DW:

1. 数据仓库的定义

特征

1. 数据库与数据仓库OLTP与OLAP

区别与联系

1. 数据仓库的体系结构

包括哪几部分

1. 数据立方体

概念及计算

1. 多维数据模型

几种类型的比较？

1. OLAP的基本操作

哪几种？

DM1 引论

* 1. 为什么进行数据挖掘？

数据信息知识之间的区别与联系

* 1. 什么是数据挖掘？

知识发现KDD的核心与步骤

* 1. 可以挖掘什么类型的数据？

数据库数据 数据仓库 事务数据

其它类型的数据：时间相关数据或序列数据，数据流，空间数据，超文本和多媒体数据，万维网等等

* 1. 可以挖掘什么类型的模式？

一般功能/常用功能

并非所有东西都是数据挖掘!

* 1. 使用什么技术？

数据库/数据仓库/统计学/机器学习

DM2认识数据(作业)

2.1 数据对象与属性类型

标称/二元(对称/非对称)/序数/数值(区间/比率的区别)

2.2 数据的基本统计描述

动机/中心趋势度量:均值、中位数、众数（模）等/离散程度度量:四分位数、四分位数极差、方差、标准差等

2.4 度量数据的相似性和相异性

二元属性的相异性计算/数据属性的相异性计算

DM3数据预处理(作业)

3.1 数据预处理

为什么进行数据预处理/数据预处理有哪些形式

3.2 数据清理

如何处理空缺值(均值/中位数)/如何处理噪声(分箱)

3.4 数据归约

维归约与数据归约的区别与联系:PPT19页

特征选择与特征提取的区别与联系:PPT20/22/23页

K-L变换与线性判别分析的区别与联系:PPT36页

主成分分析与线性判别分析的区别与联系/性质、不足:PPT36/37/38/44/45/46页

DM4 关联规则

4.1 基本概念

支持度的概念与计算/置信度的概念与计算/什么是强关联规则/关联规则挖掘的步骤

4.2 频繁项集挖掘方法

Apriori算法与P-growth算法的区别与联系:PPT21/37/38页

Apriori算法的计算

4.3 模式评估方法

强关联规则的框架模式/如何扩充

DM5 分类

5.1 基本概念

有监督与无监督学习的区别与联系

分类与预测的区别与联系

常见的分类方法有哪些

5.2 决策树归纳

什么是决策树、由决策树提取规则

5.3 贝叶斯分类

不足及解决

5.5 模型评估与选择

评估分类器性能的度量:PPT21/37/38页

理解评估一个分类器准确率的方法:PPT58页

* 1. 支持向量机

支持向量机的目标

什么是支持向量

理解支持向量机的特点与创新

6.4惰性学习法

急切与惰性学习法的区别与联系

k-最近邻分类的步骤

DM6 聚类

聚类分析的目标/典型应用

划分层次基于密度三种方法的原理及优缺点

层次聚类方法的计算