

<<医学数据挖掘>>

图书基本信息

书名：<<医学数据挖掘>>

13位ISBN编号：9787040190786

10位ISBN编号：7040190788

出版时间：2006-7

出版时间：高等教育出版社（蓝色畅想）

作者：崔雷

页数：221

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>

## <<医学数据挖掘>>

### 内容概要

本书是国内第一部关于医学数据挖掘的教材。

包括基础篇、核心篇和应用篇三个部分。

基础篇介绍数据挖掘的基本概念和理论，核心篇介绍数据挖掘的主要算法和工具，应用篇则分别介绍数据挖掘在医学临床、分子生物学、预防医学、医院管理、文本和web挖掘中的具体应用。

本书首先强调数据挖掘的基本概念和基本方法，重点介绍该领域的基本概念、基本过程和方法；各种算法以介绍其适用条件和原理为主，尽量少涉及具体算法的数学公式。

其次，本书以应用为主，介绍数据挖掘方法在医学研究和服务中的应用实例，为学生今后进一步从事这一方面的深入研究提供基础。

最后，本书在内容组织上力求全面系统，突出重点。

由浅入深、突出交叉学科的特色的同时，注重所介绍知识的层次，适合不同水平读者的学习需要。

## &lt;&lt;医学数据挖掘&gt;&gt;

## 书籍目录

基础篇 第一章 概述	第一节 数据挖掘与知识发现的基本概念	一、数据挖掘的产生
二、什么是数据挖掘和知识发现	三、数据挖掘的知识表示	第二节 知识发现和数据挖掘的步骤、算法与工具
一、知识发现和数据挖掘的基本步骤	二、知识发现和数据挖掘的算法	
三、数据挖掘的工具	第三节 数据挖掘系统的体系结构	一、数据库管理模块
挖掘前处理模块	三、挖掘操作模块	二、知识输出模块
四、模式评估模块	五、知识输出模块	第四节 数据挖掘和知识发现的应用
一、数据挖掘和知识发现在科学研究中的应用	二、数据挖掘和知识发现在商业上的应用	三、Web挖掘
三、Web挖掘	第二章 数据挖掘的对象	第一节 关系型数据库
一、关系型数据库的定义	二、关系组成与性质	三、关系型数据库的数据挖掘
第二节 数据仓库	一、数据仓库的定义和结构	二、数据仓库的特征
三、多维数据模型	四、数据仓库的数据挖掘	第三节 文本数据库
一、语种识别	二、特征提取	三、文本聚类
四、文本分类	第四节 复杂类型数据库	一、空间数据库
二、Web数据库	三、时序数据库	第三章 数据挖掘的步骤
第一节 跨行业数据挖掘过程标准	一、产生背景	二、CRISP-DM过程模型
三、数据挖掘工具	第二节 业务理解	一、确定商业目标
二、状况评估	三、确定数据挖掘目标	四、建立项目计划
第三节 数据准备	一、理解数据	二、数据选择
三、数据清洗	四、数据转换	五、数据集成
六、数据归约	第四节 模型建立和评估	一、模型的种类
二、模型的精确度	三、模型评估核心篇	第二章 关联规则与关联分析
第三章 聚类分析	第六章 决策树	第七章 内容概括相关技术
第八章 人工神经网络	第九章 遗传算法	第十章 粗糙集理论及其应用应用篇
第十一章 数据挖掘在临床领域中的应用	第十二章 数据挖掘在分子生物学领域中的应用	第十三章 数据挖掘在预防医学领域中的应用
第十四章 时间序列数据挖掘及其在医院管理中的应用	第十五章 文本挖掘及其在生物医学领域中的应用	第十六章 Web挖掘
第十七章 数据挖掘工具概述参考文献		

## 章节摘录

实表，即“死亡现象”，其中包含四个维的关键字，即地区码、时间码、性别码和死因码；两个对事实的度量，即死亡数和死亡年龄。

在星型模式中每一个维只用一个表表示，每个表包含一组属性。

比如，地区维表包括大区、省份、地市以及县等属性。

四、数据仓库的数据挖掘大部分情况下，数据挖掘都要先把数据从数据仓库中拿到数据挖掘库或数据集市（DataMart）中。

数据集市或数据挖掘库是完整的数据仓库的一个逻辑子集，常常在一个组织的部门级为满足决策分析而建立，所以俗称为“部门级的数据仓库”。

所有数据集市有机地结合在一起就构成了数据仓库。

在没有建立数据仓库的情形下，为了进行数据挖掘可以把一个或者几个事务数据库导入一个只读数据库里，将其看作数据集市进行挖掘，但是数据仓库的建立会给数据挖掘带来不少的便利。

从组织角度看，数据仓库的数据是集成化、综合性的。

数据挖掘要实现的是一个组织全局模式的知识发现。

因此，基于数据仓库的数据挖掘能更好地满足高层战略决策的要求，而且数据仓库建立的过程中本身就包括了对数据的预处理过程。

数据在装载之前，需要进行清洗、集成、变换和归约等处理，有些还要进行初步的分析处理。

在数据挖掘的整个流程中，数据的预处理过程是最费时、费力的，往往占到整个挖掘时间消耗量的一半以上。

以数据仓库为挖掘对象，数据的噪声问题、缺失值问题、不一致性问题得到解决，可以提高挖掘的成本效果性。

数据仓库是面向主题的，对人们关心的问题具有强大的决策分析支持功能。

因此其体系结构能够保证实时进行查询分析活动；而一般的联机事务处理系统则主要针对更新的实时性，对查询分析功能针对性较弱。

除此之外，数据仓库采用更加接近人类思维的数据模型，拥有可视化的用户界面，加上对查询功能的强大支持，能使数据挖掘效率更高，并在挖掘过程中做到实时交互，使决策者有可能发现更深入、更有价值的知识。

但是，数据仓库在给挖掘带来便利的同时，也给数据挖掘带来了挑战。

首先，大型的数据仓库来自于异种数据源（HeterogeneousDatabase）。

这些数据源来源于庞大的、分布式的和异种数据库，具有结构化、半结构化和非结构化的数据语义。

比如遗产数据库（legacyDatabase），由关系型数据库、层次数据库、电子表格、网络数据库、多媒体数据库等组合在一起，其中的数据规律一般很难被简单的查询分析发现。

这要求数据挖掘的算法必须更快、更有效。

其次，数据仓库包含了大量陈旧历史的数据。

这种数据在时间轴上的特征也增加了数据挖掘的难度。

第三节 文本数据库 随着信息技术的不断进步，对超大文本集合的电子存储已经成为可能，于是形成了文本数据库（TextDatabase）或者称全文数据库（Full-textDatabase）。

文本数据库包含计算机能够读取的整个文本，或者从文献里有目的节选的摘要，比如新闻数据库、电子邮件数据库、学术期刊数据库等。

由于整个社会信息化程度的不断提高，人们越来越依赖使用电子文本记录和发表信息，同时为方便查询和检索，历史沉积下来的各类印刷文本正在不断地被电子化，再加上网络技术的普及，以电子文本为载体保存下来的信息越来越多。

大多数文本数据库所存放的数据都是半结构化数据（semi-structured data），即它们既不是完全结构化也不是完全无结构的。

比如，一个文档可能包含结构字段，诸如标题、作者、出版时间、长度和图书分类号等；也可能包含

## <<医学数据挖掘>>

大量无结构的文本成分，诸如摘要和内容等。

近年来在数据库研究领域对半结构化数据集进行建模和操作已有许多研究成果。

此外信息检索技术，如文本索引方法，也已应用到非结构化文档的处理上。

人们使用文本数据库主要基于三个基本的目的：首先，用户需要便利地获得全文文本，即文献检索。

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>