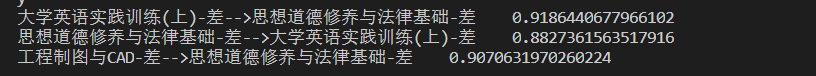
## 遇到的问题1

对字段进行编码之后大概有9000+差，4000+中，1200+良，675个优，可以看出数据分布是非常不平均的，在这种情况下生成的规则是极其不均匀的，只能得到课程差评之间的规则，如下图所示



解决方案

采用抽样的方式来生成规则，考虑到apriori算法的特殊性，它是无监督学习，因此不能简单使用上采样或下采样的方式进行优化。比较可行的方案是对于不同的成绩等级（优良中差）分别生成规则，这样使得每一类在它所包含的数据集中都比较均衡，且能生成不同级别的规则，既能充分利用数据，又解决不平衡问题

数据库名称

early\_warning\_system简称 EWS

科目及成绩信息管理

1. CourseNum课程字典表

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 名称 | 字段名 | 类型长度 | 取值范围 | 缺省值 | PK/FK | 备注 |
| 课程号 | CNNum | varchar(20) | 字母和数字 | 无 | PK | NOT NULL |
| 名称 | CNName | varchar(100) | 汉字 | 无 |  | NOT NULL |
| 学分 | CNCredit | int | 数字 | 无 |  | NOT NULL |
| 学时 | CNHour | int | 数字 | 无 |  | NOT NULL |
| 选课属性 | CNAttribute | varchar(20) | 汉字 | 无 |  | NOT NULL |

1. SubGrade(各科成绩表)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 名 称 | 字段名 | 类型长度 | 取值范围 | 缺省值 | PK/FK | 备注 |
| 学号 | SGNum | int | 数字 | 无 | PK | NOT NULL |
| 姓名 | SGName | varchar(20) | 汉字 | 无 |  | NOT NULL |
| (科目1) | SGSub1 | varchar(64) | 汉字 | 无 |  | NOT NULL |