

<<数据挖掘原理、算法及应用>>

图书基本信息

书名：<<数据挖掘原理、算法及应用>>

13位ISBN编号：9787560627311

10位ISBN编号：7560627315

出版时间：2012-1

出版时间：西安电子科技大学出版社

作者：李爱国，库向阳 编著

页数：255

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>

<<数据挖掘原理、算法及应用>>

内容概要

《数据挖掘原理、算法及应用》以各类数据挖掘算法为核心，以智能数据分析技术的发展为主线，结合作者自身的研究和应用经验，阐述数据挖掘研究领域的主要理论和典型算法。全书共分8章：第1章为绪论；第2-5章分别介绍数据挖掘的主要技术、各类典型算法及其编程实现，包括数据预处理技术，关联规则挖掘技术、分类技术，聚类技术等几大类技术和其中包含的典型算法；第6-8章分别简要介绍一些数据挖掘的应用专题，包括时间序列数据挖掘、Web挖掘、空间数据挖掘等。

书籍目录

第1章 绪论

- 1.1 数据挖掘的概念和定义
- 1.2 数据挖掘的历史及发展
- 1.3 数据挖掘的研究内容及功能
- 1.4 数据挖掘的常用技术及工具
- 1.5 数据挖掘的应用热点
- 1.6 小结
- 习题

第2章 数据预处理

- 2.1 数据预处理的的目的
- 2.2 数据清理
- 2.3 数据集成和数据变换
- 2.4 数据归约
- 2.5 数据离散化和概念分层
- 2.6 特征选择与提取
- 2.7 小结
- 习题

第3章 关联规则挖掘

- 3.1 基本概念
- 3.2 关联规则挖掘算法
- 3.3 Apriori改进算法
- 3.4 不候选产生挖掘频繁项集
- 3.5 使用垂直数据格式挖掘频繁项集
- 3.6 挖掘闭频繁项集
- 3.7 挖掘各种类型的关联规则
- 3.8 相关分析
- 3.9 肇于约束的关联规则
- 3.10 矢量空间数据库中关联规则的挖掘
- 3.11 小结
- 习题

第4章 分类和预测

- 4.1 分类和预测的基本概念和步骤
- 4.2 基于相似性的分类算法
- 4.3 决策树分类算法
- 4.4 贝叶斯分类算法
- 4.5 人工神经网络(ANN)
- 4.6 支持向量机
- 4.7 预测
- 4.8 预测和分类中的准确率
- 4.9 评估分类器或预测器的准确率
- 4.10 小结
- 习题

第5章 聚类方法

- 5.1 概述
- 5.2 划分聚类方法

<<数据挖掘原理、算法及应用>>

5.3 层次聚方法

5.4 密度聚类方法 聚类结构

5.5 基于网格聚类方法 变换聚类 聚类方法

5.6 神经网络聚类方法：SOM

5.7 异常检测

5.8 小结

习题

第6章 时间序列数据挖掘

6.1 概述

6.2 时间序列数据建模

6.3 时间序列预测

6.4 时间序列数据库相似搜索

6.5 从时间序列数据中发现感兴趣模式

6.6 小结

习题

第7章 Web挖掘

7.1 挖掘的分类及基数据来源

7.2 Web日志挖掘

7.3 Web内容挖掘

7.4 小结

习题

第8章 复杂类型数据挖掘

8.1 空间数据挖掘 空间OLAP

8.2 文本数据挖掘

8.3 多媒体数据挖掘 遥感影像分类

8.4小结

习题

参考文献

<<数据挖掘原理、算法及应用>>

章节摘录

版权页：第1章 绪论 1.1 数据挖掘的概念和定义 数据挖掘(Data Mining)是近年来随着人工智能和数据库技术的发展而出现的一门新兴技术。

它是从大量的数据中筛选出有效的、可信的以及隐含信息的高级处理过程。

数据挖掘包含丰富的内涵，是一个多学科交叉的研究领域。

仅从从事研究和开发的人员来说，其涉及范围之广是其他领域所难以企及的，既有大学里的专门研究人员，也有商业公司的专家和技术人员。

研究背景的不同会使他们从不同的角度来看待数据挖掘的概念。

因此，理解数据挖掘的概念不是简单地下个定义就能解决的问题。

1.1.1 从商业角度看数据挖掘技术 数据挖掘是一种新的商业信息处理技术。

数据挖掘技术把人们对数据的应用从低层次的联机查询操作提高到决策支持、分析预测等更高级的应用上。

通过对特定数据进行微观、中观乃至宏观的统计、分析、综合和推理，发现数据间的关联性、未来趋势以及一般性的概括知识等，这些知识性的信息可以用来指导高级商务活动，如顾客分析、定向营销、 workflow 管理、商店分布和欺诈监测等。

原始数据只是未被开采的矿山，需要挖掘和提炼才能获得对商业目的有用的规律性知识。

这正是数据挖掘这个名字的由来。

因此，从商业角度看，数据挖掘就是按企业的业务目标，对大量的企业数据进行深层次分析，以揭示隐藏的、未知的规律并将其模型化，从而支持商业决策活动的技术。

从商业应用角度刻画数据挖掘，可以使人们更全面地了解数据挖掘的真正含义。

1.1.2 数据挖掘的技术含义 谈到数据挖掘，必须提到另外一个名词：数据库中的知识发现 (Knowledge Discovery in Databases, KDD)，即将来加工的数据转换为有用信息的整个过程。

KDD这个术语首次出现在1989年8月在美国底特律召开的第十一届国际人工智能联合会议的专题讨论会上。

随后，在近十年的发展过程中，KDD专题讨论会逐渐发展壮大。

1999年在美国圣地亚哥举行的第五届KDD国际学术大会，参加人数近千人，投稿280多篇，近年来的国际会议涉及的范围更广，如数据挖掘与知识发现(Data Mining and Knowledge Discovery, DMKD)的基础理论、新的发现算法、数据挖掘与数据仓库及OLAP的结合、可视化技术、知识表示方法、Web中的数据挖掘等。

此外，IEEE、ACM、IFIS、VLDB、SIGMOD等其他学会、学刊也纷纷把DMKD列为会议议题或出版专刊，成为当前国际上的一个研究热点。

关于KDD和Data Mining的关系，有许多不同的看法。

我们可以从这些不同的观点中了解数据挖掘的技术含义。

1) 将KDD看成数据挖掘的例子之一 这一马点在数据挖掘发展的早期比较流行，并且可以在许多文献中看到这种说法，其主要观点是数据库中的知识发现仅是数据挖掘的一个方面，因为数据挖掘系统可以在关系数据库(Relational Database)、事务数据库(Transaction Database)、数据仓库(Data Warehouses)、空间数据库(Spatial Database)、文本数据(Text Data)以及诸如Web等多种数据组织形式中挖掘知识。

从这个意义上来说，数据挖掘就是从数据库、数据仓库以及其他数据存储主式中挖掘有用知识的过程。

2) 数据挖掘是KDD不可缺少的一部分 为了统一认识 Fayyad、Piatetsky-Shapiro 和 Smyth 在1996年出版的权威论文集《知识发现与数据进展》中给出了KDD和数据挖掘的最新定义：KDD是从数据中辨别有效的、新颖的、潜在有用的、最终可理解的模式的过程；数据挖掘是KDD中通过特定的算法在可接受的计算效率限制内生成特定模式的一个步骤。

这种观点得到了大多数学者的认同。

它将KDD看做是一个广义的范畴，包括数据清理、数据集成、数据选择、数据转换、数据挖掘、模式

<<数据挖掘原理、算法及应用>>

生成及评估等一系列步骤，这样，我们可以把KDD看做是由一些基本功能构件组成的系统化协同工作系统。而数据挖掘则是这个系统中的一个关键的部分。

源数据经过清理和转换等步骤成为适合挖掘的数据集，数据挖掘在这种具有固定形式的数据集上完成知识的提炼，最后以合适的知识模式用于进一步的分析决策工作。

将数据挖掘作为KDD的一个重要步骤看待，可以使更容易聚焦研究重点，有效解决问题。

目前，人们对于数据挖掘算法的研究基本属于这样的范畴。

3) KDD与Data Mining的含义相同有些人认为，KDD与Data Mining只是对同一个概念的不同叫法。

事实上，在现今的许多文献（如技术综述等）中，这两个术语仍然不加区分地使用着。

有人说，KDD在人工智能界更流行，而Data Mining在数据库界使用更多，也有人说，一般在研究领域称之为KDD，在工程领域则称之为数据挖掘。

实际上，数据挖掘的概念有广义和狭义之分。

广义的定义是，数据挖掘是从大型数据集（可能是不完全的、有噪声的、不确定性的、各种存储形式的）中，挖掘隐含在其中的、人们事先不知道的、对决策有用的知识的过程。

狭义的定义是，数据挖掘是从特定形式的数据集中提炼知识的过程。

综上所述，数据挖掘概念可以从不同的技术层面上来理解，但是其核心仍然是从数据中挖掘知识。

所以，有人说叫知识挖掘更合适。

本书也在不同的章节使用数据挖掘的广义或狭义概念，读者要注意根据上下文加以区分。

当然，在可能混淆的地方，我们将明确说明。

1.2 数据挖掘的历史及发展数据挖掘可以看做是信息技术悄然演化的结果。

像其他新技术的发展历程一样，数据挖掘也必须经过概念提出、概念接受、广泛研究和探索、逐步应用和大量应用等阶段。

从目前的现状看，大部分学者认为数据挖掘的研究仍然处于广泛研究和探索阶段。

<<数据挖掘原理、算法及应用>>

编辑推荐

《数据挖掘原理、算法及应用》由西安电子科技大学出版社出版。

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>