由此逐渐产生了知识发现技术，也就是数据挖掘的前身。
知识发现源自于人工智能的机器学习领域其实质是在一个已知状态的数据集上，通过设定一定的学习算法，从数据集中获取所谓的知识。
而与此同时数据库技术也已经发展到一定的阶段，并得到了广泛应用，各个企业都已经积累了无数的数据资源，迫切需要有一种技术能够帮助他们从数据中发掘出其内在的规律。
数据挖掘技术正好能满足这一需求，它实质上就是知识发现技术在数据库领域中的应用，其主要应用领域涵盖了商务管理、生产管理、市场分析、工程设计和科学探索等。
因此，可以说数据挖掘是一个从数据到知识的过程。
谈到学习数据挖掘，《数据挖掘：概念与技术》就是一本全面而深入地介绍数据挖掘实用技术的经典教材。
本书作者美籍华人韩家炜教授有着丰富的理论和实际经验。
在本书中，他从数据库的角度针对数据挖掘是什么、数据挖掘能解决什么问题、数据挖掘的关键问题和数据挖掘涉及的相关技术等问题做了深入的分析，由浅入深地把读者引入数据挖掘这个未知领域。
首先，数据挖掘是什么。
书中指出，数据挖掘是指从大量数据中提取或挖掘知识的过程。
用通俗的语言说，数据挖掘就像淘金者从石头或沙子中寻找金子的过程。
另外一个非常重要的同时也在本书中进行了详细说明的是数据挖掘的对象（也就是大量数据）在什么地方？
书中指出，数据挖掘主要针对关系型数据库、数据仓库、高级数据库等。
对于数据库开发人员，这些都是已经比较熟悉的领域了。
而本书中的数据挖掘主要针对的是数据库系统。
其次，数据挖掘能解决什么问题。
挖掘来的数据是用来解决企业中的问题的，否则挖掘就没有任何意义了。
书中指出，数据挖掘可以被用来进行市场分析和管理、风险分析和管理、缺陷分析和管理等。
具体来讲，比如对大学里的学生课程数据，可以用数据挖掘的方式来获得学生姓名、地址、状态和相关课程等信息及其相互关系。
再次，数据挖掘的关键问题。
本书从数据挖掘方法和用户交互、数据挖掘的性能以及不同数据库类型的数据挖掘等多角度出发，把数据挖掘做了非常透彻的分解，并对分解后的问题逐一解决。
第四方面，数据挖掘涉及的技术。
对于数据挖掘涉及的OLAP和数据仓库、数据清理、数据集成和变换、数据简化、数据挖掘查询语言、数据库支持的数据挖掘规则、数据挖掘分类和预测、数据聚合方法、对多媒体数据库和万维网的数据库挖掘方法等技术，本书作者都从理论和实践的角度做了深入的剖析，让读者不仅知其然，而且还知其所以然。
数据挖掘是可以由多个层面来提供支持的，包括数据库层面、应用程序层面等。
当前，包括Oracle、IBM、Sybase和Microsoft等在内的多家数据库提供商都从数据库层面提供了数据挖掘的支持。
对于应用层面讲，包括SAS、SPSS、COGNOS等软件在内的应用程序都对数据挖掘提供了很好的支持。
本书从理论和实际相结合的角度出发，在书中提供了大量的理论、模型、算法和实际的应用例子，包括多维数据模型、三层数据仓库架构、数据立方体的实际查询例子等。

借助本书，读者可以理论联系实际，把数据挖掘的应用提高一个层次。
另外，书中提供了大量的例子、习题以及一些基本概念的比较。
如数据库与数据仓库的区别、用于数据挖掘的OLAP与数据仓库技术等。
所以，即使您对数据库比较陌生，也能根据本书内容的安排，循序渐进地领悟到数据挖掘的真谛。
同时，您可以根据书中提供的习题，检验自己对书中内容的理解。
“实践是检验真理的唯一标准”，书中提供的例子在实际开发应用中都会用到，让您不会为了读书而读书。
总之，该书图文并茂，理论联系实际，非常适合作为从事数据挖掘研究和应用开发工作的相关人员的参考书，也特别适合作为高年级本科生和研究生的专业课教材。

编辑推荐

《数据挖掘概念与技术》(英文版)(第2版)适合作为高等院校计算机及相关专业高年级本科生的选修课教材，特别适合作为研究生的专业课教材，同时也可供从事数据挖掘研究和应用开发工作的相关人员作为必备的参考书。
本书更新和改进了原本已十分丰富和全面的第1版内容，并增添了新的重要课题，例如挖掘流数据、挖掘社群网络和挖掘空间、多媒体和其他复杂数据。
本书将是一本适用于数据挖掘和知识发现课程的优秀教材。
——Gregory Piatetsky-Shapiro, KDnuggets的总裁　　本书第2版最完整、最全面地讲述了数据挖掘领域的重要知识和技术创新。
相比内容已经相当全面的第1版，第2版展示了该领域的最新研究成果，例如挖掘流、时序数据和序列数据以及挖掘空间、多媒体、文本和Web数据。
本书是数据挖掘和知识发现领域内所有教师、研究人员、开发人员和用户都必读的一本书。
——Hans-Peter Kriegel, 德国慕尼黑大学　　我们产生和收集数据的能力正在快速增长。
除了大多数商业、科学和政府事务的日益计算机化会产生数据之外，数码相机、发布工具和条码的广泛应用也会产生数据。
在数据收集方面，扫描的文本和图像平台、卫星遥感系统和国际互联网已经使我们的生活被巨大的数据量所包围。
这种爆炸性的数据增长促使我们比以往更加迫切地需要新技术和自动化工具来帮助我们将这些数据转换为有用的信息和知识。

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