

“以赛促学，以赛促教”在教学中的应用

——以数据挖掘概念与技术课程为例

王 苗 董利飞

重庆三峡学院 重庆 404130

摘 要: 数据挖掘概念与技术是信息与计算科学的基础专业课之一,在数据呈指数化增长的今天,亟需高质量的数据挖掘技术人才。而目前数据挖掘概念与技术的教学主要采用理论教学结合实践教学的模式,理论教学主要以算法理论知识为主,实践教学主要以程序设计语言实现算法为主,但均存在一些问题。本文主要分析了数据挖掘的教学现状,指出了教学过程中存在的不足,并提出了“以赛促学,以赛促教”的提升数据挖掘概念与技术教学质量举措。

关键词: 数据挖掘概念与技术;以赛促学;以赛促教

The application of “promoting learning and promoting teaching by competition” in teaching

— takes the data mining concept and technology course as an example

Miao Wang, Lifei Dong

Chongqing Three Gorges University, Chongqing, 404130

Abstract: Data mining concept and technology is one of the basic professional courses of information and computing science. In today's exponential growth of data, high-quality data mining technicians are urgently needed. At present, the teaching of data mining concept and technology mainly adopts the mode of combining theory teaching with practice teaching. Theory teaching mainly focuses on algorithm and theoretical knowledge, and practice teaching mainly focuses on programming language to realize algorithm, but there are some problems. This paper mainly analyzes the teaching current situation of data mining, points out the shortcomings in the teaching process, and puts forward the improvement of data mining concept and technical teaching quality measures of “promoting learning and promoting teaching by competition”.

Keywords: data mining concept and technology; to promote learning; to promote teaching by competition

1 引言

数据挖掘概念与技术是信息与计算科学专业基础平台课程之一,是一门实践性课程,交叉关联多个学科,是人工智能和数据库领域研究的热点问题。数据挖掘是

项目基金: 重庆市教育科学“十四五”规划2021年度重点项目“地方性本科院校新工科课程思政建设模式探索与实践”(2021-GX-131);重庆三峡学院教学改革项目“新工科背景下地方应用型本科院校安全工程专业课程建设模式探索与实践”(JGQN2101)。

作者简介: 王苗(1990-),女,汉族,湖北仙桃人,助教,硕士,重庆三峡学院,研究方向:数学与应用数学、大数据科学与技术等方面教学和科研工作。

指从数据库的大量数据中揭示出隐含的、先前未知的并有潜在价值的信息的非平凡过程,是一种决策支持过程,它主要基于人工智能、机器学习、模式识别、统计学、数据库、可视化技术等,高度自动化地分析数据,作出归纳性的推理,从中挖掘出潜在的模式,帮助决策者调整市场策略,减少风险,作出正确的决策。本课程主要讲授数据挖掘的基本概念,原理、方法和技术,具体包括:数据的预处理、分类预测、关联挖掘、聚类分析等内容。通过学习,使学生理解数据挖掘的基本流程,掌握数据挖掘的基本理论和技术,熟悉数据挖掘成果的表述;掌握数据挖掘的基本方法,能熟练地应用数据挖掘技术对现实数据进行有效的分析,能够结合python软件从大量统计数据中获取有价值的信息。

数据挖掘概念与技术先修课程为《概率论与数理统计》、《线性代数》、《数据库原理与应用》、《程序设计基础》等专业课程,因而学生已经具备了一定的数学基础知识、基础专业技能和程序设计技能。该课程的教学模式为理论课教学结合实践课教学,总学时为64学时,包含40学时理论课程及24学时实践课程,教师的授课方式主要还是以传统的讲授法为主,理论课的在讲授的过程中主要还是以算法为主,案例讲解为辅,实验课则是以学生通过python编码实现相应算法为主,在实际场景中的应用为辅,但均存在案例陈旧,落后于时代等相关问题。同时,学生被动的接受相关知识,参与度低,课堂上与课堂下学习脱节,对知识的理解不够深入,仅仅停留在表面,自主学习能力不够,对学习的兴趣不够。

“以赛促学,以赛促教”顾名思义就是以比赛的形式促进学习,以比赛的形式提高教师的教学能力,这是一个很好的外在反馈系统,可以让学生与教师形成良好的学习闭环。比赛是一种学习形式,是一种外在的评判标准,能较好的检验学生的学习成果,使得学生能够充分准备,克服拖延的坏习惯,提高学习效率,提高学生的自主学习能力,同时,在比赛过程中形成“比学赶超”的氛围,也能够满足学生自我实现的需求。比赛是一种反馈形式,通过学生在参与比赛的过程中出现的问题,帮助教师充分认识教学过程中的不足,帮助教师进一步深化课程改革、优化课堂教学,切实提高教师的教学质量与教学水平,为进一步提高教师专业素养,以教师素养提升带动学生素养的落地,从而构建具有明显特色的高质量课堂教学模式。

2 数据挖掘课程教学的现状

数据挖掘是一个反复迭代的人机交互处理过程。该过程需要经历多个步骤,并且很多决策需要由用户提供。从宏观上看,数据挖掘过程主要由三个部分组成,即数据整理、数据挖掘和结果的解释评估。

数据挖掘概念与技术基本理论主要由六个模块组成,包含数据预处理方法、关联规则挖掘理论和算法、分类方法、聚类方法、时间序列和序列模式挖掘等,有Apriori关联规则、Knn、C4.5、Naive Bayes、最大期望算法(EM)、K-means等经典算法^[1]。而实践课主要由三个部分组成,也即数据挖掘过程的三个组成部分,即依托python进行数据处理、数据挖掘、结果的可视化。

数据挖掘概念与技术主要内容由各种不同类型的抽象算法所构成,学生要能够较好的理解相关算法,则需对数学能力有一定的要求;教师要能够深入浅出的进行授课,除了需要对算法有深刻的认识和理解外,还需要结合学科发展及时代要求,明确数据挖掘在实际中的应用。数据挖掘概念与技术实践部分主要为数据挖掘的三个宏观步骤,在数据挖掘的过程中,需要学生不断的尝

试多种算法,对结果进行比对分析,是一个不断试错的过程,需要学生较好的掌握python基础,有较强的编程能力,能够了解不同算法适用的场景;教师要讲授好实验课,对编程能力提出了更高的要求,也需要教师有紧跟时代发展的综合性案例。目前,已有较多学者研究了数据挖掘概念与技术在教学过程中存在的问题^[2-5],本文主要从以下几个方面进行分析。

2.1 课程难度大,学生基础存在差异

数据挖掘就是从大量的、不完全的、有噪声的、模糊的、随机的实际应用数据中,提取隐含在其中的人们事先不知道的、但又是潜在有用的信息和知识的过程,是一类深层次的数据分析方法,对学生的数学能力要求较高,而学生的数学基础参差不齐,对算法的理解只能停留在表面,对各个算法在实际中的应用无从下手,更谈不上优化跟改进,学习起来比较吃力。在实践教学方面,对学生要求较高,除了要求学生掌握数据挖掘基本算法理论外,还要求学生有一定的编程能力,但部分学生对程序设计语言的语法知识掌握不够,程序设计能力不足,难以自主的实现算法的编码,更有甚者,对给出的代码都无法读懂,产生畏难情绪,缺乏学习的积极性。

2.2 教学案例陈旧

数据挖掘概念与技术是综合应用性较强的一门课程,由于目前数据呈爆炸式增长,亟需数据挖掘技术从大量数据中挖掘出有用的信息,这使得数据挖掘成为热门,可供选择的教材很多,网络资源亦颇丰。数据挖掘概念与技术理论部分主要以算法理论为主,案例讲解为辅,实践教学主要以算法实现为主,各算法在实际中的应用案例为辅,合适的案例能够帮助学生更好的理解算法,提高学生的学习兴趣,但目前教师授课所使用的案例大部分来自于教材及参考书,实验部分也多以验证性实验为主,难以获取当前最新的应用案例,增加了案例教学备考的难度,导致案例陈旧,难以激发学生的学习兴趣。

2.3 实践教学不足

数据挖掘概念与技术是一门实践性较强的学科,内容多,涵盖广,算法知识也较为抽象复杂。理论教学主要以教师讲授为主,学生被动接受相关知识;在实践教学方面,主要以验证性实验为主。要掌握数据挖掘概念与技术算法的核心思想,仅仅依靠教师课堂上的理论讲解及简单的验证性实验,远远不够。但目前数据挖掘概念与技术的授课模式理论教学与实践教学结合不够,光说不练假把式,且不同知识点之间的验证性实验相互独立,难以帮助学生更加系统的理解各算法的理论知识,更不用说将所有算法知识串联起来。同时,在实践教学过程中,所用到的数据体量往往较小,且大部分都是事先预处理过的,不利于学生掌握数据预处理方法,不利于学生理解数据挖掘算法,更不用说将数据挖掘算法应

用到实战场景中去, 学生也难以理解数据挖掘的本质。在实践教学过程中, 也难以真正做到以学生为中心, 因材施教。

3 提升数据挖掘概念与技术教学质量举措

针对数据挖掘概念与技术教与学的过程中出现的问题, 提出了“以赛促学, 以学促教”的教学模式, 用以提高教师的教学质量与教学水平, 增加学习在理论层面对知识的理解, 在实践层面的程序设计能力, 提高学生学习兴趣, 加强学生自主学习的能力, 培养适应现代技术发展的数据挖掘技术人才。

3.1 以赛促学, 赛学融合

以赛促学, 通过参加竞赛, 激励学生学习数据挖掘的积极性, 提高学生分析、解决实际问题的综合能力, 同时也能在一定程度上提高学生的数据分析应用创新能力。

目前数据挖掘竞赛有许多, 有“泰迪杯”全国大学生数据挖掘竞赛, 数据竞赛类网址有 kaggle, 阿里巴巴天池大数据比赛、DataCastle 等, 教师可以在教学过程中讲授些竞赛案例, 引导学生参加各类数据挖掘竞赛, 有针对性的做一些竞赛类培训, 增强学生的学习兴趣。竞赛题目均具有较大的灵活性, 学生可以充分发挥其创造能力, 在参加数据挖掘竞赛的过程中, 会遇到很多的困难跟问题, 需要不断去尝试不同的数据预处理方法、数据挖掘算法, 由此来获得一个合理的、满意的结果。这是一个不断试错的过程, 需要不断的修正, 再尝试, 再修正, 在不断的探索中加深对数据挖掘理论知识的认识, 也是在培养学生的探索精神。竞赛是相互学习交流的过程, 本身也有高低之分, 即使再优秀的学生也会存在不足, 通过竞赛可以使学生更好的发现自身的不足, 明白“只有更好, 没有最好”的道理。

3.2 以赛促教, 赛教融合

在信息爆炸及数据指数化增长的今天, 为适应时代需求, 各高校应当根据人才培养需求, 致力于培养一批高质量的数据挖掘技术人才服务于社会。数据挖掘概念与技术的教学内容也应跟随时代发展不断调整, 以此来适应并服务于当前的社会环境, 因而对教师的要求也更高。

以赛促教, 可以推动数据挖掘技术在高校的推广和应用, 提高教师的知识技能及教学水平。当前数据挖掘应用主要集中在电信(客户分析), 零售(销售预测), 农业(行业数据预测), 网络日志(网页定制), 银行(客户欺诈), 电力(客户呼叫), 生物(基因), 天体(星体分类), 化工, 医药等方面, 教师能够获取的理论及实践教学案例有限, 不能满足实际教学需求。而挑战赛的题目主要是来源于企业, 管理机构和科研院所等的实际问题, 在带领学生参加比赛的过程中, 教师可以丰富自己的案例库, 形成自己的知识储备, 在教学的过程

可以将经典的竞赛题来进行讲解, 同时选取参赛过程中的代表性案例, 涵盖数据挖掘领域的各个不同方向。同时, 教师也能够清晰的了解到数据挖掘技术在实际案例的合理运用, 结合学生在竞赛过程中的表现, 明确在教学过程中的存在哪些优势与不足, 从各角度来总结重点知识, 形成自己的知识网络及知识储备库。

3.3 以赛促改

以赛促改, 改进教学过程中实践教学不足的问题, 推进数据挖掘概念与技术教学改革。在数据挖掘概念与技术的教学过程中, 教师应更加注重理论结合实践, 学好理论能够加深对实验结果的理性认识, 做好实验则能够增加对理论的感性体验, 因此, 二者能够相互促进, 共同提升课程的教学效果。各高校及相应学院可以设立与数据挖掘相关的比赛及实践项目, 加强学生对理论知识的理解, 同时也能增强学生的编程能力。同时也可以组织学生参加全国大学生数据挖掘大赛, 并对学生进行相应的培训, 使学生清晰的了解数据挖掘技能在大数据场景下的具体应用, 理解数据挖掘的本质。通过参加比赛, 学生能够从真实的案例中吸取经验教训, 牢记在实践活动中用到的算法相关的理论知识, 同时也能提升学生的程序设计能力。在教学的过程中, 教师也可以结合竞赛题, 将相关的知识点逐一分析讲解, 促进理论和实践相互融合, 提高学生综合应用知识的能力。

4 结束语

在信息技术快速发展的今天, 数据已经渗透到当今每一个行业和业务职能领域, 成为重要的生产因素。“大数据”在物理学、生物学、环境生态学等领域以及军事、金融、通讯等行业存在已有时日, 人们对于海量数据的挖掘和运用, 预示着新一波生产率增长和消费者盈余浪潮的到来。但目前数据挖掘概念与技术的教学虽然已在各高校普及, 但已经落后于时代的发展, 无法满足当前时代发展的需求, 亟需改革。本文对数据挖掘概念与技术一课的教学中的问题进行了分析, 提出了“以赛促学, 以赛促教”的改革方式, 竞赛有始终, 在竞赛的过程中让有形的知识转化为无形的数据挖掘技术学习的全覆盖, 更好的调动学生的学习兴趣, 提高教师的教学水平。时代在发展, 数据在增长, 亟需高质量的数据挖掘技术人才, 数据挖掘概念与技术的教学模式也应不停探索与发展。

参考文献:

- [1]周志华.机器学习[M].北京:清华大学出版社, 2016.
- [2]黄浩, 余琨, 刘树波.以实践和创新能力培养为导向的数据挖掘课程教学方法研究[J].教育教学论坛, 2019(49): 130-132.
- [3]刘波, 蔡燕斯, 钟少丹.大数据背景下数据挖掘课程实践教学的探索[J].高教学刊, 2019(18): 124-128.