**数据挖掘第二次作业报告**

计算机学院 2120171017 黄焯恒

**一、关联规则挖掘**

**数据集1 Building\_permit.csv**

**二、实现过程**

**1.对数据集预处理，找出频繁项集**

**选取标称属性：**Permit Type，Street Number，Estimated Cost，Proposed Use，Existing Construction Type，然后进行预处理，通过1\_Pre\_process.py程序进行预处理，生成预处理数据Pre\_data.csv

使用Apriori算法找出频繁项集，并保存在fre\_items.txt中，结果如下：

frequent\_item\_set support

frozenset(['Permit\_Type:8']) 0.899

frozenset(['Existing\_Construction\_Type:5.0']) 0.570

frozenset(['Existing\_Construction\_Type:5.0', 'Permit\_Type:8']) 0.508

frozenset(['Proposed\_Use:1\_family\_dwelling']) 0.233

frozenset(['Proposed\_Use:1\_family\_dwelling', 'Existing\_Construction\_Type:5.0']) 0.227

frozenset(['Proposed\_Use:apartments']) 0.216

frozenset(['Proposed\_Use:1\_family\_dwelling', 'Permit\_Type:8']) 0.214

frozenset(['Proposed\_Use:1\_family\_dwelling', 'Permit\_Type:8', 'Existing\_Construction\_Type:5.0']) 0.210

**2.导出关联规则，计算其支持度和置信度**

使用2\_rule\_mining.py导出关联规则，并保存在confident\_order.txt，运行结果如下：

rule support confidence lift

frozenset(['Proposed\_Use:1\_family\_dwelling','Permit\_Type:8']) ==> frozenset(['Existing\_Construction\_Type:5.0']) 0.210 0.985 1.728

frozenset(['Proposed\_Use:1\_family\_dwelling']) ==> frozenset(['Existing\_Construction\_Type:5.0']) 0.227 0.975 1.712

frozenset(['Proposed\_Use:1\_family\_dwelling', 'Existing\_Construction\_Type:5.0']) ==> frozenset(['Permit\_Type:8']) 0.210 0.926 1.030

frozenset(['Proposed\_Use:1\_family\_dwelling']) ==> frozenset(['Permit\_Type:8']) 0.214 0.917 1.020

frozenset(['Proposed\_Use:1\_family\_dwelling']) ==> frozenset(['Existing\_Construction\_Type:5.0', 'Permit\_Type:8']) 0.210 0.903 1.776

frozenset(['Existing\_Construction\_Type:5.0']) ==> frozenset(['Permit\_Type:8']) 0.508 0.892 0.992

**3.使用Lift对规则进行评价**

使用Lift对规则进行评价，并保存在lift\_order.txt，运行结果如下：

rule support confidence lift

frozenset(['Proposed\_Use:1\_family\_dwelling']) ==> frozenset(['Existing\_Construction\_Type:5.0', 'Permit\_Type:8']) 0.210 0.903 1.776

frozenset(['Proposed\_Use:1\_family\_dwelling', 'Permit\_Type:8']) ==> frozenset(['Existing\_Construction\_Type:5.0']) 0.210 0.985 1.728

frozenset(['Proposed\_Use:1\_family\_dwelling']) ==> frozenset(['Existing\_Construction\_Type:5.0']) 0.227 0.975 1.712

frozenset(['Proposed\_Use:1\_family\_dwelling', 'Existing\_Construction\_Type:5.0']) ==> frozenset(['Permit\_Type:8']) 0.210 0.926 1.030

frozenset(['Proposed\_Use:1\_family\_dwelling']) ==> frozenset(['Permit\_Type:8']) 0.214 0.917 1.020

frozenset(['Existing\_Construction\_Type:5.0']) ==> frozenset(['Permit\_Type:8']) 0.508 0.892 0.992