## 作业3

**截止日期：11月.18，**  **2020**

**提交要求：**

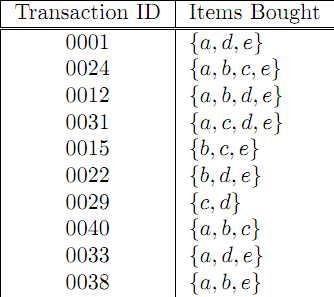
**请 将您的解决方案提交到我们的班级网站。**

**第一部分： 书面作业**

1.考虑表 1 中显示的数据集（min\_sup= 40%， min\_conf = 75%）。。

1. 通过将每个交易 ID 视为市场篮子来查找使用 Apriori 的所有频繁项目集。
2. 使用第 （a） 部分的结果b计算关联规则[a， b = b🡪c]c和[c]a🡪的置信} and {度。置信度是对称测量吗？
3. 列出与以下 metarule 匹配的所有强关联规则（支持 s和置信度 c），其中 X 是代表客户的变量，项目 i表示表示项的变量（例如"A"，"B"等）：

表1.市场篮子交易示例。



2.考虑表 1 中显示的数据集 （min\_sup= 40%）%）。

（a） 使用 FP 增长查找所有 频繁的项目集。 请显示所有FP树和所有条件模式基。

（b） 比较前一和FP增长的效率。

3.表 2 给出了用户-产品评级矩阵。

表 2. 用户-产品 Rating 矩阵

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 产品 1 | 产品 2 | 产品 3 | 产品 4 |
| 用户 1 | 1 | 1 | 5 | 3 |
| 用户 2 | 3 | ? | 5 | 4 |
| 用户 3 | 1 | 3 | 1 | 1 |
| 用户 4 | 4 | 3 | 2 | 1 |
| 用户 5 | 2 | 2 | 2 | 4 |

1. 根据余 弦相似性 列出用户2前3名最相似的用户。
2. 预测用户 2的产品评级 2。

**第二部分： 实验室**

**问题 1**

了解市场篮子分析的使用，以便与客户提出产品购买建议。

数据集包含来自大型超市的交易记录。每笔交易都是由持有会员卡的人做的。为了简单起见，我们将此超市数据中的类别总数限制为 20 个类别。如果客户购买了某个产品的字段值，则字段值为 1，如果客户没有购买，则字段值为 0。名为"事务"的文件有46243个事务的数据。

数据可从类网页获得。

您的子离子应仅包含标有"手持"标记的可交付成果。

市场篮子分析的目标是发现单个产品，或倾向于在交易中同时发生的产品组。企业可以从市场篮子分析中获得的知识用于识别经常一起销售的产品，以确定建议和交叉销售以及向上销售的机会。它还可用于提高促销活动的效率。

在"事务"数据集上运行 Apriori。将"COD"的"类型"设置为"无类型"，将所有其他 20 个类别的"方向"设置为"两者"，将它们的"类型"设置为"标志"。将"最小前置支持"设置为 7%，将"最小置信度"设置为 4%，将建模节点（Apriori 节点）中的"最大前置支持数"设置为4。 通常，您应该通过尝试这些参数的不同值来探索，以查看您得到的规则类型。

* **交手**：模型生成的关联规则的列表。

分别按提升、支持和置信度对规则进行排序，以查看识别的规则。**交手**: **：F或每个情况**，选择前5个规则（注意：确保没有冗余规则在5个规则），并给出2-3行注释。许多规则在逻辑上是多余的，因此在仔细考虑后必须取消这些规则。