“中大课程通”应用选课体系建模

1. 问题重述

“中大课程通”是一个集一键选课、查看课表、课程交流等功能的应用。其中一键选课功能需要建立一个选课模型，以对某类可选课程进行量化考虑，找到当前条件下，学生最值得选的课程。

1. 名词定义

可选度：GS（Gain of Selecting），最终权衡课程是否值得选的指标；

校区：CA（CAmpus），某门课程开课的校区；

选中率：SR（Selection Rate），某门课程当前的选中率；

上课时段：ST（Schooltime），上课的时间段，共划分为四个时间段：早课、上午、下午和晚上；

学分：CR（CRedit），某门课程的学分；

开课学院：CO（COllege），开设某门课程的学院，仅对公共课程有效；

1. 模型建立

3.1决策树模型介绍

决策树模型就是为了求出一系列规则，按照规则划分数据，得到预测结果。可以把决策树看做有决策块和终止块组成；

3.2决策树的优点：

1）决策树易于理解和解释.人们在通过解释后都有能力去理解决策树所表达的意义；

2）对于决策树，数据的准备往往是简单或者是不必要的.其他的技术往往要求先把数据一般化，比如去掉多余的或者空白的属性；

3）能够同时处理数据型和常规型属性。其他的技术往往要求数据属性的单一。

4）决策树是一个白盒模型。如果给定一个观察的模型，那么根据所产生的决策树很容易推出相应的逻辑表达式；

5）易于通过静态测试对模型进行评测。表示有可能测量该模型的可信度；

6）在相对短的时间内能够对大型数据源做出可行且效果良好的结果；

7）可以对有许多属性的数据集构造决策树；

8）决策树可很好地扩展到大型数据库中，同时它的大小独立于数据库的大小。

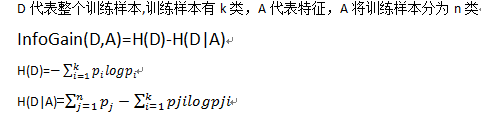
3.3 决策树建模过程：

1）特征选择；

2）决策树生成

3）剪枝

根据信息增益进行特征选择，这就涉及到信息熵的概念，信息熵描述的是数据的混乱程度，信息增益描述的加入特征后数据混乱程度的减小程度，所以信息增益的值等于信息熵减去条件熵的值：



每一步选择信息增益最大的特征作为决策块，最终生成决策树。

3.4 选课体系建模

根据决策树模型，我们统计每类课程各特征与是否选课的关系表如下：

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ID | CA | ST |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

表3-1