# 基于实例的高校《数据挖掘》课程教学模式探索

孟丽丽　于洋

天津商业大学理学院

**摘　要：**《数据挖掘》是计算机与数学知识相交叉的应用课程。文章针对《数据挖掘》课程教学模式中存在的“三重三轻”现象，结合近几年教学的实际授课经验，通过构建“一中心+五维一体”的课堂教学模式，形成以实例为驱动力的运行模式，为《数据挖掘》课程教学改革提供思路。

**关键词：**数据挖掘；实例；课程思政；教学探索

【DOI】10.12252/j.issn.2096-627X.2024.05.129

## 引言

随着经济、科技和信息技术的飞速发展，特别是网络技术的发展，数据的产生与应用有了很大的提高。

很多高校开设了《数据挖掘》课程，该门课程指运用有效的方法和工具收集、处理数据并获取信息的过程，

目的是从杂乱无章的商务数据中提炼出有用数据，用于研究数据的内在规律和特点，从而指导及优化企业运营在工程技术、人工智能、经济管理等领域有广泛应用。

本文以《数据挖掘》课程为例，针对该课程教学模式中存在的“三重三轻”现象，结合近几年教学的实际授课经验，通过构建“一中心+五维一体”的课堂教学模式，形成以实例为驱动的运行模式，为实现大数据时代下《数据挖掘》课程教学改革主旨指明方向与可行路径。

## 一、以“学生为本”，构建教学体系

《数据挖掘》作为数据科学与大数据专业的核心专业课，是计算机与数学知识相交叉的应用课程，与数据库、人工智能、统计学等课程息息相关。数据挖掘原理与算法涉及的学科领域很宽泛，其最终目的是在数据中挖掘出可供人们利用的知识和信息，因此数据挖掘技术涉及数据库技术、统计学、机器学习、神经网络、知识系统、信息检索、高性能计算和可视化等领域知识。整本教材涵盖了数据分析基础和实现人工智能的关键算法，所以课程难度较大。

### （一）课程教学模式现状

传统数据挖掘课程的教学模式基本以讲授教学为主，在教学中存在的问题主要有以下三个方面：

（1）重“知识”轻“思想”。通常大学教师在教学过程中单纯重视知识的传授，使得数学知识变得单调乏味，不利于大学生对知识的掌握。

（2）重“经典”轻“时效”。随着社会发展和社会进步，经典教材中许多例题比较古典陈旧，不能与时俱进，特别存在部分内容过分追求逻辑的严密性和理论体系的完整性，内容理论化，不能调动大学生的学习兴趣和对知识的掌握。

（3）重“传统”轻“创新”。大学教师在教学时往往采用板书和 PPT演示相结合的讲授方法，教学方法较单一，学生比较被动的接受知识。

### （二）构建“一中心+五维一体”教学模式

本着育人之本在于立德铸魂的精神，《数据挖掘》的教学目标主要是：

（1）知识目标：掌握数据挖掘一般理论和技术，具有进一步学习的能力，为后续课程打下基础。

（2）能力目标：能够在实际问题中熟练应用数据挖掘技术，解决遇到的各种问题。

实现《数据挖掘》课程的教学目标，需要大学教师在日常的教学过程中，要始终坚持“一中心+五维一体”的教学模式（见图1）。一中心，即以学生为中心，主要体现在“五个维度”，分别是思想品行，包括培养学生的政治倾向，思想信仰，道德品格，科学精神和人文素养；二是专业水平，主要包括数学水平、数据处理和数据分析水平、科学计算水平。三是学习能力，包括信息搜索、自我学习和逻辑推理能力。四是思维能力，包括创造性思维和系统性思维能力。五是职业素养，包括职业道德、职业技能和沟通能力。同时，教师在教学组织和教学手段方面要推进网络化课程平台建设，促进学生个性化学习，尝试开展项目式学习，强调教师角色转变，由知识的传授者转变为项目的合作者；提高学生课堂参与度。

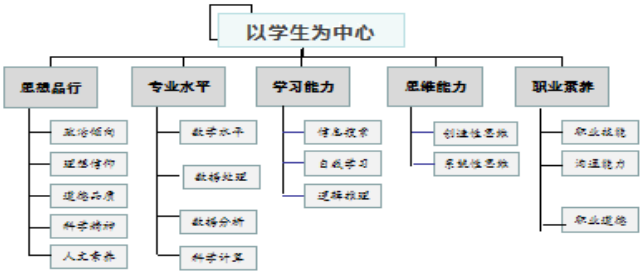


图1 一中心+五维度教学模式

为实现上面的建设过程，需要进一步明确数据挖掘课程教学基本要求，修订教学大纲；重新编写课程教案。主要从以下方面实施。

（1）拓展教学内容

从知识的起源、人物的成长、学科的发展，挖掘其中的课程案例；从当前的社会热点问题出发引发学生思考；通过知识的重组，将课程内容由点到面、由面到体进行扩展，实现教学内容在深度和广度上的延伸，不断拓展教学内容。

（2）采用情景教学法

针对课程中各知识内容创设相应的问题情境，教师由传统的讲授型角色转变为引导式角色，将课程知识融入情景之中，学生能自发、自觉的产生对情景的呼应。

（3）注重翻转教学法

《数据挖掘》这门课知识点比较明确，很多教学内容只需讲清楚一个概念、一道公式、一道例题、一个算法，其学科特点便于翻转课堂的实施。利用翻转课堂使得学生在完成案例的基础上，修改、优化、精心打造自己的案例，临场答辩，并完成“分析问题---算法设计---解决问题---实践应用”整个过程的学习。

（4）改进课程的考核和评价方法

在多元化教学中，教学采用的教学方法多，形式多样化，因此，课程的考核与评价方法与传统教学有很大的差别。课程教学中，要建立将传统考核方法与过程性评价、刑成性评价等多种形式相结合的考核与评价体系，从而体现多远化在课程教学中的基础性地位。该课程在考核中采用期末考核成绩和平时考核成绩各占总评成绩的70%和30%。期末考核采用闭卷考试，平时成绩包括以下几部分：

1.爱课程（中国大学MOOC）平台上的学习完成度和课前案例的预习度。

2.线下上机操作考试放在课程的中后段进行，要求学生现场写程序并上机运行得到结果。

3.翻转课堂答辩小组成绩，在每个案例后进行。

4.线下考勤及课程参与度，包括雨课堂参与度、雨课堂小测验成绩等。

**二、案例贯穿教学全程，做“实”案例教学运行模式**

在数据挖掘课程教学模式中，探索采用案例教学法并且重新组织授课内容，把主讲教材内容分成四个模块：数据分析基础模块、分类算法模块、回归模型模块和集成模块。对每一个模块的方法理论介绍时，首先要介绍的不是理论本身，而是这个理论适用的特定场景，这个场景具体到一个非常鲜活的案例，并且这个案例要足够大。

教学运行模式具体分为三阶段进行。首先，基于大数据实训平台的课前预习布置，通过案例背景与问题引导使学生提前了解课程内容，促使学生自主思考，推进在课堂教学中的问题讨论与合作学习。其次，课堂中以案例贯穿教学全过程，通过完成一个实际问题研究的全流程进行数据挖掘方法的知识点介绍，在讲案例故事的同时把理论方法穿插其中，教会学生学以致用，通过理论讲授和操作实践的结合使得学生牢固掌握知识。最后，紧接翻转课堂，鼓励学生自主选题进行数据挖掘方法的实践研究，加强学生对案例的分析和理解，促使学生更好地理解和应用所学知识，举一反三解决感兴趣的实际问题。下面以具体案例展示教学运行模式全程。

（一）第一阶段：课前实例预习

在大数据实训平台给出案例问题与目标：本实例的研究目标是对《红楼梦》前80回和后40回的作者是否一致进行判断，利用聚类分析的角度尝试解决这个问题。同时给出问题思考与引导：聚类分析希望达成的实际目标是什么？聚类分析方法需要哪些数据信息？《红楼梦》文本数据应如何应用于数据挖掘方法中？

（二）第二阶段：课堂实例讲授

以《红楼梦》文本数据的预处理为起点，至《红楼梦》回目的聚类分析与结果对比，贯穿数据挖掘应用步骤全过程，系统介绍数据挖掘案例教学的完整思路与知识内容。

（1）文本数据的预处理

①介绍python中相关数据库的准备与《红楼梦》文本数据的读取。②讲授文本数据的正则表达式匹配技术。引出实现数据信息字符串的快速查找、提取、删除与替换的知识点，并在本例中应用于《红楼梦》回目提取与章回文本切割。③重点讲授文本数据的分词及去噪声技术。在本例中对《红楼梦》章回文本进行分词，去掉停用词、单字词等。④讲授文本数据的向量化方法。在本例中将《红楼梦》章回文本内容处理为分词向量矩阵的形式，初步形成用于聚类分析的文本数据特征。⑤重点讲授TF-IDF文本特征提取技术。在本例中将《红楼梦》章回文本内容进一步处理为TF-IDF矩阵的形式，完成用于聚类分析的文本数据特征。

（2）文本数据的可视化

在介绍按词频排序的条形图与词云图的文本数据可

视化方法之后，直观展示《红楼梦》章回文本所涵盖的主要信息。

（3）聚类分析与结果对比

在讲授聚类分析的基本知识和常用算法（如KMeans聚类、Ward聚类、DBSCAN、MeanShift等）之后，在本例中对《红楼梦》章回文本内容进行实际应用，观察KMeans聚类和Ward聚类方法的聚类效果，验证最初问题提出时《红楼梦》章回作者的假设，给出最终的聚类分析结论。从聚类过程来看，kmeans方法相比ward方法计算耗时更少。从聚类结果来看，kmeans方法需事先设定类别总数与初始质心，这些参数的设定将直接影响聚类效果，而ward聚类不需设定初始参数，且可直观看到不同层次类别总数下的章回聚类情况。由此引导学生总结不同聚类方法的优缺点与适用环境，思考如何因地制宜地选择聚类方法进行实际应用。在引导学生思考不同聚类方法的思想差异与应用表现的过程中，引导学生深刻理解世界上的事物不是孤立的存在，分析问题一定要多方面综合考虑并重视其各种构成因素之间的关系，避免顾此失彼，培养学生的大局意识。

（三）第三阶段：翻转课堂实践

在学生完成本例练习后，引导学生自选实际问题进行上述聚类分析方法的实践应用，鼓励学生组成小组，进行答辩，给每小组得分，算作平时成绩的一部分。

## 三、教学模式效果与反思

（1）通过案例教学法使学生从失败的研究中总结经验教训和警示，从当前的社会实际问题出发引发学生思考。同时，将书本知识贯穿于实践应用全过程，通过知识的重组，将知识由点到面、由面到体进行系统扩展，实现教学内容在深度和广度上的延伸，拓展教学内容。

（2）通过翻转课堂激发学生自主应用知识的兴趣，提升学生学习的主观能动性，学生在翻转课堂环境中比传统课堂环境中的学习效果更佳。翻转课堂开阔了学生的眼界，启迪了学生的思维。在教育信息化发展过程中，翻转课堂教学模式是新的机遇和挑战。

（3）学生实践应用过程中存在对聚类知识掌握不够熟练，对文本数据挖掘步骤理解不够充分的问题。对此，应事先组织学生对重点、难点知识进行有针对性地预习，初步构建本节课内容的基本知识框架，调动学生的学习积极性，引导学生自主思考，进一步带着问题上课学习。

## 结语

本文对《数据挖掘》课程教学模式中存在的“三重三轻”现象，通过构建“一中心+五维一体”的课程教学模式，以《红楼梦》文本数据挖掘问题作为授课内容主线，让实例教学始终围绕着典型的算法和前沿知识展开的教学方式，注重翻转课堂，在今后的教学工作中，我们还应该不断站在学科发展的前列，从知识传授为主向能力培养为主转变，从课堂学习为主向多种学习方式转变。经常更新实例，使其更好地融入教学。

## 参考文献

[1]万通明，陈小秋，马苗.基于《机器学习》中决策树学习的思政教育[J].产业与机器论坛，2021，17（20）：129-130.

[2]赵青明.教育部印发《高等学校课程思政建设指导纲要》，全面推进高校课程思政建设[J].新教育，2020，19（7）：32.

[3]周志华.机器学习[M].北京：清华大学出版社，2016.

[4]刘云霞.统计学专业本科生开设“数据挖掘”课程的探讨[J].吉林工程技术师范学院学报，2010（6）：20-22.

[5]中南大学.本科生院召开“开放式精品示范课堂计划”课程建设研讨会[EB/OL].（2016-3-25）[2020-8-1]. http: //news.csu.edu.cn/info/1004/126648.htm.

[6]吴岩. 应对危机 化危为机 主动求变 做好在线教学国际平台及课程资源建设[J]. 中国大学教学，2020（4）：4-16.

[7]谭征，孙红霞，王立宏.普通院校本科生开设数据挖掘课程的教学研究[J].福建电脑，2011，27（10）：38-39.

作者简介：孟丽丽（1979—），河南漯河人，天津商业大学理学院副教授，硕士，研究方向为应用统计。

于洋（1992—），天津市人，天津商业大学理学院讲师，博士，研究方向为数据挖掘。

基金项目：本文为天津市教育科学规划项目，“NEET”理念下数据科学与大数据技术专业人才培养范式与实现路径研究”（CIE220074），教育部东方艾学协同育人项目“基于NEET理念的大数据实践教学平台建设”（2022年），天津商业大学课程思政示范课“数据挖掘”建设项目（2022年）的成果。