**数据挖掘之关联规则挖掘**

**-结果与分析**

**曹文强**

**2120150977**

**北京理工大学计算机学院**

1. **数据预处理结果**

首先将数据转换成csv格式方便读入，如下图

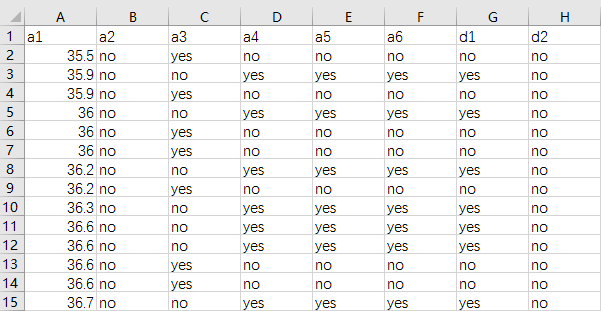


图 1 csv数据

第一个属性是数值属性，我将其简化为二值属性，分别表示发烧和不发烧。

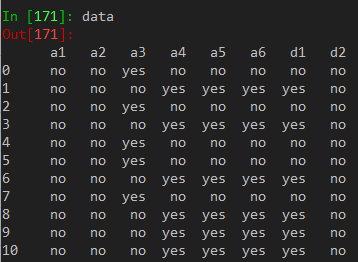


图 2 温度二值化后的数据

然后，将矩阵式的数据转换成事务型的列表数据，如下：

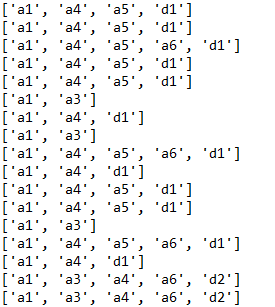


图 3 事务化的数据

1. **频繁规则统计与计算**

我只计算了最大到3频繁项集，计算过程不赘述，结果如下。

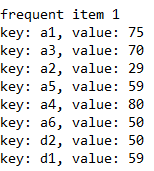


图 4 1-频繁项集

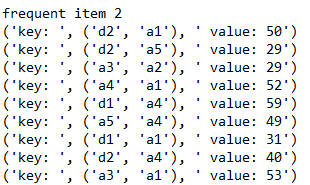


图 5 2-频繁项集

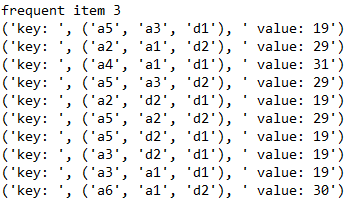
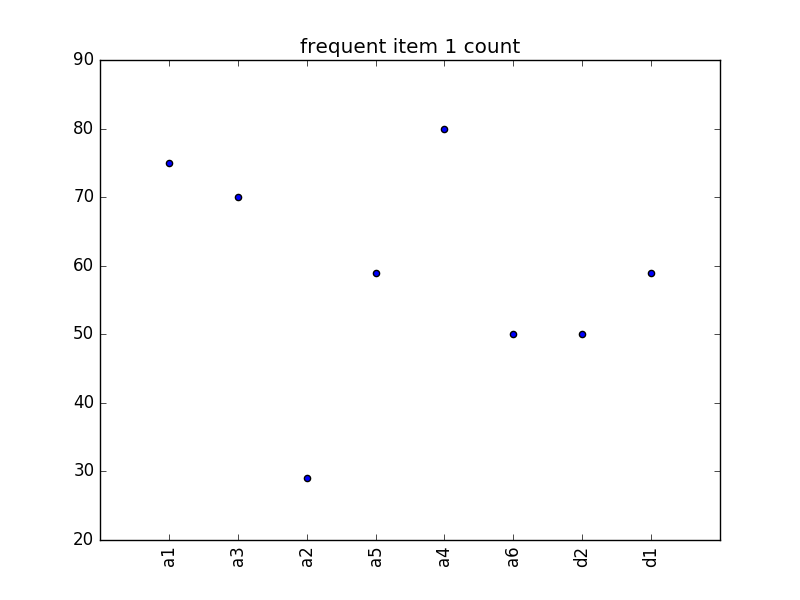
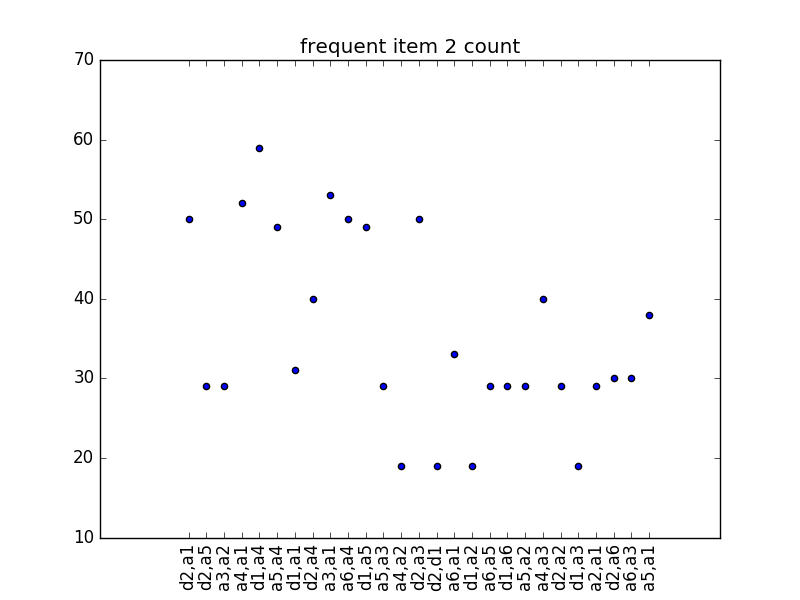
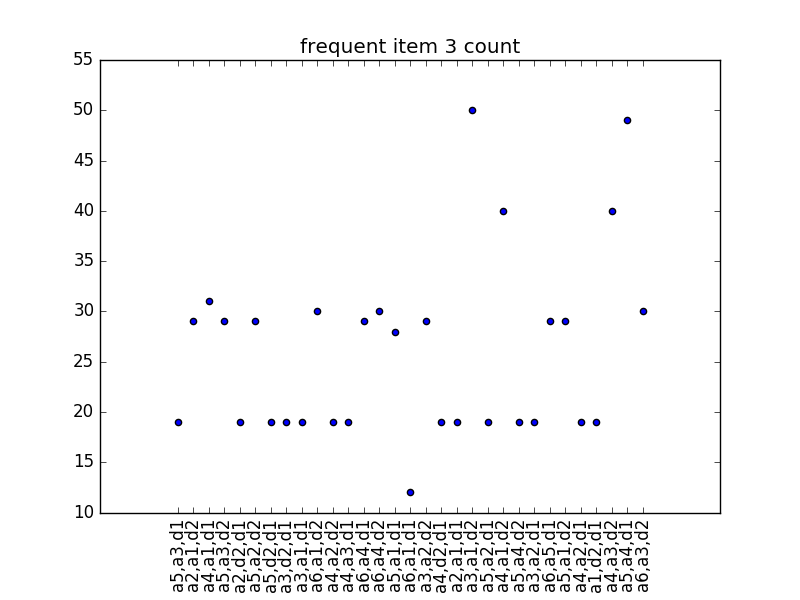


图 6 3-频繁项集

这些频繁项集的可视化目标主要是频数，如下：





****

1. **规则、可信度、支持度**

规则从3-频繁项集中选取，数据中，最后两个属性是结果属性，前6个属性是特征属性，所以这里规则的选取应该是从特征到结果的映射，如下所示：

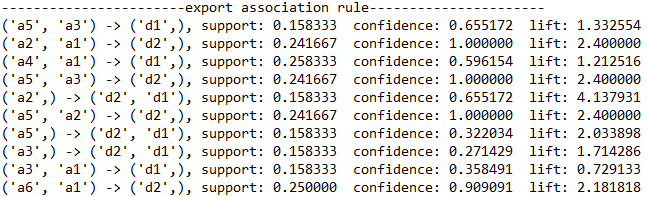


图 7 规则、可信度、支持度

相关标准的计算方法在另一个文档中已经说明。可视化如下(横坐标是规则，纵坐标是lift值)：

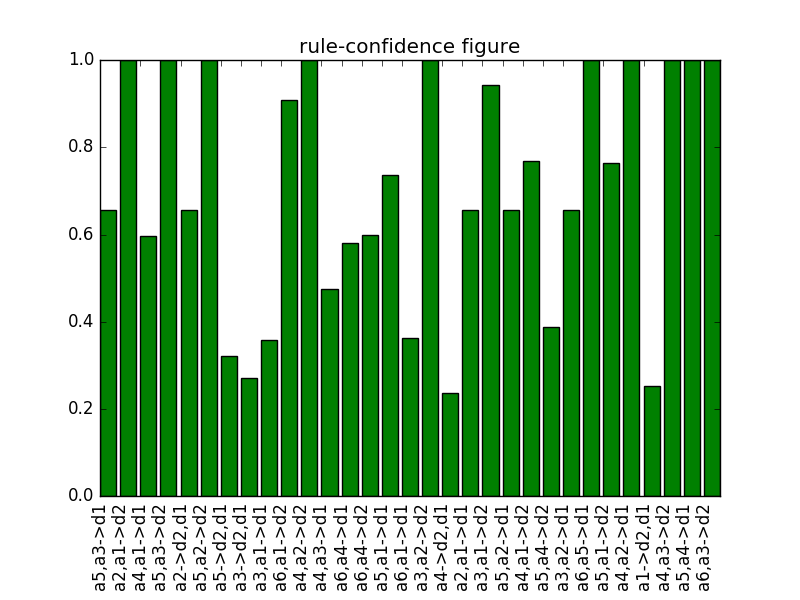


图 8 规则-可信度图示

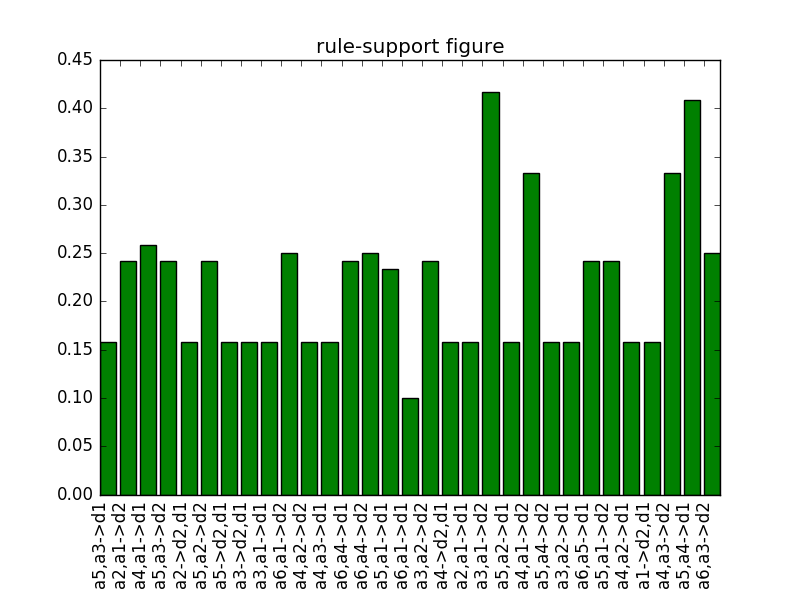


图 9 规则-支持度图示

1. **Lift评价因素**

Lift支持度，也叫做提升度，它是可信度与期望可信度的比值，反映了关联规则中的相关性，其可视化如下：

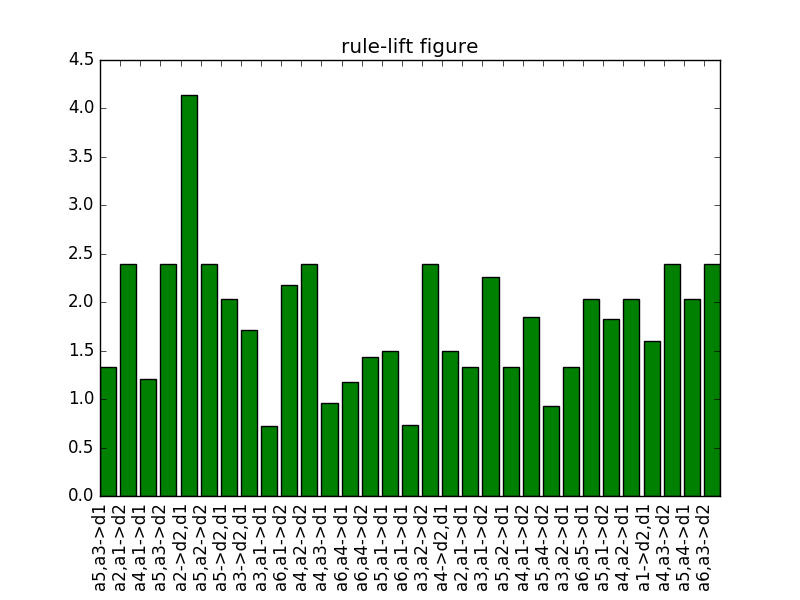


图 10 规则-提升度图示

提升度大于1表示正相关，小于1表示负相关，等于1表示不相关，图中所示，绝大多数规则都是大于1的，并且a2->d1,d2已经超过了4，说明其正相关性非常高。