上海理工大学光电信息与计算机工程学院

**《数据挖掘》实验报告**

****

**专　　业 智能科学与技术**

**姓 名　　 高浩琦**

**学　 号 2035060413**

**年　　级 2020级**

**指导教师 孙占全**

**成 绩：**

**教师签字：**

# 实验三 Apriori关联分析算法

1. 实验目的和要求
2. 学会在具体应用上使用Apriori算法
3. 运用Python编程实现Apriori算法计算过程，并对数据进行分析。
4. 通过改变支持度阈值，得到不同的频繁项集和关联规则。
5. 实例分析，读取story.txt文件，每行为一个事件，事件包含多个项，项之间用逗号分割，最小支持度0.2，最小置信度0.7。可自己尝试不同的阈值。

二、实验算法

from apyori import apriori

import numpy as np

def info(results):

    for result in results:

        # for i in result.ordered\_statistics:

        # print(result.items)

        for i in result.ordered\_statistics:

            if i.items\_base and i.items\_add:

                print(list(i.items\_base) , "====>" , list(i.items\_add), "confidence:",i.confidence,"lift:",i.lift)

            else:

                print("skip")

f = open("story.txt",'r',encoding="utf-8")

orders = []

for i in f:

    orders.append(i.strip("\n").split(","))

print(orders)

results = apriori(transactions=orders, min\_support = 0.2 ,min\_confidence = 0.7)

support = np.arange(1,10,1)/10

confidence = np.arange(1,10,1)/10

for i in support:

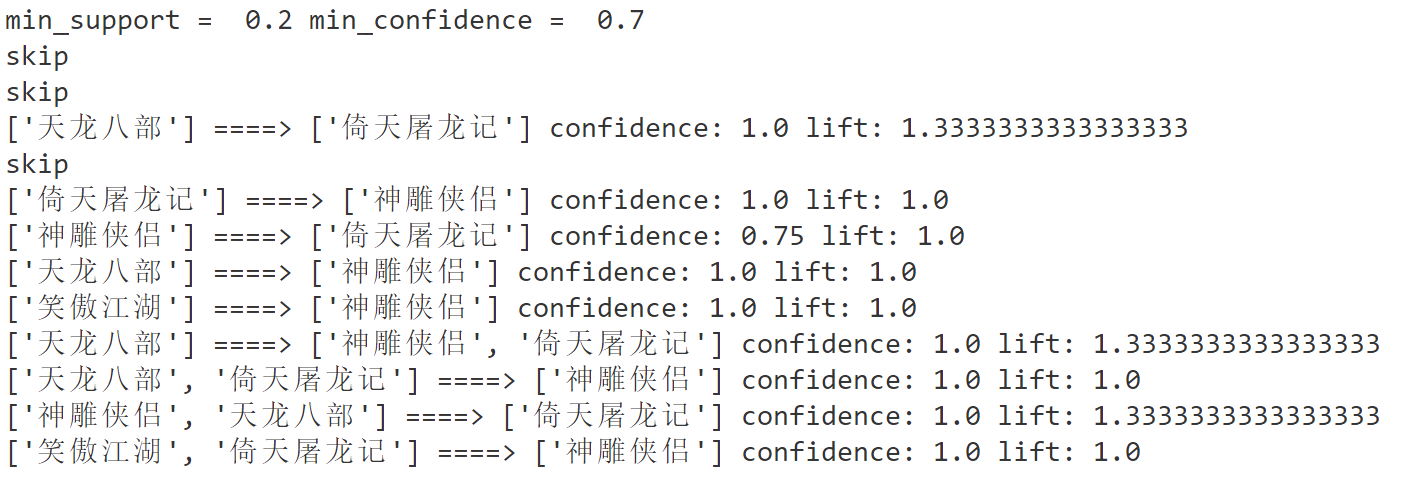
    for j in confidence:

        results = apriori(transactions=orders, min\_support = i ,min\_confidence = j)

        print("min\_support = ",i,"min\_confidence = ",j)

        info(results)

        print()

三、实验过程记录

四、结果与分析

从是按结果来看, Apriori 算法能够广泛应用于各种领域，通过对数据的关联性进行了分析和挖掘，挖掘出的这些信息在决策制定过程中具有重要的参考价值。

五、自评

Apriori 算法的基本思想是：首先找出所有的频集，这些项集出现的频繁性至少和预定义的最小支持度一样。然后由频集产生强关联规则，这些规则必须满足最小支持度和最小可信度。然后使用第1步找到的频集产生期望的规则，产生只包含集合的项的所有规则，其中每一条规则的右部只有一项，这里采用的是中规则的定义。一旦这些规则被生成，那么只有那些大于用户给定的最小可信度的规则才被留下来。为了生成所有频集，还使用了递归的方法。