

DW:

1. 数据仓库的定义
哪些特征: PPT 4-6 页
2. 数据库与数据仓库/OLTP 与 OLAP
区别与联系: PPT 9-13 页
3. 数据仓库的体系结构
包括哪几部分: PPT 16 页
4. 数据立方体
概念及计算: PPT 32-36 页
5. 多维数据模型
几种类型的比较
6. OLAP 的基本操作
哪几种: PPT 44 页

DM1 引论

- 1.1 为什么进行数据挖掘?
数据信息知识之间的区别与联系: PPT 5 页
- 1.2 什么是数据挖掘?
数据库中的知识发现 KDD 的核心与步骤: PPT 8-9 页
- 1.3 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
- 1.4 可以挖掘什么类型的模式?
一般功能/常用功能: PPT 17-21 页
并非所有东西都是数据挖掘!
- 1.5 使用什么技术?
数据库/数据仓库/统计学/机器学习: PPT 24-26 页

DM2 认识数据

- 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): PPT 6-7 页
- 2.2 数据的基本统计描述(作业)
动机/中心趋势度量:均值、中位数、众数/离散程度度量:四分位数、四分位数极差、方差、标准差: PPT 9-11 页
- 2.4 度量数据的相似性和相异性(作业)
二元属性的相异性计算/数值属性的相异性计算: PPT 37-41 页

DM3 数据预处理

- 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: PPT 3-4 页
- 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): PPT 6-10 页
- 3.4 数据归约
维归约与数据归约的区别与联系: PPT 19 页
特征选择与特征提取的区别与联系: PPT 19,20,23 页
K-L 变换与主成分分析的区别与联系: PPT 36 页
主成分分析与线性判别分析的区别与联系: PPT 36/37/38/44/45/46 页

DM4 关联规则

4.1 基本概念

基本概念：PPT 3-8 页/关联规则挖掘的步骤：PPT 11-12 页

4.2 频繁项集挖掘方法

Apriori 算法与 P-growth 算法的区别与联系：PPT 21/37/38 页

Apriori 算法(计算)

4.3 模式评估方法

强关联规则的框架模式/如何扩充：PPT 41 页

DM5 分类

5.1 基本概念

有监督与无监督学习的区别与联系：PPT 10 页

分类与预测的区别与联系：PPT 2-5 页

分类的步骤：PPT 6-9 页

常见的分类方法有哪些

5.2 决策树归纳

什么是决策树：PPT 11 页/由决策树提取规则：PPT 35 页

5.3 贝叶斯分类

不足及解决：PPT 38-39 页

5.5 模型评估与选择

评估分类器性能的度量：PPT 52-55 页

评估分类器准确率的方法：PPT 58 页

5.6 支持向量机

支持向量机的目标：PPT 6 页

什么是支持向量：PPT 6 页

理解支持向量机的特点与创新：PPT 21-22 页

6.4 惰性学习法

急切与惰性学习法的区别与联系：PPT 51 页

k-最近邻分类(计算)

DM6 聚类

聚类分析的目标/典型应用：PPT 2 页

划分层次基于密度三种方法的一般特点：PPT 9 页

层次聚类方法(计算)

题型:填空/选择/简答/计算