

第四章 作业

- 1. 给定一个规则  $I_1 \rightarrow I_2$ ，给出该规则的置信度定义，并解释它的含义。
- 2. 简述 Apriori 算法和 FP-tree 算法流程，比较两个算法的优缺点。
- 3. 设最小支持度为 33.3%，最小置信度为 50%。按照 Apriori 算法的步骤，给出每次扫描表 3-1 中的数据库后得到的所有频繁项集。在频繁项集的基础上，产生所有的强关联规则。

表 3-1

TID	故障类型
1	A, B, C, D, E
2	A, B, D, E
3	B, C, D
4	C, D, E
5	A, C, E
6	A, B, D

- 4. 某商店统计了上个季度 10000 笔交易记录，给出如表 3-2 所示的统计信息：

表 3-2

1. {牙刷}在 6000 个事务中出现；
2. {防晒霜}在 5000 个事务中出现；
3. {凉鞋}在 4000 个事务中出现；
4. {太阳镜}在 2000 个事务中出现；
5. {牙刷，防晒霜}在 1500 个事务中出现；
6. {牙刷，凉鞋}在 1000 个事务中出现；
7. {牙刷，太阳镜}在 250 个事务中出现；
8. {牙刷，防晒霜，凉鞋}在 600 个事务中出现。

回答如下问题：

（1）规则“牙刷→防晒霜”与“{牙刷，防晒霜}→{凉鞋}”的置信度分别是多少？

(2) {牙刷}和{防晒霜}是独立的吗?

(3) 计算  $Lift(\text{牙刷}, \text{太阳镜})$ 。

5. 给定如表 3-3 所示的一个事务数据库,画出 FP-tree 树的生成过程。

表 3-3

TID	Itemset
1	a, b, c
2	b, c, d, e
3	a, c, e
4	b, c, d
5	b, c, d, e