# 数据挖掘作业二 关联规则集挖掘 报告

姓名：严西敏 学号：2120151051

## 数据源

本次实验采用的数据源为UCI的“急性炎症”数据集。其地址为<http://archive.ics.uci.edu/ml/datasets/Acute+Inflammations>。其共包含120跳条数据，均为对于尿路感染和原发性肾炎的诊断。

## 程序实现

本次实验采用Java语言作为我们的实现语言，项目的总体框架如下所示。

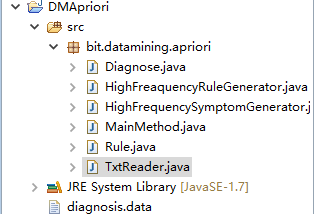


图 程序的总体框架图

下面我们来简单介绍以下每个程序文件的作用。

Diagnose：诊断的POJO实体类。

Rule：规则的POJO实体类。

TxtReader：从diagnosis.data中读取数据，进行数据预处理功能。

HighFrequencySymptomGenerator：发现高频症状。

HighFreaquencyRuleGenerator：发现高频规则。

MainMethod：程序入口，完成一些初始化功能。

## 实验结果

### 对d1 decision: Inflammation of urinary bladder的分析

输入测试集大小：120

共找到17个频繁症状集

共找到2个频繁疾病

共找到17个可以置信的规则，如下所示

< 1no > ----->no 置信程度：0.5604395604395604

< 1no > ----->yes 置信程度：0.43956043956043955

< 2yes > ----->no 置信程度：0.7285714285714285

< 3no > ----->no 置信程度：1.0

< 4no > ----->no 置信程度：0.8360655737704918

< 2no > ----->yes 置信程度：0.8

< 3yes > ----->yes 置信程度：0.7375

< 4yes > ----->yes 置信程度：0.8305084745762712

< 1no 2yes > ----->no 置信程度：1.0

< 1no 4no > ----->no 置信程度：0.8360655737704918

< 1no 2no > ----->yes 置信程度：0.8

< 1no 3yes > ----->yes 置信程度：0.6557377049180327

< 2yes 4no > ----->no 置信程度：1.0

< 2no 3yes > ----->yes 置信程度：1.0

< 3yes 4yes > ----->yes 置信程度：1.0

< 1no 2yes 4no > ----->no 置信程度：1.0

< 1no 2no 3yes > ----->yes 置信程度：1.0

程序所用时间：36毫秒

### 对d2 decision: Nephritis of renal pelvis origin 的分析

输入测试集大小：120

共找到17个频繁症状集

共找到2个频繁疾病

共找到12个可以置信的规则，如下所示

< 1no > ----->no 置信程度：0.7692307692307693

< 2yes > ----->yes 置信程度：0.7142857142857143

< 4no > ----->no 置信程度：0.6557377049180327

< 2no > ----->no 置信程度：1.0

< 3yes > ----->no 置信程度：0.5

< 3yes > ----->yes 置信程度：0.5

< 1no 4no > ----->no 置信程度：0.6557377049180327

< 1no 2no > ----->no 置信程度：1.0

< 1no 3yes > ----->no 置信程度：0.6557377049180327

< 2yes 3yes > ----->yes 置信程度：1.0

< 2no 3yes > ----->no 置信程度：1.0

< 1no 2no 3yes > ----->no 置信程度：1.0

程序所用时间：36毫秒

以上结果都是在置信度为0.4 支持度为0.4的情况下取得。

## 实验结果分析

由上述的实验结果可以看到，本算法可以快速地找出满足置信度和支持度限制的规则。

## 可视化展示

以下是挖掘出来的支持度和置信度的分布图。

