
实验三 关系挖掘实验

实验目的

- 1、理解 Apriori 算法思想与流程；
- 2、应用 Apriori 思想解决问题；
- 3、PCY 算法解决问题。

实验内容

以 Groceries.csv 作为输入文件，完成频繁项挖掘，产生关联规则。具体，完成以下任务：

1) 编程实现 Apriori 算法，要求使用给定的数据文件 Groceries.csv 进行实验，获得频繁项集以及关联规则。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。输出各阶频繁项集的数量，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；

2) 在 Apriori 算法的基础上，使用 PCY 或 PCY 的几种变式 multiHash、multiStage 等算法（备注：选择一种算法进行实现就行）对二阶频繁项集的计算阶段进行优化。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。输出各阶频繁项集的数量，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。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。