|  |
| --- |
| 西南民族大学学生实验报告  教学单位：计科学院 实验室名称：BS226 实验时间：2019年12月1日 |
| 实验项目名称：实验6—Apriori 算法 实验成绩： 教师签名： |
| 实验项目报告内容（1、实验目的；2、实验主要过程与结果；3、分析讨论）。  **一、实验目的**   1. 了解Apriori原理； 2. 学会使用Aprioi算法来发现频繁集； 3. 用算法实现从频繁集中挖掘关联规则；   **二、实验主要过程与结果**   1. 使用Apriori算法来发现频繁集    1. 生成候选项集   创建一个用于构建初始集合的函数createC1()程序清单如图1所示，创建一个通过扫描数据集以寻找交易记录子集的函数scanD()，程序清单如图2所示，在函数loadDataSet()中创建一个用于测试的简单数据集，程序清单如图1所示。  在main函数中加入如图3所示语句。  运行程序，结果如图4所示。    图 1    图 2    图 3    图 4   * 1. 组织完整的Apriori算法   构建aprioriGen()函数，输入频繁项集列表 Lk 与返回的元素个数 k，然后输出所有可能的候选项集 Ck，函数清单如图5所示。  构建apriori()函数，找出数据集 dataSet中支持度 >= 最小支持度的候选项集以及它们的支持度。即频繁项集。函数清单如图6所示。  在main函数中加入如图7所示测试代码，运行结果如图8所示。    图 5    图 6    图 7    图 8   1. 从频繁集中挖掘关联规则   构建calcConf()函数，用于计算可信度，函数清单如图9所示。构建rulesFromConseq()函数，用于递归计算频繁项集的规则，函数清单如图10所示。构建generateRules()函数，用于调用前两个函数，生成关联规则，函数清单如图11所示。  在主函数中加入如图12所示测试语句，运行程序，结果如图13所示。    图 9    图 10    图 11    图 12    图 13   1. 示例：发现国会投票中的模式    1. 收集数据：构建美国国会投票记录的事务数据集   构建getActionIds()函数，用于收集美国国会议案中actionId，函数清单如图14所示。  在控制台测试函数，显示actionId，部分结果如图15所示。  构建getTransList()函数，用于填充票务数据，函数清单如图16所示。  在控制台调试函数，如图17所示，无法获得数据。    图 14    图 15    图 16    图 17   * 1. 测试算法：基于美国国会投票挖掘关联规则   在控制台测试函数，如图18所示，无法获取到数据。    图 18   1. 示例：发现毒蘑菇的相似特征   在主函数中写入测试语句，如图19所示。部分运行结果如图20所示。    图 19    图 20  **三、分析讨论**   1. 关联分析是一种在大规模数据集中寻找关系的任务。这些关系可以分为频繁项集（经常出现在一块的物品的集合）和关联规则（暗示两种物品之间可能存在很强的关系）。 2. 使用Apriori原理可以有效的减少数据库上进行检查的集合的数目。 |

注：实验报告的内容及格式可由学院根据学科专业特点确定；全校各专业必须使用学校统一封面。