学校代码： 11059

学 号：\_\_\_\_\_\_\_\_\_\_



Hefei University

**毕业设计（论文）**

**BACHELOR DISSERTATION**



论文题目： 基于决策树算法的学生成绩分析的研究与实现

学位类别： 工学学士

学科专业： 软件工程专业

作者姓名： 李想

导师姓名： 殷荣网

完成时间： 2018年5月

**本科生毕业设计（论文）诚信承诺书**

**本人慎重承诺和声明**：

我承诺在毕业论文撰写过程中严格遵守学院及系部有关规定，恪守学术规范，在指导教师的安排与指导下独立完成所规定的毕业论文工作，决不弄虚作假，不请别人代做毕业论文或抄袭别人的成果。所撰写的毕业论文是在指导老师的指导下自主完成，文中所有引文或引用数据、图表均注解并说明来源，如有违规行为发生，我愿承担一切责任，接受学院及系部的处理。

学生（签名）：

年 月 日

**基于决策树算法的学生成绩分析的研究与实现**

**中 文 摘 要**

学生成绩分析的研究作为学校教学管理的一个重要组成部分，在学校教育中占有很重要的作用。学生成绩分析的研究与实现系统可分为学生成绩数据录入与保存、C4.5决策树生成、分析与显示学生成绩的影响因素3个部分，其中学生成绩数据录入与保存是学生成绩分析与研究的重要步骤，是后续决策树生成的关键环节。C4.5决策树的实现对学生成绩与影响因素之间联系的构建分析,从所有影响因素中分析出哪一个因素是该学生该课程的成绩主要影响因素。

在本文中作者分析出影响学生课程成绩主要因素的具有如下特征：（1）课后上机学习时间；（2）课前了解；（3）课程学习；（4）平时成绩。

针对学生成绩固有的特征，本文着重介绍了学生成绩分析的痛点与难点，学生成绩管理系统的作用，以及针对学生成绩分析的科学算法；决策树的由来演化，ID3、C4.5决策树的优劣对比等。之后阐述了现在常用的分析方法，并对这些方法进行分析，总结出各种方法的优缺点，然后在此基础上提出采用决策树C4.5算法，并使用java语言编码实现系统，使得该系统具有一定的实用推广价值。

**关键词：****决策树；C4.5；学生成绩分析**

**Research and implementation of student achievement analysis based on decision tree algorithm**

**ABSTRACT**

As an important part of school education management, student achievement analysis plays an important role in school education. The research and implementation system of student achievement analysis can be divided into 3 parts: student achievement data entry, C4.5 decision tree generation, analysis and display of students' achievements, among which student achievement data entry is an important step for student achievement analysis and research, and is the key link for the subsequent decision tree to become a tree. The C4.5 decision tree realizes the analysis and construction of the relationship between the students' achievement and the influencing factors. From all the influencing factors, which factor is the main influencing factor of the student's achievement.

In this paper, the author analyses the main factors that affect the students' curriculum performance as follows: (1) learning time after class; (2) understanding before class; (3) course learning; (4) normal performance.

In view of the inherent characteristics of students' achievements, this paper mainly introduces the pain points and difficulties of student achievement analysis, the role of the student achievement management system, and the scientific algorithm for the analysis of students' performance, the origin of the decision tree and the comparison of the advantages and disadvantages of the ID3 and C4.5 decision trees. Secondly, it introduces the analysis methods commonly used now, analyzes these methods and summarizes the advantages and disadvantages of various methods. On the basis of this, the decision tree C4.5 algorithm is adopted and the system is implemented using the Java language. Finally, the system has been tested. The test results show that the system has a good human-computer interaction mode. It has high analysis accuracy and fast analysis speed. The system has a certain practical value for the students to select the given feature selection and show the results of the analysis to the users.

**KEY WORDS: decision tree; C4.5; student achievement analysis; education**

**目 录**

[中文摘要 I](#_Toc232955056)

[Abstract II](#_Toc232955057)

[目 录 III](#_Toc232955058)

[第一章 绪论 1](#_Toc232955059)

[1.1 课题的来源及意义 1](#_Toc232955060)

[1.2 课题主要研究的问题 1](#_Toc232955061)

[1.3 系统设计的目标及基本思路 1](#_Toc232955062)

[1.3.1 设计目标 1](#_Toc232955063)

[1.3.2 基本思路 1](#_Toc232955064)

1.4 内容简述 2

[第二章 学生成绩分析与研究常用的技术 3](#_Toc232955065)

[2.1 统计分析法 3](#_Toc232955066)

[2.2 决策树分析法 3](#_Toc232955069)

[2.2.1 决策树分类 3](#_Toc232955070)

[2.2.2 ID3 4](#_Toc232955071)

[2.3.3 C4.5 4](#_Toc232955072)

2.3.4 ID与C4.5的区别 4

[第三章 项目分析与设计 5](#_Toc232955073)

[3.1 需求分析 5](#_Toc232955074)

[3.2 功能设计 5](#_Toc232955074)

[3.2.1 超级管理员 6](#_Toc232955072)

[3.2.2 教师 4](#_Toc232955072)

[3.2.3 学生 4](#_Toc232955072)

[3.3 用例图分析 6](#_Toc232955071)

[3.4 业务对象分析 7](#_Toc232955071)

[3.5 数据库设计 4](#_Toc232955072)

[3.5.1 E-R图设计 4](#_Toc232955072)

[3.5.2 数据库表设计 4](#_Toc232955072)

[3.6 C4.5决策树的设计与分析 4](#_Toc232955072)

[3.6.1 创建决策树 4](#_Toc232955072)

[3.6.2 匹配决策树 4](#_Toc232955072)

[第四章 项目实现 1](#_Toc232955081)0

[4.1 环境准备 1](#_Toc232955082)0

4.4.1 硬件环境 10

4.4.2 软件环境 10

[4.2 项目展示 1](#_Toc232955087)0

4.2.1 登录页面 10

4.2.2 教师管理 11

4.2.3 学生管理 12

4.2.4 课程管理 13

4.2.5 专业管理 14

4.2.6 班级管理 14

4.2.7 职称管理 15

[4.2.8 管理员信息 1](#_Toc232955091)5

[第五章 系统测试 1](#_Toc232955093)6

[5.1 系统测试 1](#_Toc232955094)6

[5.2 系统登录测试 1](#_Toc232955095)6

5.3 系统功能测试 18

5.2.1 测试过程 18

5.2.2 测试结果 18

[5.3 测试分析 2](#_Toc232955096)2

[结 论](#_Toc232955102) [2](#_Toc232955102)3

[参考文献 2](#_Toc232955103)4

[致 谢 25](#_Toc232955104)

**第一章 绪论**

**1.1 课题的来源及意义**

随着国家日益昌盛，国家对祖国的未来的教育越加注重，而学生的教育就成为了重中之重，学生成绩作为学生学习成长的关键指标，学生、学校教师对学生的成绩管理分析越受到人们的重视。如何有效地进行学生成绩的管理日益成为学生、老师、国家相关部门所关注的焦点。本系统课题针对这一问题，有效实现学生成绩分析自动化管理且提高效率，提供学生与老师的改进方向，提高学生的学习效率，老师的教育成效，以期达到双赢的目的。

在基于决策树算法的学生成绩分析的研究与实现中，如何准确的对学生成绩的好坏影响因素分析是最为重要和关键的问题。学生的课程的课后上机学习时间长短，课前了解的程度，课程学习的优良，平时成绩的优良，以及考试的成绩都是分析学生成绩好坏影响的指标。其中如何生成学生成绩的C4.5决策树成为了首要问题和重要的步骤。

**1.2 课题主要研究的问题**

（1）采用何种方法对学生成绩的分析与研究？

（2）采用哪些计算机算法分析能满足要求？

（3）系统的技术流程是怎样的？

**1.3 系统设计的目标及基本思路**

1.3.1 设计目标

本系统主要应用于学生成绩的分析与研究，根据这一实际应用背景，确定了系统设计的两个目标：实用性、高速性。

1.3.2 基本思路

由于学生考试成绩与学生生活息息相关，可以从平时学生的学习规律、方法、时间等方面分析；其主要特征主要包括：

（1）每日下课后上机学习的平均时间（单位：小时。分为:<5,5-10,>10）。

（2）每科上课前的准备工作（对课程的了解程度。分为：不了解，了解一点，基本了解）。

（3）课堂的学习态度（分为：差，一般，良，优）。

（4）平时的成绩（分为：不合格，合格，良，优）。

**1.4 内容简述**

第一部分主要叙述关于系统课题的背景原因和研究意义，给用户留下基本印象，以方便系统的使用与传播，达到引人入胜的效果。

第二部分主要叙述课题涉及的相关技术，分别阐述了传统技术的基本情况和局限性，以及决策树的分类，又简单介绍了ID3的基本信息和C4.5的信息和应用，重点阐述了C4.5的基本原理C4.5相对ID3的不同和优势。

第三部分主要叙述了系统的需求分析与设计，让用户对系统有基本了解，了解系统有哪些功能，可能对用户的帮助和方便之处，给用户留下深刻印象。

第四部分主要叙述了系统实现的展示，分别阐述展示了每种功能和主要截图，让顾客可以更好的获得系统的相关功能使用。

第五部分主要展现的是系统测试，确保系统的正确运行与安全性。之后对于系统中对于某部分数据得到相应结论进行适当分析，以确保系统的适用性和实用性。

**第二章 学生成绩分析与研究常用的技术**

**2.1统计分析法**

统计分析法就是使用数学方式，继而建立数学模型，然后对通过调查获取的各种数据及资料进行数理统计和分析，最终形成定量的结论。

优点：

方法简单，工作量小。

缺点：

定额的准确性差，可靠性差。不能深层次分析事务的发展权重规律，在此则表示不能分析出学生成绩优劣的影响指标。

**2.2 决策树分析法**

决策树分析法是指分析每个决策或事件（即自然状态）时，能引出两个或多个事件和不同的决策结果，并把这种决策或事件的分支画成图形，这种图形很像一棵树的枝干，故称决策树分析法。

优点：

1. 决策树算法中建立决策树模型的过程比较简单；

2)　决策树模型可以可视化，非常直观；

3) 应用范围较广，可用于分类与回归，而且经常用于做多类别的分类；

4) 能够处理数值型和连续的样本特征。

缺点：

1. 容易过拟合（overfitting）。

2) 不能确保建立全局最优的决策树。

2.2.1 决策树的分类

主要是分为ID3、C4.5算法。

2.2.2 ID3

在决策树学习中，ID3（Iterative Dichotomiser 3）是一种由Ross Quinlan发明的用来从数据集中生成决策树的算法。ID3是C4.5算法的前身，其常用于机器学习和自然语言处理领域。

2.2.3 C4.5

C4.5就是一个决策树算法，它是决策树（决策树也就是做决策的节点间像一棵树一样的组织方式，其实是一个倒树）核心算法ID3的改进算法，所以基本上了解了一半决策树构造方法就能构造它。决策树构造方法其实就是每次选择一个好的特征以及分裂点作为当前节点的分类条件。

2.2.4 C4.5与ID3的区别

1、用信息增益率来选择属性。ID3采取属性用的是子树的信息增益，ID3行使的是熵（entropy， 熵是一种不纯度度量准则），而C4.5用的是信息增益率。

2、在决策树构造过程中修剪，因为一些元素少的节点可能使构造决策树过度拟合（过拟合），并且如果不考虑这些节点，则可能更好。

3、对非离散数据也能处理。

4、能够对不完整数据进行处理。

**第三章** **项目分析与设计**

**3.1 需求分析**

本系统主要面向老师和学生两大对象，老师主要是通过系统查询自己所教导的课程及学习这门课程所有学生的成绩，并得到课程的分析结果和每一名学生的成绩和分析结果，对自己所教导的学生成绩有基本认知，并在之后的教导中吸取经验，再接再励。学生主要是通过系统查询自己所有课程的成绩和分析结果，对自己本身的学习状态有一个基本认知，进而改善学习方法，提高成绩。

**3.2 功能设计**

根据需求分析可得本系统主要有超级管理员、老师和学生三大角色，下面分别展示每一个角色的系统功能。

3.2.1 超级管理员

1、教师信息管理：增加、删除、修改、查询；

2、学生信息管理：增加、删除、修改、查询；

3、专业信息管理：增加、删除、修改、查询；

4、班级信息管理：增加、删除、修改、查询；

5、职称信息管理：增加、删除、修改、查询；

6、课程信息管理：增加、删除、修改、查询；

7、教师课程信息管理：增加、删除、修改、查询；

8、成绩信息管理：增加、删除、修改、查询；

9、学习调查信息管理：增加、删除、修改、查询；

10、分析建议信息管理：增加、删除、修改、查询。

超级管理员拥有所有权限，可以管理所有的信息得增删改查。

3.2.2 教师

1、教师信息管理：查询自身；

2、教师课程信息管理：查询自身；

3、教师课程成绩信息管理：增加、删除、修改、查询；

4、学习调查表信息管理：增加、修改、删除、查询；

5、教师课程学生成绩信息管理：查询自身。

教师账户主要是修改查询自身信息，查询自身课程，为每个课程下的学生增删改查学习调查表信息，以及下面学生的成绩管理查询。

3.2.3 学生

1、学生信息管理：查询自身；

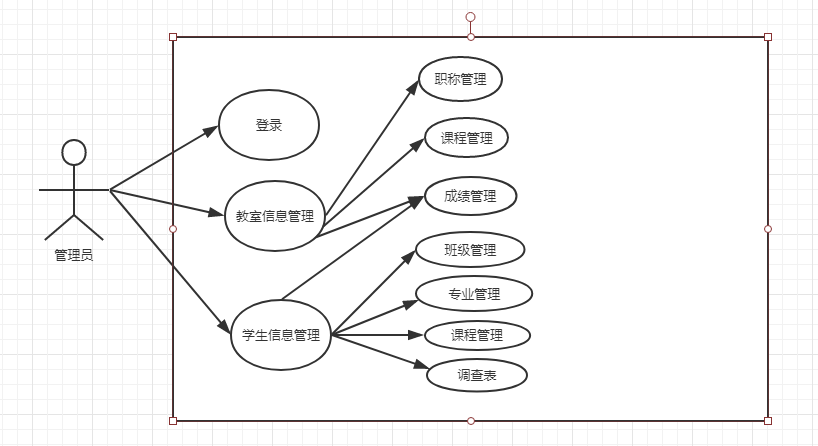
2、学生课程管理：查询自身；

3、学习调查表信息管理：查询自身；

4、教师课程学生成绩分析建议：查询自身。

学生账户主要是修改查询自身信息，查询自身课程，自身学习调查表的平时情况以及自身成绩和分析建议。

**3.3 用例图分析**



3.3.1 教师信息管理用例表

|  |  |
| --- | --- |
| 用例名称 | 教师信息管理 |
| 参与者 | 管理员、教师 |
| 用例描述 | 包括管理员对于所有老师的基本信息的增删改差和老师对自身的查 |
| 前置条件 | 超级管理员登录后进入老师信息管理模块 |
| 后置条件 | 无 |
| 基本路径 | 管理员或老师登录 进入老师信息管理模块 使用相应功能 |

3.3.2 学生信息管理用例表

|  |  |
| --- | --- |
| 用例名称 | 学生信息管理 |
| 参与者 | 管理员、学生 |
| 用例描述 | 包括管理员对于所有学生的基本信息的增删改差和学生对自身的查 |
| 前置条件 | 超级管理员登录后进入学生信息管理模块 |
| 后置条件 | 无 |
| 基本路径 | 管理员或学生登录 进入学生信息管理模块 使用相应功能 |

3.3.3 登录用例表

|  |  |
| --- | --- |
| 用例名称 | 登录 |
| 参与者 | 管理员、学生、老师 |
| 用例描述 | 包括管理员、学生、老师输入账户密码登录 |
| 前置条件 | 所有用户必须使用正确的账户密码 |
| 后置条件 | 无 |
| 基本路径 | 输入账户密码 点击登录 |

**3.4 业务对象分析**

本系统是使用C4.5的决策树针对学生的成绩的分析与研究，便需要建立一个完整的学生成绩管理系统加分析系统，针对学校的主体关系，创建如下业务对象：

1、角色(Role)，账户的角色分类，赋予账户权限；本项目定三类角色，分别为：超级管理员，教师，学生。

2、账户(Account)，账户信息，账户登陆，账户的各类操作，是活动的主题。

3、专业(Major)，学生所在的专业。

4、班级(StudentClass)，学生所在的班级。

5、职称（Positional），教师的职称，如：讲师，教授，副教授。

6、课程(Course)，课程信息；如：语文，数学，英语。

7、教师课程(TeacherCourse)，教师自己拥有的课程信息。

8、成绩(Score)，学生成绩信息，每个成绩对应一个学生与一个教师的课程。

9、学习调查表（Survey），教师对学生平时的成绩打分。（学号，课程编号，课后上机学习时间，课前了解，课堂学习，平时成绩）

10、分析建议（Analyse）,根据决策树分析教师课程下的学习成绩分析，找到影响学生成绩的最大因素，给予学生建议。

对象关系：

1、角色与账户，一对多，一个角色多个账户，一个账户一个角色，例如账户张三角色为学生，角色教师有张教师，李教师等。

2、专业与班级，一对多，一个专业多个班级，一个班级一个专业，比如班级通信A班专业为通信工程，计算机专业有计算机A班，计算机B班等。

3、班级与账户，一对多，一个班级多个学生账户，一个学生账户一个班级，比如学生账户张三的班级为班级通信A班，计算机A班有学生账户李四、王五等。

4、职称与账户，一对多，一个职称多个教师账户，一个教师账户一个职称，比如教师账户王教师的职称为副教授，讲师有教师账户张教师、李教师等。

5、课程与教师账户,多对多，一个课程多个教师账户，一个教师账户多个课程，比如教师账户王教师的课程有高等数学、概率论等，课程语文有教师账户李教师、唐教师等。

6、教师课程与成绩，一对多，一个教师课程多个学生成绩，一个学生成绩一个教师课程，比如张教师-语文有学生李四的成绩、王五的成绩等，赵六的成绩对应何教师-数学。

7、教师账户与学习调查表，一对一，一个教师账户对应一个学习调查表，比如张教师对李四的学习调查。

8、学生账户与学习调查表，一对多，一个学生账户对应多个学习调查表，比如学生张三有2018年3月份的学习调查，2018年4月份的学习调查。

9、成绩与分析建议，一对一，一个教师课程的学生成绩对应一个分析建议，比如张教师-数学下的李四的成绩有的分析建议。

**3.5 数据库分析**

3.5.1 E-R图设计

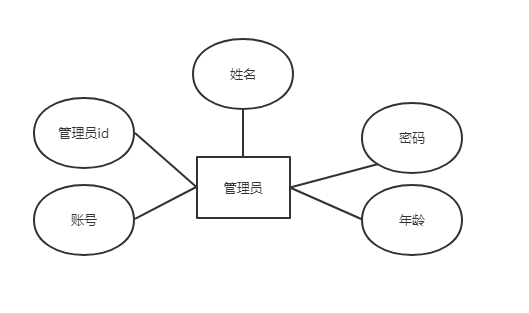


图3.5.1.1 管理员E-R图

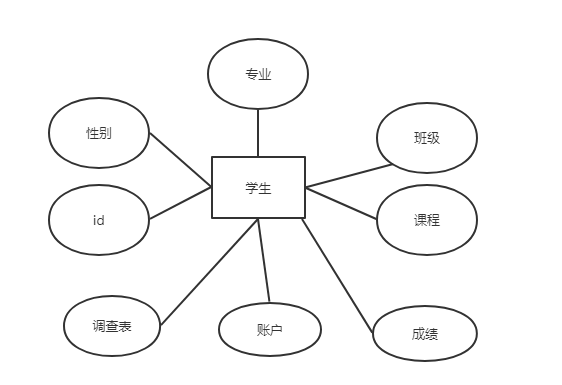


图3.5.1.2 学生E-R图

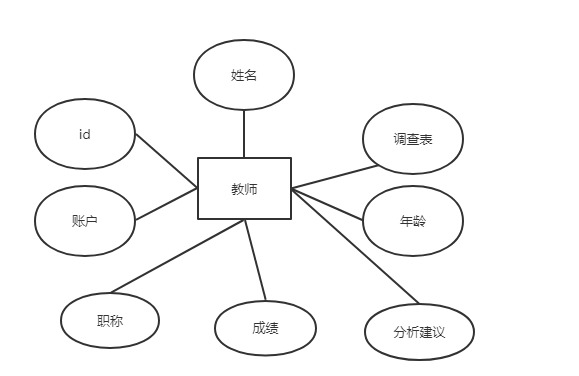


图3.5.1.3 教师E-R图

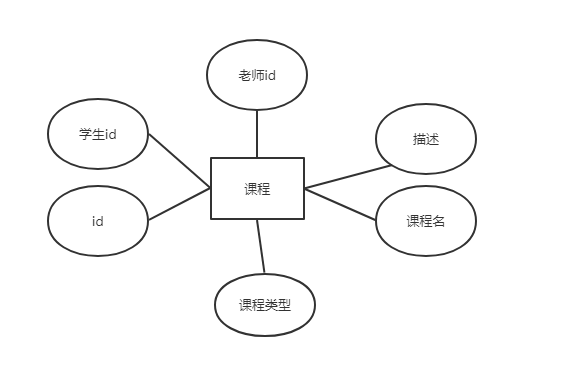


图3.5.1.4 课程E-R图

3.5.2 数据库表设计

根据业务对象分析和E-R图一共设计10个表，分别为账户表(account)、分析建议表（analyse）、课程表(course)、专业表(major)、职称表（positional）、角色表(role)、成绩表(score)、班级表(studentclass)、学习调查表（survey）、教师表课程(teachercourse)。

表3.5.2.1 帐户表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 字段名 | 数据类型 | 数据长度 | 非空 | 说明 | Key |
| account\_id | int | 15 | Y | 账户id | P\_K |
| age | int | 15 | N | 年龄 |  |
| create\_date | datetime | 32 | N | 创建时间 |  |
| login\_date | datetime | 32 | N | 登录时间 |  |
| major\_id | int | 15 | N | 专业id |  |
| password | int | 15 | N | 密码 |  |
| positional\_id | int | 15 | N | 职称id |  |
| realname | varchar | 255 | N | 姓名 |  |
| role\_id | int | 15 | N | 角色id |  |
| sex | varchar | 255 | N | 性别 |  |
| student\_class\_id | int | 15 | N | 学生班级id |  |

表3.5.2.2 分析建议表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 字段名 | 数据类型 | 数据长度 | 非空 | 说明 | Key |
| analyse\_id | int | 15 | Y | 分析id | P\_K |
| business\_id | int | 15 | N | 结论id |  |
| conclusion | varchar | 255 | N | 结论 |  |
| score | int | 15 | N | 分数 |  |
| sub\_type | int | 15 | N | 种类 |  |
| type | int | 15 | N | 种类 |  |
| update\_date | datetime | 32 | N | 更新时间 |  |

表3.5.2.3 课程表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 字段名 | 数据类型 | 数据长度 | 非空 | 说明 | Key |
| course\_id | int | 15 | Y | 课程id | P\_K |
| duty | varchar | 255 | N | 描述 |  |
| name | varchar | 255 | N | 课程名 |  |
| type | int | 15 | N | 类型 |  |
| update\_date | datetime | 32 | N | 更新时间 |  |

表3.5.2.4 专业表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 字段名 | 数据类型 | 数据长度 | 非空 | 说明 | Key |
| major\_id | int | 15 | Y | 专业id | P\_K |
| duty | varchar | 255 | N | 描述 |  |
| name | varchar | 255 | N | 专业名 |  |
| update\_date | datetime | 32 | N | 更新时间 |  |

表3.5.2.5 职称表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 字段名 | 数据类型 | 数据长度 | 非空 | 说明 | Key |
| positional\_id | int | 15 | Y | 职称id | P\_K |
| duty | varchar | 255 | N | 描述 |  |
| name | varchar | 255 | N | 职称名 |  |
| update\_date | datetime | 32 | N | 更新时间 |  |

表3.5.2.6 角色表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 字段名 | 数据类型 | 数据长度 | 非空 | 说明 | Key |
| role\_id | int | 15 | Y | 角色id | P\_K |
| duty | varchar | 255 | N | 描述 |  |
| name | varchar | 255 | N | 角色名 |  |
| update\_date | datetime | 32 | N | 更新时间 |  |

表3.5.2.7 角色表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 字段名 | 数据类型 | 数据长度 | 非空 | 说明 | Key |
| score\_id | int | 15 | Y | 分数id | P\_K |
| score | float | 15 | N | 分数 |  |
| student\_account\_id | int | 15 | N | 学生id |  |
| teacher\_course\_id | int | 15 | N | 老师id |  |
| update\_date | datetime | 32 | N | 更新时间 |  |

表3.5.2.8 学生表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 字段名 | 数据类型 | 数据长度 | 非空 | 说明 | Key |
| student\_class\_id | int | 15 | Y | 学生id | P\_K |
| duty | varchar | 255 | N | 描述 |  |
| name | varchar | 255 | N | 角色名 |  |
| update\_date | datetime | 32 | N | 更新时间 |  |

表3.5.2.9 老师表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 字段名 | 数据类型 | 数据长度 | 非空 | 说明 | Key |
| survey\_id | int | 15 | Y | 调查表id | P\_K |
| before\_class\_understand | int | 15 | N | 课前了解 |  |
| classroom\_leaning | int | 15 | N | 课堂学习 |  |
| operate\_computer\_leaning | int | 15 | N | 课后上机 |  |
| peacetime\_score | int | 15 | N | 平时表现 |  |
| student\_account\_id | int | 15 | N | 学生id |  |
| teacher\_course\_id | int | 15 | N | 老师id |  |
| update\_date | datetime | 32 | N | 更新时间 |  |

表3.5.2.10 老师表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 字段名 | 数据类型 | 数据长度 | 非空 | 说明 | Key |
| teacher\_class\_id | int | 15 | Y | 老师id | P\_K |
| duty | varchar | 255 | N | 描述 |  |
| name | varchar | 255 | N | 角色名 |  |
| update\_date | datetime | 32 | N | 更新时间 |  |

**3.6 C4.5决策树设计与分析**

3.6.1 创建决策树

1,首先创建影响属性集合4类， A={课后上机学习时间 （小时），课前了解，课堂学习，平时成绩}，成绩类别集合4类，L={差，一般，良，优秀}。

比如数据集如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 课后上机学习时间 （小时） | 课前了解 | 课堂学习 | 平时成绩 | 成绩 |
| <5 | 不了解 | 差 | 差 | 差 |
| <5 | 不了解 | 差 | 差 | 差 |
| 5-10 | 了解一点 | 良 | 一般 | 一般 |
| 5-10 | 了解一点 | 一般 | 一般 | 一般 |
| 5-10 | 基本了解 | 差 | 差 | 差 |
| 5-10 | 不了解 | 一般 | 一般 | 一般 |
| >10 | 不了解 | 一般 | 一般 | 一般 |
| >10 | 了解一点 | 良 | 良 | 良 |
| >10 | 基本了解 | 优秀 | 良 | 优秀 |
| <5 | 不了解 | 一般 | 一般 | 一般 |
| <5 | 基本了解 | 差 | 一般 | 一般 |
| <5 | 了解一点 | 良 | 一般 | 良 |

1. 计算类别信息熵（熵越大，不确定性就越大）

Info(D)=

-（差/总\*log2(差/总)）

-（一般/总\*log2(一般/总)）

-（良/总\*log2(良/总)）

-（优秀/总\*log2(优秀/总)）

即：

Info(D)=-3/12\*log2(3/12)-6/12\*log2(6/12)-2/12\*log2(2/12)-1/12\*log2(1/12)

1. 计算每个属性的信息熵（信息熵越大，类别越不纯）

Info(课后上机学习时间 （小时）)=

5/12\*[-2/5\*log2(2/5)-2/5\*log2(2/5)-1/5\*log2(1/5)]

+4/12\*[-3/4\*log2(3/4)-1/4\*log2(1/4)]

+3/12\*[-1/3\*log2(1/3)-1/3\*log2(1/3)-1/3\*log2(1/3)]

Info(课前了解)=

5/12\*[-2/5\*log2(2/5)-3/5\*log2(3/5)]

+4/12\*[-2/4\*log2(2/4)-2/4\*log2(2/4)]

+3/12\*[-1/3\*log2(1/3)-1/3\*log2(1/3)-1/3\*log2(1/3)]

Info(课堂学习)=

4/12\*[-3/4\*log2(3/4)-1/5\*log2(1/4)]

+4/12\*[-4/4\*log2(4/4)]

+3/12\*[-1/3\*log2(1/3)-2/3\*log2(2/3)]

+1/12\*[-1/1\*log2(1/1)]

Info(平时成绩)=

3/12\*[-3/3\*log2(3/3)]

+7/12\*[-6/7\*log2(6/7)-1/7\*log2(1/7)]

+2/12\*[-1/2\*log2(1/2)-1/2\*log2(1/2)]

4,计算信息增益（熵-信息熵）

Gain(课后上机学习时间 （小时）)=Inf(D)-Info(课后上机学习时间 （小时）)

Gain(课前了解)=Inf(D)-Info(课前了解)

Gain(课堂学习)=Inf(D)-Info(课堂学习)

Gain(平时成绩)=Inf(D)-Info(平时成绩)

5,计算属性的分裂信息度量（越大，越不确定，越后选择）

H(课后上机学习时间 （小时）)=-5/12\*log2(5/12)-4/12\*log2(4/12)-3/12\*log2(3/12)

H(课前了解)=-5/12\*log2(5/12)-4/12\*log2(4/12)-3/12\*log2(3/12)

H(课堂学习)=

-4/12\*log2(4/12)-4/12\*log2(4/12)-3/12\*log2(3/12)-1/12\*log2(1/12)

H(平时成绩)=-3/12\*log2(3/12)-7/12\*log2(7/12)-2/12\*log2(2/12)

6,计算信息增益率

IGR(课后上机学习时间 （小时）)=Gain(课后上机学习时间 （小时）)/H(课后上机学习时间 （小时）)

IGR(课前了解)=Gain(课前了解)/H(课前了解)

IGR(课堂学习)=Gain(课堂学习)/H(课堂学习)

IGR(平时成绩)=Gain(平时成绩)/H(平时成绩)

1. 根据增益率排序选择属性

选择最大的增益率为节点，循环执行，直到所有节点到位，组成树。

3.6.2 匹配决策树

新创建的学生学习调查表，根据决策树匹配，找到当前教师课程下所有学生成绩的最大影响因素。动态匹配决策树，找到影响学生最大因素的树枝节点。

**第四章 项目实现**

**4.1 环境准备**

本环境皆为本地测试实验环境，非服务器生成环境。

4.1.1硬件环境

电脑主机一台，显示器一台，鼠标键盘等外设;

最低配置：CPU:Intel® Pentium 4 CPU:1.8GHz;

内存:1G;

硬盘:500G.

实际配置：

CPU:Intel® I5-6500 4CPU:3.2GHz;

内存:8G;

硬盘:500G.

4.1.2软件环境

操作系统：Windows7系统;

系统运行环境与工具：JDK8;

本地服务器运行环境：tomcat8;

数据库：SqlServer2008;

IDE:Eclipse;

前端框架：html5+jquery(javascript框架)+bootstrap(css框架);

系统框架：Struts2+Spring+Hibernate。

**4.2 项目展示**

4.2.1 登录页面

登录页面用于安全保护，用户输入用户帐号密码，点击登录后会自动判断帐号密码的正确性，正确会登录到系统首页。

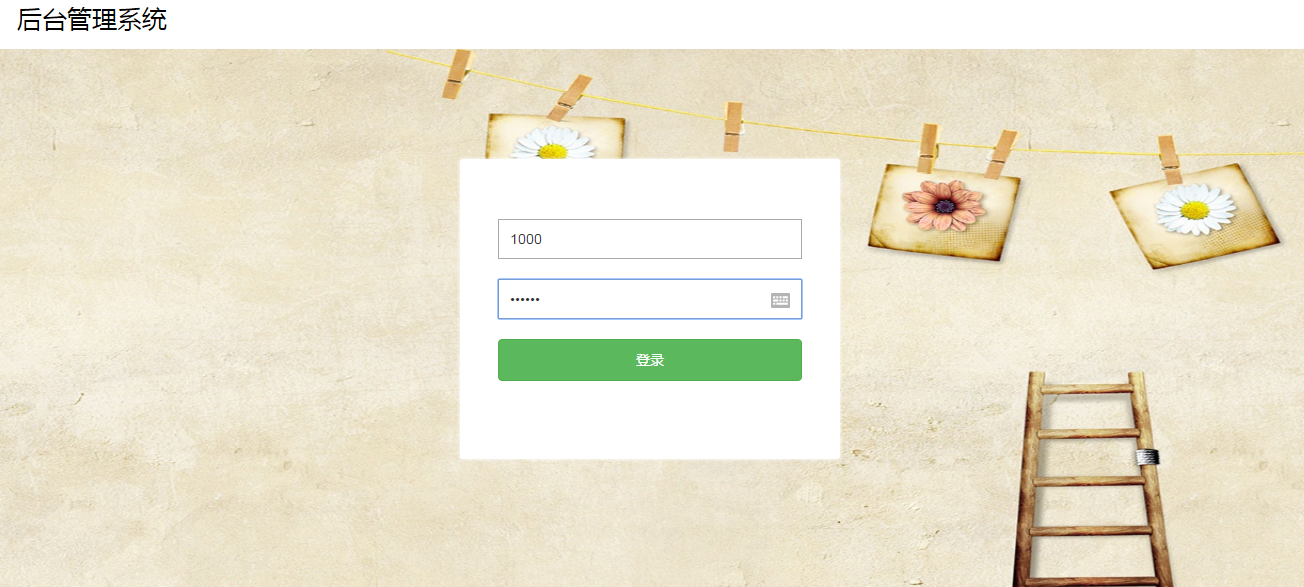


图4.2.1.1 登录页面

4.2.2 教师管理

教师管理主要展示的是所有教师的基本信息和每名老师所教导的课程，如图4.2.2.1所示，并可以查询到每个学生的学习调查表，清楚的知道每个学生的学习情况，如图4.2.2.2所示，也可点击进入该课程的学生成绩和本课程总成绩的因素分析，得出总的影响因素，也可以查询到每一名学生成绩的主要影响因素，如图4.2.2.2所示。



图4.2.2.1 教师管理主页面



图4.2.2.1 学习调查表页面



图4.2.2.3 老师课程成绩及分析

4.2.3 学生管理

学生管理主要显示的是某专业某班级的所有学生成绩，如图4.2.3.1所示，并可以点击查看该课程的课程分析，如图4.2.3.2所示，知道影响因素，从而了解到自己的不足，进而加强学习，拼搏出一个好成绩。



图4.2.3.1 学生管理主页面



图4.2.3.2 学生成绩分析

4.2.4 课程管理

课程管理主要展现的是有关课程的增删改差以及相应的课程的描述。



图4.2.4.1 课程管理页面

4.2.5 专业管理

专业管理主要介绍的是有关专业的增删改差以及相应的专业的基本描述，使学生进入对应专业，方便查询，对一些专业有相应的了解。



图4.2.5.1 专业管理页面

4.2.6 班级管理

班级管理主要展现的某专业下班级的增删改差，使得学生可以对号入座，方便管理查询，以及一些描述，对此班级有适当了解。



图4.2.6.1 班级管理页面

4.2.7 职称管理

职称管理主要介绍的是关于老师的职称，使得学生可以更好的选择适合自己的老师，也更好的了解老师的能力，信任自己的老师可以把自己教导的很好。



图4.2.7.1 职称管理页面

4.2.8 管理员信息

管理员信息主要是对于管理员信息和帐号密码的修改。

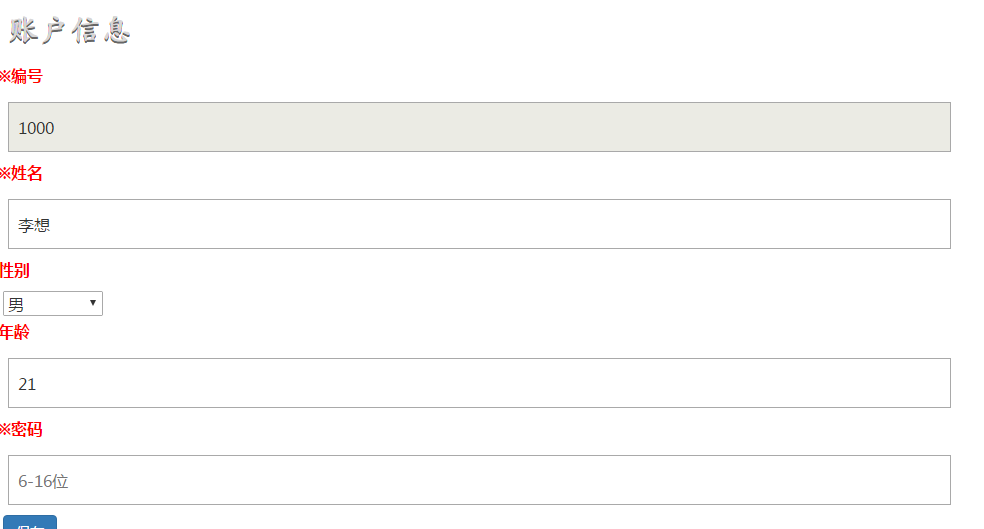


图4.2.8.1 管理员信息页面

1. **系统测试**

**5.1 系统测试**

软件测试是分析某个软件项以发现现存和要求的条件之差别并评价此软件项目的特性。本系统采用黑盒测试和白盒测试。

**5.2 系统登录测试**

此处主要针对系统登录的测试进行主要讨论方向，使用等价类划分法是测试登录的一个好方法。确定测试用例。

表5.2.1登录模块等价分类

|  |  |  |
| --- | --- | --- |
| 输入数据 | 有效等价类 | 无效等价类 |
| 帐号 | ①帐号为数字；  ②长度在9； | ③含非法字符格式错误  ④长度错误 |
| 密码 | ⑤密码为数字；  ⑥长度6-16； | ⑦含非法字符  ⑧长度错误 |

表5.2.2帐号测试用例

|  |  |
| --- | --- |
| 测试数据 | 测试范围 |
| 20185021 | ① ④ |
| 201850213 | ② |

表5.2.2密码测试用例

|  |  |
| --- | --- |
| 测试数据 | 测试范围 |
| 123456 | ⑤ ⑥ |
| 201 | ⑧ |

表5.2.2测试结果表

|  |  |  |
| --- | --- | --- |
| 测试数据 | 期望结果 | 实际结果 |
| 201850213 123456 | 输入有效 | 正确 |
| 20185021 123456 | 输入无效 | 错误 |
| 201850213 201 | 输入无效 | 错误 |
| 20185021 201 | 输入无效 | 错误 |

**5.3 系统功能测试**

5.3.1 测试过程

（1）启动项目，登陆超级管理员账户，增加专业、班级、职称、课程、教师、学生,退出系统。

（2）登陆教师账户，增加教师课程，增加当前教师课程下的学生，增加学生的学习调查表信息以及学生考试成绩，退出系统。

（3）登陆学生账户，查看自身成绩与与决策树自动生成的分析与建议。

5.3.1 测试结果

（1）系统主界面



系统登陆界面



系统主界面

（2）通过对学生调查表与考试成绩建立决策树，得到的影响学生成绩的因素的数据统计。









**5.4 测试结果分析**

在测试登录中，可以很清楚的看到，必须正确保证输入帐号密码的准确性才能进入系统，进而使用系统，此部分可确保系统的安全性。

在测试功能结果中**，**可以看出学生的成绩受课后上机时间、课堂时间、课前了解和平时成绩的影响，由于每个学生的学习调查表的不同，会对成绩产生相应的影响。由此，可以让学生可以更好的了解自身情况，可以改变相应的学习计划，提高学习的效率，从而努力获得一个良好的成绩；而老师可以看到整个课程受哪方面影响比较大，从而可以督促学生，加强自身对于该方面的注意，以此来提高整体成绩的水平。

**结 论**

本文通过构建企业级别的基本系统后台，让我完全了解熟悉了整个项目的构建过程，熟悉了企业应用系统项目的开发流程与作用。

通过对比两种决策树算法，分析他们的优缺点，并进行了综合分析，根据各种定位算法的优缺点和算法的某些特征，最终采用C4.5算法分析学生成绩的影响因素。

C4.5算法不是最好最完美的算法，但在多维度分析学生考试成绩的影响关系上面却是最直观最有效的算法。

从此项目立项到结束，困难而激情，让我学会了解决问题的方法，相信在今后踏入社会，走向人生道路上能并积薄发，越走越远。

在此，我特别感谢我的同学，我的导师，是你们的帮助与激励，才能让我完成此项目，增加自身的成就感，责任感。

鞠躬！谢谢！

2018-5-18

**参考文献**

[1]赵德才.决策树方法使用中的改进策略 [J]. 科技信息. 2010 (08)

[2] 董欢.决策树技术在高校学生成绩分析中的应用研究 [D]. 西安电子科技大学. 2012

[3] 康振华,周金刚.数据挖掘在高校就业工作建设中的应用 [J]. 现代计算机. 2006 (05)

[4] 倪现君.基于数据挖掘分类技术的高校教学方法研究 [J]. 科学技术与工程. 2006 (04)

[5] 张棪,曹健. 面向大数据分析的决策树算法[J]. 计算机科学,2016,43(S1):374-379

[6] 马瑜,王有刚.ID3算法应用研究 [J]. 信息技术. 2006 (12)

[7] 徐成香.基于数据挖掘技术的学生信息系统开发 [J]. 硅谷. 2009 (16)

[8] J. R. Quinlan.Induction of decision trees[J] 1986,Machine Learning(1):81～106

[9] 彭军勇.学生成绩分析系统设计与实现 [J]. 电脑编程技巧与维护. 2009 (11)

[10] 吴陈,林炎钟.C4.5算法在高校教师评价中的应用研究 [J]. 信息技术. 2011 (01)

[11] 徐鹏,林森.基于C4.5决策树的流量分类方法 [J]. 软件学报. 2009 (10)

**致 谢**

首先，我要感谢合肥学院给我提供的良好的学习环境，同时要感谢计算机科学与技术系所有老师对我的培养和关怀。在本次毕业设计中，论文指导老师殷荣网老师对我进行了悉心的指导，给我提供了许多帮助，给我指明了论文的研究方向。并且老师严谨的治学态度也深深的感动了我。

其次，我要感谢我的家人，是他们的无私支持使我能顺利完成大学的学习，是他们在我遇到难题时给我及时的帮助与无限的动力，所以，我以此毕业论文作为我大学几年来的学习成果向父母汇报，感谢你们。

在我几年的大学生活中，得到了老师们的悉心教导，学到了丰富的知识和做人的道理，为我走向社会、迎接挑战打下了坚实的基础。此时此刻我才真正体会到老师们的苦心和所学知识的重要。

最后，我要再次向给予我帮助的所有老师和同学表示深深的谢意，感谢大家在学习、生活中对我的帮助，谢谢大家！

**附录：**