

**课 程 实 验 报 告**

**课程名称： 大数据分析**

**专业班级：**

**学 号：**

**姓 名：**

**指导教师：**

**报告日期：**

**计算机科学与技术学院**

**目录**

[实验三 关系挖掘实验 1](#_Toc90230186)

[**3.1 实验目的** 1](#_Toc90230187)

[**3.2 实验内容** 1](#_Toc90230188)

[**3.3 实验过程** 1](#_Toc90230189)

[3.3.1 编程思路 1](#_Toc90230190)

[3.3.2 遇到的问题及解决方式 2](#_Toc90230191)

[3.3.3 实验测试与结果分析 2](#_Toc90230192)

[**3.4 实验总结** 2](#_Toc90230193)

# 实验三 关系挖掘实验

## **3.1 实验目的**

1、加深对Apriori算法的理解,进一步认识Apriori算法的实现；

2、分析Apriori算法的缺点，使用pcy等变式对Apriori算法进行优化。

## **3.2 实验内容**

**必做：**

1. 实验内容

编程实现Apriori算法，要求使用给定的数据文件进行实验，获得频繁项集以及关联规则。

2. 实验要求

以Groceries.csv作为输入文件。

输出1~3阶频繁项集与关联规则，各个频繁项的支持度，各个规则的置信度，各阶频繁项集的数量以及关联规则的总数。

固定参数以方便检查，频繁项集的最小支持度为0.005，关联规则的最小置信度为0.5。

**加分项：**

1. 实验内容

在Apriori算法的基础上，要求使用pcy或pcy的几种变式multiHash、multiStage等算法对二阶频繁项集的计算阶段进行优化。

2. 实验要求

以Groceries.csv作为输入文件。

输出1~4阶频繁项集与关联规则，各个频繁项的支持度，各个规则的置信度，各阶频繁项集的数量以及关联规则的总数。

输出pcy或pcy变式算法中的vector的值，以bit位的形式输出。

 参数不变，频繁项集的最小支持度为0.005，关联规则的最小置信度为0.5。

## **3.3 实验过程**

### 3.3.1 编程思路

### 3.3.2 遇到的问题及解决方式

### 3.3.3 实验测试与结果分析

## **3.4 实验总结**