



基于决策树算法的实践类课程分层教学研究

陈 颢, 林献坤

(上海理工大学 机械工程学院, 上海 200093)

摘要: 分层教学是满足不同学生群体的学习需求最有效的教学模式, 根据学生学习能力不同, 对学生进行精准分类是成功实施分层教学的关键。该文以液压与气动技术课程设计与实验课程为例, 分析了影响学生课程设计能力的主要因素, 根据支撑本课程的知识体系和培养目标, 确定了学生的分类属性, 采用决策树智能算法, 以往届学生的数据为样本, 建立了决策树分类模型, 并分析了模型的有效性。

关键词: 分层教学; 实践类课程; 决策树; 智能算法

中图分类号: G642.0

文献标志码: A

DOI: [10.12179/1672-4550.20230228](https://doi.org/10.12179/1672-4550.20230228)

Research on Stratified Teaching of Practical Courses Based on the Decision Tree Algorithm

CHEN Sa, LIN Xiankun

(School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: To meet the different learning needs of students, stratified teaching is the most effective teaching mode. Accurate classification for students is the key to the successful implementation of stratified teaching according to different learning abilities of students. Taking the course of Projects for Hydraulic and Pneumatic Technology as an example, this paper analyzes the main factors which affect students' curriculum design ability. The classification attributes of students are determined according to the knowledge system and training objectives that support this course. The decision tree intelligent algorithm is used to classify the students, and the classification model is established with the data of previous students. Finally, the effectiveness of the model is verified.

Key words: stratified teaching; practical course; decision tree; intelligent algorithm

随着高等教育从精英化向大众化转变, 高校招生规模日益变大, 学生个体差异也越来越大, 优化教学方法, 改变传统的“整齐划一”的教学模式, 因人而异地制定教学内容, 即实现分层教学, 是提高高校人才培养质量的重要途径^[1-3]。从多年前的卓越工程师教育培养计划到现在的创新实验班计划, 上海理工大学一直在探索如何在专业教学中有效地实现分层教学, 如何构建在统一的国家基本教育质量标准的分层教育理念和实际操作方法, 实现高校人才培养的多样性、多层次性。

对大学教学中进行分层教学已有很多探讨和研究, 文献 [4] 采用系统性文献分析法, 对近

20 年来中外文献中分层教学的教学成果进行了分析, 建立了分层教学效果分析的理论框架, 为改进分层教学设计提供了理论依据; 文献 [5] 基于新工科理念, 提出了利用线上资源和线下课堂混合的分层教学模式, 解决了学生知识储备参差不齐、课内学时有限与新工科成果导向多元化之间的矛盾; 文献 [6] 将小规模限制性在线课程 (small private online courses, SPOC) 与传统课堂相结合设计分层教学, 为学生提供了更灵活更符合学生需求的学习内容; 文献 [7] 从课程目标的确立、课程内容的组织、教学方式的设计及考核方式的制定 4 个方面阐述了“以学生为中心”的分层教学设计和实践; 文献 [8] 在大型开放式网络课程 (massive

收稿日期: 2023-04-21; 修回日期: 2023-10-09

基金项目: 上海理工大学 2020 年度教师教学发展研究项目 (CFTD201005)。

作者简介: 陈颢 (1976-), 女, 硕士, 讲师, 主要从事精密测量方面的研究。E-mail: chensaa@163.com

open online courses, MOOC)背景下,以大学计算机基础课程为例,采用螺旋分层式教学模式设计实验任务,提高了学生的学习兴趣。

本文以机械工程专业中的液压与气动技术课程设计与实验课程为例,具体研究实现分层教学中的学生分类问题。因为在具体实施分层教学过程中,对学生进行合理准确的分类是有效实现分层教学的前提和基础。

1 实践类课程进行分层教学的重要性

液压与气动技术课程设计与实验课程是机械设计制造及其自动化专业的一门专业必修实践课,包括实验和课程设计两部分,实验16学时,课程设计32学时。实验内容包含液压元件的拆装实验和气压回路控制实验。通过拆装实验,直观观察元件的内部结构,了解其工作原理和工艺要求;通过回路控制实验,深入巩固所学液压及气动专业知识。课程设计为一加工液压换向阀阀芯的专用机床设备设计,包括液压系统油路原理图及尾架、纵横刀架机构、后刀架机构及集成块的结构设计等,综合能力强的学生,可增设伺服系统设计内容。下面具体分析本课程作为一门实践课所具有的特点。

课程设计要求完成课程设计说明书一份,完成加工设备液压油路图的设计,绘制加工设备尾架、纵横刀架的装配图、液压站装配图及相关零件图。根据上述要求可以看出,学生既要掌握液压与气动技术的理论知识,又要具有机械结构设计能力、绘图能力以及力学分析计算能力,该要求已经超出了单一课程的知识空间范围,是对学生综合能力的锻炼和考察,体现了实践类课程综合性的特点。

课程设计中所有的设计任务均需学生自主完成,老师主要起到引导和有针对性的指导作用。学生在课程设计过程中需要自主规划设计方法,独立思考设计思路,自主完成设计内容,体现出实践类课程高度的自主性特点。

实践类课程的多样性主要表现在设计过程的多样性、评价标准的多样性。每个学生都有不同的思维方式和不同的设计过程,学生最终的设计方案是多种多样的。同时对应的评价标准也需要多样性,以此激励学生的创造力和想象力。

根据上述特点,对实践环节进行分层教学是

非常必要的,以此满足不同层次学生的学习需求。实现分层教学的关键点是如何精准地对学生进行分类,本文采用基于决策树的智能算法,通过收集学生相关属性的评价成绩,获取先验样本,通过机器学习建立预测模型,实现学生分类。

2 决策树分类算法描述

决策树算法是一类常见的机器学习方法,该算法通过对训练样本中的数据表现出的属性进行总结分类,根据数据的属性特征建立一个确切的描述与分类的模型,实现对不可预测的数据进行分类^[9-10]。决策树分类算法是以给定数据样本为基础的归纳学习方法,在给定已知属性的数据集的情况下,采用自上向下的递归方式产生一棵树结构。

决策树分类算法首先要选择具有最高信息增益的描述属性作为给定数据集的分支属性,创建决策树中的一个节点,根据该描述属性的不同取值再创建分支,之后对各分支中的样本子集递归调用上述方法,建立该节点的各个子节点^[11-12]。从根节点开始逐层向下判断,直到叶节点,得到该数据样本的分类树。

建立决策树模型的关键是如何选择最优划分属性,使自上而下的分支节点能最快速地达到统一,使所有节点具有相同属性。通常通过计算数据样本某一属性的信息熵和信息增益的数值,作为衡量样本纯度的指标,判断哪个样本属性作为当前的分支属性。

设所需分类的样本集合为 D ,样本中某一属性 i 的样本所占 D 集合样本数量的比例为 $p_i(i=1, 2, 3, \dots, k)$,集合 D 的信息熵为:

$$\text{Ent}(D) = - \sum_{i=1}^{|k|} p_i \log_2 p_i \quad (1)$$

$\text{Ent}(D)$ 信息熵的大小表征了样本的纯度,信息熵越小,则集合的纯度越高。

接着计算集合 D 中任一样本属性 n 的信息增益,设该属性具有 j 个可能取值情况,记作 $n^1, n^2, \dots, n^j$,如果采用属性 n 对集合 D 进行分类,则可把集合 D 分为 j 个分支,产生 j 个分支节点,每个分支节点包含的属性为 n 的样本记作 D^n ,根据式(1)可计算出 D^n 的信息熵。由于不同的分支节点包含的样本数目不同,对分支节点赋予权重为

D^n/D ，以此计算出用属性 n 对集合 D 进行分类时，所对应的信息增益值为：

$$\text{Gain}(D, n) = \text{Ent}(D) - \sum_{n=1}^j \frac{|D^n|}{D} \text{Ent}(D^n) \quad (2)$$

信息增益越大，代表使用属性 n 来对集合 D 进行分类所获得的“纯度提升”越大。选择具有最高信息增益的描述属性作为给定数据集 D 的分支属性，之后对各分枝中的样本子集递归调用上述方法，建立当前节点的各个子节点。当某个分枝上的所有数据样本都属于同一个类别时分类停止。

3 基于决策树的学生分类方法研究及结果分析

3.1 学生分类属性的确定

由于液压与气动技术课程设计与实验课程涉及对学生多方面能力的考察，对学生进行分类时，需要考虑的因素也是多方面的，即分类属性的定义和划分就需要多方面考虑，避免出现分类误差，否则既影响后续学生进行课程设计的质量，也失去了对课程进行分层教学的意义。对本课程来说，学生的分类属性要根据支撑本课程的知识体系和本课程的培养目标进行确定。

本课程的培养目标之一是对已学过的液压与气动技术的基本理论知识进一步巩固和深化，掌握液压和气压基本元件的结构和工作原理，掌握常用液压系统的设计步骤和元件的参数计算。基本理论知识的掌握程度是进行本课程毕业设计的必要条件，所以将其作为对学生进行分类的属性之一。依据学生的液压与气动技术课程成绩，把基本理论知识的掌握情况分为 4 个等级：优、良、中、差。

本课程设计需要绘制液压站的装配图和零件图，或者绘制非标集成块零件图及装配外形图，所以绘图能力也作为学生分类的属性之一，分为 4 个等级：优、良、中、差，等级的划分可以参考学生的前修课程——机械制图的成绩。

本课程需要学生具有一定的动手能力，拆装实验和回路控制试验都是实际操作的实验内容，通过动手操作，学生能更加深刻体会流体作为传递介质的优点，更好地掌握液压传动系统设计和计算的一般方法。所以学生实验动手能力也作为

学生分类的属性之一，其数据来源可根据学生的相关实验成绩得到。

本课程的另一培养目标是通过课程设计与实验的综合训练，使学生能够综合运用所学过的液压与气动技术的相关理论知识，初步具有针对液压与气动设备进行设计、制造、工艺规划的能力。培养学生液压系统总体设计能力和液压元件的结构设计能力，需要学生具有一定的自主设计能力，因此将创新和设计能力也作为对学生分类的属性之一，该属性可从学生是否参加过相关专业的竞赛或者是否参加了创新实验课题获得，分为有、无两类。

本课程设计需要学生掌握常用液压元件，比如液压泵、液压缸、各种液压阀的选型，学会查找和使用各种设计手册及产品样本等相关技术资料，所以查阅文献能力也要作为学生分类属性之一，该属性可参考学生的前修课程——文献阅读的成绩，分为 3 个等级：强、中、弱。

故对学生分类属性可划分为：基本理论知识、绘图能力、创造和设计能力、实际操作能力、查阅文献能力。根据样本中学生本课程的最终成绩，将学生分为高级别和低级别两类。样本数据如表 1 所示。

表 1 根据学生分类属性划分的数据样本

学号	基本理论知识	绘图能力	创造和设计能力	实际操作能力	查阅文献能力	分类结果
1	优	良	否	强	中	高级
2	良	中	否	中	中	低级
3	中	优	否	中	中	低级
4	中	良	否	中	中	低级
5	良	优	否	中	强	高级
6	差	中	否	中	弱	低级
7	中	良	否	强	强	高级
8	良	良	否	强	中	高级
9	优	优	是	强	强	高级
10	优	中	否	中	中	低级
11	中	中	否	中	中	低级
12	良	中	否	中	强	高级
13	良	良	否	强	强	低级
14	中	差	否	弱	中	低级
15	优	优	是	强	中	高级
16	中	中	否	中	中	低级
17	中	良	否	中	中	低级
18	良	优	否	中	强	高级
19	差	差	否	弱	弱	低级
20	中	中	否	中	强	低级

通过对上述样本进行机器学习,可得到一棵分类树模型,应用该模型可实现对后续学生的准确分类,具体算法实现如下所述。

3.2 学生分类的决策树节点信息增益的计算

根据表1样本创建决策树,其根节点包含了样本中的所有属性:基本理论知识、绘图能力、创造和设计能力、查阅文献能力。样例中的分类结果显示:综合能力高的学生占了总学生人数占8/20,综合能力低的学生占了总学生人数的12/20,从而可以求出根节点的信息熵为:

$$\begin{aligned} \text{Ent}(D) &= -\sum_{i=1}^{|D|} p_i \log_2 p_i = \\ &= -\left(\frac{8}{20} \log_2 \frac{8}{20} + \frac{12}{20} \log_2 \frac{12}{20}\right) = 0.971 \end{aligned}$$

计算出属性集合{基本理论知识;绘图能力;创新和设计能力;实际操作能力;查阅文献能力}中每个属性的信息增益,本文以“基本理论知识”为例,分析其信息增益的计算过程。属性“基本理论知识”有4个可能的取值:{优,良,中,差},使用该属性对样本集合 D 进行划分时,获得4个子集,分别为 D^1 (基本理论知识=优), D^2 (基本理论知识=良), D^3 (基本理论知识=中), D^4 (基本理论知识=差)。

子集 D^1 中包含的学生编号为{1、9、10、15},其中正例(综合能力高的学生) $p_1=3/4$,反例(综合能力低的学生) $p_2=1/4$;子集 D^2 包含学生编号为{2、5、8、12、13、18},其中综合能力高的学生所占比例 $p_1=4/6$,综合能力低的学生所占比例 $p_2=2/6$;子集 D^3 包含学生编号{3、4、7、11、14、16、17、20},其中综合能力高的学生所占比例 $p_1=1/8$,综合能力低的学生所占比例 $p_2=7/8$,子集 D^4 中包含的学生编号为{6、19},其中综合能力高的学生所占比例 $p_1=0$,综合能力低的学生所占比例 $p_2=1$ 。根据式(1)计算出采用属性“基本理论知识”划分之后4个分支节点的信息熵分别为:

$$\text{Ent}(D^1) = -\left(\frac{3}{4} \log_2 \frac{3}{4} + \frac{1}{4} \log_2 \frac{1}{4}\right) = 0.811$$

$$\text{Ent}(D^2) = -\left(\frac{4}{6} \log_2 \frac{4}{6} + \frac{2}{6} \log_2 \frac{2}{6}\right) = 0.918$$

$$\text{Ent}(D^3) = -\left(\frac{1}{8} \log_2 \frac{1}{8} + \frac{7}{8} \log_2 \frac{7}{8}\right) = 0.544$$

$$\text{Ent}(D^4) = -\left(\frac{0}{2} \times \log_2 \frac{0}{2} + \frac{2}{2} \times \log_2 \frac{2}{2}\right) = 0$$

根据式(2)计算出属性“基本理论知识”的信息增益为:

$$\begin{aligned} \text{Gain}(D, \text{基本理论知识}) &= \text{Ent}(D) - \sum_{n=1}^4 \frac{|D^n|}{D} \text{Ent}(D^n) = \\ &= 0.971 - \left(\frac{4}{20} \times 0.811 + \frac{6}{20} \times 0.918 + \frac{8}{20} \times 0.544 + \frac{2}{20} \times 0\right) = \\ &= 0.3158 \end{aligned}$$

同理,可以推算出其他各项属性的信息增益分别为:

$$\text{Gain}(D, \text{绘图能力}) = 0.423$$

$$\text{Gain}(D, \text{创新和设计能力}) = 0.145$$

$$\text{Gain}(D, \text{实际操作能力}) = 0.289$$

$$\text{Gain}(D, \text{查阅文献能力}) = 0.204$$

通过对信息增益大小比较,可得样本属性“绘图能力”的信息增益最大,所以“绘图能力”被首选为划分属性,对学生样本进行第一次划分即对根节点进行划分,如图1所示。

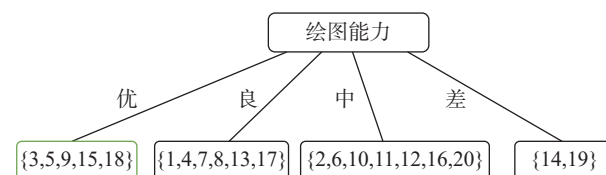


图1 以信息增益最大的属性对根节点划分

3.3 决策树学习模型的建立

递归调用上述方法建立该节点的各个子节点,对每个分支节点进行进一步划分。继续把图1中的每个分支节点进行如上算法的计算,得到每个分支节点集合 D^n 在可用属性下的信息增益。以第一个分支节点(“绘图能力=优”)为例,计算该节点集合 D^1 {3、5、9、15、18}的5个样例,可用属性集合为{基本理论知识;创新和设计能力;实际操作能力;查阅文献能力},在“绘图能力=优”下可以计算出上述属性的信息增益为:

$$\text{Gain}(D^1, \text{基本理论知识}) = 0.722$$

$$\text{Gain}(D^1, \text{创新和设计能力}) = 0.171$$

$$\text{Gain}(D^1, \text{实际操作能力}) = 0.171$$

$$\text{Gain}(D^1, \text{查阅文献能力}) = 0.322$$

进一步增加学生样本数量，采用上述算法，可获得完整的决策树学习模型如图 2 所示。

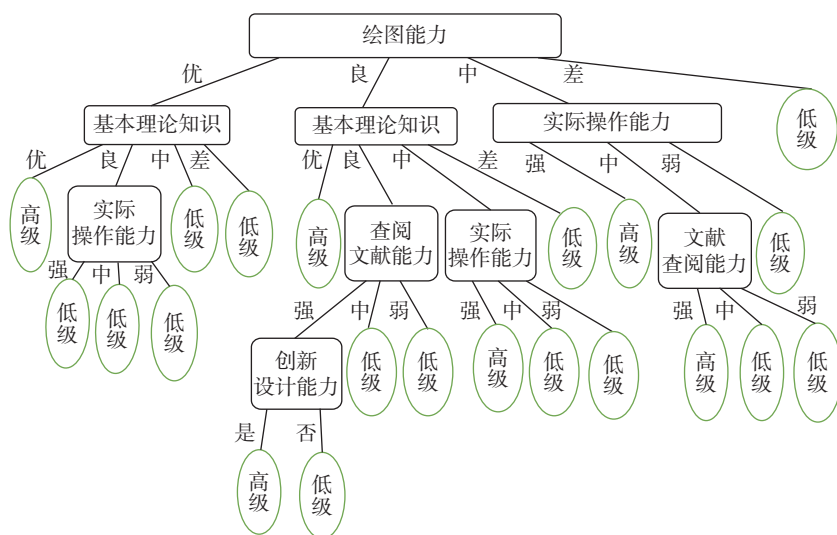


图2 决策树分类模型

决策树分类模型能够直观地展示出学生的各项属性对分类的影响,由图2可以看出,对分类影响比较大的是“基本理论知识”和“绘图能力”这两项。本实践课所涉及的基本理论知识包括液(气)压系统的基本组成、液(气)压元件的工作原理、基本液(气)压回路等。从图中可见,即便绘图能力为良好,基本理论知识的成绩优秀,也可直接被划分为高级别。而绘图能力作为工科学生尤其是机械专业学生所必须学习的一门基础

4 结束语

本文所述分类算法的准确性很大程度上依赖于学习样本数据, 样本数据量越大、类型越丰富, 所得到的分类模型对未来数据的预测越准确, 因而后续可以通过增加学生样本数量, 根据课程性质和培养目标合理设置数据分类属性, 增加分类数据类型, 以提高模型的正确率。

[1] 周光明,彭红军. 尊重差异因材施教构建“和谐”教学新模式: 高校分层教学探索[J]. 中国大学教育, 2007(6): 69-71.

(下转第 139 页)