

《关联规则七章》PPT课件

制作人：PPT创作者
时间：2024年X月

目录

- 第1章 简介
- 第2章 关联规则挖掘的数据预处理
- 第3章 Apriori算法
- 第4章 FP-growth算法
- 第5章 关联规则挖掘的进阶技术
- 第6章 总结与展望
- 第7章 附录

● 01

第1章 简介

关联规则的概念

关联规则是一种数据挖掘技术，用于发现数据中的相关性和模式。通过关联规则可以找到数据集中的频繁项集和关联规则。

关联规则的应用

市场营销分析

分析顾客购买行为
和促销策略制定

医学领域应用

发现病症和治疗方
法关联关系

关联规则算法

Apriori算法

逐层搜索频繁项集
发现关联规则

FP-growth算法

使用树形结构存储数据
减少候选集生成

01 支持度和置信度

评价关联规则好坏的重要指标

02 Lift值

判断关联规则是否具有实际意义

03

关联规则的应用场景

关联规则不仅局限于市场营销和医学领域，还可以应用于推荐系统、网络安全等多个领域。通过挖掘数据间的关联关系，可以为决策提供重要依据。



• 02

第二章 关联规则挖掘的数据 预处理

数据清洗

数据清洗是关联规则挖掘中的重要步骤，需要清除缺失值和异常值，处理重复数据和不一致数据，确保数据质量

数据变换

标准化处理

对数据进行标准化处理，统一数据的尺度和范围

离散化处理

将连续型数据转换为离散型数据，便于关联规则挖掘的分析

特征选择

相关性分析

通过相关性分析等方法，选择与关联规则挖掘相关的特征

特征筛选

筛选出对模型准确性有影响的特征，提高挖掘效率

01 训练集

用于模型训练，学习数据的关联规则

02 测试集

用于验证模型的泛化能力和准确性

03

数据预处理总结

数据清洗

清除缺失值和异常值
处理重复数据和不一致数据

数据变换

标准化处理
离散化处理

特征选择

相关性分析
特征筛选

数据集划分

训练集
测试集

关联规则挖掘的数据预处理

数据预处理是关联规则挖掘的首要步骤，包括数据清洗、数据变换、特征选择和数据集划分，通过这些步骤可以准备好数据集用于关联规则挖掘



● 03

第3章 Apriori算法

Apriori算法原理

Apriori算法是一种经典的关联规则挖掘算法，其原理是基于先验知识进行频繁项集的生成，通过逐层搜索来减少候选项集的生成数量。



Apriori算法流程

初始化频繁一
项集

第一步

剪枝操作

第三步

组合产生下一
层候选项集

第四步

生成候选项集

第二步

Apriori算法优缺点

优点

简单易懂
实现便捷

缺点

需要多次扫描数据集
计算开销较大

Apriori算法改进

为了提高Apriori算法的性能，可以通过剪枝策略和数据结构的优化来优化算法。此外，还可以使用并行化和分布式计算来加速算法的运行，进一步提高效率。

Apriori算法改进方法

剪枝策略优化

减少候选项集数量

并行化处理

加速算法运行

数据结构优化

提高算法效率

● 04

第4章 FP-growth算法

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：
<https://d.book118.com/755344214104011131>