

数据挖掘在大学英语教学质量分析中的应用研究

——基于中外合作跨境“2+2”双文凭项目（会计）的建立与认证课题

李 灿¹

（广州华立科技职业学院，广东、广州，511325）

摘要：《教育部关于推进高等职业教育改革创新引领职业教育科学发展的若干意见》（教职成〔2011〕12号）文件提出，鼓励和支持高等职业学校加强国际交流与合作，积极参与职业教育国际标准和规则的研究与制定，积极开展中外合作办学，引进优质教育资源，提升办学水平。本文基于中外合作跨境“2+2”双文凭项目（会计）的建立与认证课题，通过将数据挖掘技术应用于大学英语教学质量分析之中，试图发现、显化隐藏在教学过程中影响大学英语教学质量的重要因素群，为教学质量分析和有针对性改进教学提供一种新的思路和方法，以期构建大数据背景下更为科学、客观、针对性强的质量分析评估体系。

关键词：数据挖掘；大学英语教学；质量分析

引言

基于云、物联网、数据库网络技术的成熟应用，数据信息作为一种可捕捉、可量化、可传递的数字存在，已经引起教育领域的变革。如何区分数字与数据？如果一学生成绩100分，那么100就是“数字”，100背后的因素，如智力水平、家庭背景、学习态度等则是“数据”。数据的集中以信息技术的成熟为基础，它既是过程性和综合性的考虑，更是真实世界背后潜在逻辑关系的考量。在大数据时代，知识到智慧的跨越无处不在。在教育领域，经历了平台开放、内容开放、校园开放后的教育资源也将迎来教育的新一轮主流模式，即教育资源极其丰富、视频成为主要载体、翻转课堂普及、MOOC教学盛行等，并渗透到教育核心环节，革新教学思维，影响教育模式、重构教育质量评价方式（喻长志2013:188）。

一、大数据时代数据挖掘的基本理论

大数据最早“见诸笔墨”是在1980年，未来学家托夫勒的论著《第三次浪潮》誉其为“第三次浪潮的华彩乐章”。此后2012年美国政府启动国防部、能源部、卫生部、国家自然科学基金、地勘局及NASA“六部共研”的“大数据研究和发展”计划，标志着大数据时代的来到。它对这个时代所带来的影响力堪比望远镜让人类领略宇宙风采一般，其通过对潜藏在数据背后有用信息的挖掘分析让人们可以更好地诠释世界，其基于数据分析的决策将成为这个时代的核心特征^[1]。

数据挖掘（Data Mining, DM）是指从数据集中提取并显化潜藏于数据背后的有用信息。继1989年第11届国际人工智能会议上首现KDD术语（Knowledge Discovery in Database）后，学术界和工业界1995年成立的数据挖掘与知识发现大会将其发展成数据挖掘领域的国际顶级会议^[2]。1996年Fayyad和Piatetsky-Shapiro等在KDD国际会议上给予完整定义：KDD是从数据矿山中淘出“知识金块”的过程；DM则是“智慧挖掘”过程中的某个特定步骤，两者没有明确界线，均指从数据中挖掘知识的过程^[3]。

¹作者简介：李灿（1973-），男，汉族，广州，硕士，讲师，国际合作与交流中心主任，研究方向：外语教学和管理、国际合作办学

*基金项目：1. 2013年度广东省教育厅高等学校质量工程项目——跨境“2+2”双文凭项目（会计）的建立与认证（20130201070）研究成果。

2. 2014年度广东省教育厅教学成果奖培育项目。

二、数据挖掘在教育领域的应用状况

随着各行业对数据处理深度需求的增长,数据挖掘在教育领域的应用也不断扩大,通过数据挖掘显化潜藏在教学管理工作背后的有用信息能为教育决策服务。传统的教育起源于工业化并兴盛于工业化,其铃声、班级、标准化课堂、统一教材的流水线场景集中映射了工业化过程中经济批量和物流集中等特征于一体,并为工业时代制造了标准化人才。随着数据挖掘在教育领域的应用,其在学生信息管理、高校教学质量评估、高校教学管理水平以及学生学习能力等方面的影响必将引起教育领域的革命。

三、传统大学英语教学质量分析指标体系的缺陷

从20世纪90年代起,随着大学英语教学手段的多样化,如何提高大学英语教学质量日益引起各大高校的关注并陆续启动了外语教学质量保障体系,同行评价和学生评价两种方式在高校中得到广泛运用。传统的教学质量保障指标体系通常由多级指标构成,如一级指标包含教师教学方法、教学内容、教学素养、教学效果等方面,二级指标则是对一级指标的具体化^[4],并根据指标所起作用及导向力度分别赋予不同的权重形成完整的指标体系。事实上传统的质量分析方法仍存在一定缺陷:

(一) 传统的质量分析指标客观性欠缺

基于制定者经验的传统教学质量指标构成中主观、定性的成分较多,指标的权重分配也缺乏客观基础,甚至同一质量指标体系“依瓢画葫芦”套用在不同学科、不同类型的课程上。而实际上,只有显化隐藏在大学英语教学过程不易被发现的影响因素,并将这些因素纳入教学质量指标体系,才能设计出科学、客观的教学质量分析指标。

(二) 传统的质量分析结果导向性不足

Marsh H W在《教育心理学杂志》上阐述:传统教学质量评估不管是采用同行或学生评价,都不可避免地受主观因素影响。以往的评价方式固然能够获取各项指标的评价结果,然而至于这些指标如何影响、在多大程度上影响教学效果,则难以判断。作为一种“结果型”的评价方式,传统质量分析所得的只是纯粹的“数字”,却没能得到影响该数据背后因素的“数据”,即传统的质量分析结果只能回答教学效果“what”(怎么样)的问题,至于“why”(为什么会这样)和“如何提高”(how)等问题则无能为力^[5]。

四、数据挖掘在大学英语教学质量分析的支撑作用

(一) 校园网 E-learning 数据平台准备

E-learning 数据平台配备听、说、读、写、译等技能学习模块,各模块配置海量图文并茂、声色俱全的英语听说录音及视频教学片段,具体表现为“一库三台”。一库是网络化英语教学资源库,三台是网络化英语教学平台、学习平台及测试平台。“一库三台”最大化地丰富教学内容设计,解决老师时间有限、资源匮乏、手段单一的客观问题,使教学资源得到最优化共享,达到教和学的平衡。

教学过程中,学生遵循“准备阶段-确定项目任务-制定计划-实施计划-检查评估”等步骤选择不同模块,通过查阅、整理、汇总、演示后在数据平台接受检验。整个过程通过“三台”来完成。这在宏观上促使学生主动登陆平台进行线上搜索充实相关背景知识,观摩视频提高语言表达能力,线下通过团队协作解决疑难问题;在微观上通过教师进行“适当集中”教学为学生创造更高层次的语言活动条件。最后,数据挖掘学生上机情况检测大学英语教学质量。

(二) 数据挖掘应用过程

单元模块教学中,各模块均提供海量语言实训素材且知识点全面覆盖,并根据难易程度按同一级别下难度系数分三个等级,难度系数分别为0.3、0.5、0.7,学生上机登录输入性别,班级,专业,入学英语成绩,学习目的(通过考试,找工作,出国留学)后进行搜索、选择、整理、演示完成单元模块学习与测试。数据平台录入参与学生的登录频率,学习内容,对学习内容难度的评价来完成日常的GPA积累。数据平台同时学生单次成绩和历次成绩,成

绩查询目录依次是年度、班级、考试号。数据平台对所录入的数据通过决策树, 聚类分析、神经网络以及回归分析等技术进行数据分析挖掘出影响学生成绩背后的深层因素^[6]。

(三) 决策树算法在大学英语教学质量分析中的应用

决策树是通过节点分支连接成“层次树”, 它的每个分支结点代表一个分类问题, 每个叶子结点是带分类的数据分割, 根结点是数据集合空间, 从决策树的根结点到叶子结点的每一路径相对应类别预测, 用于分类学生的表现行为和期望。E-learning 数据平台通过录入学生的性别, 班级, 专业, 入学英语成绩, 学习目的(通过考试, 找工作, 出国留学)、出勤情况、上机频率, 将学生单元学习模块成绩分别录入到“高”与“低”两个数据集合, 并依据高、低两个不同训练子集所生成决策树进行修剪, 从而建立起与其成绩对应的或“高”或“低”的决策树模型。根据决策树生成分类规则可观察到“上机频率”对成绩较差学生影响是最大的, 至于那些从不上机或偶尔出勤的学生, 假如基础也一般的话, 那成绩也定然为“差”。这样, 借助决策树算法可以找出影响学生成绩背后因子, 从而用于指导教学, 提高教学质量。

这样, 数据挖掘在大学英语教学教学质量分析中的应用可帮助检测学生的参与情况、行为特征、学习成果、时间安排、空间支配自并对学生进行分类。这些分析结果对大学英语教学质量分析有着重要的参考价值。因此, 通过数据挖掘技术应用于教学质量分析可以客观、准确地发现、显化影响教学质量的重要因素群, 而这些“因素群”恰恰是构建科学、合理、客观的大学英语教学质量指标体系的前提和基础。基于此, 通过与传统的教学质量指标相结合, 其在教育领域的应用能为教学质量分析提供一种新的思路和方法, 有助于构建科学、客观、针对性强的质量分析指标体系。

五、结论与展望

数据挖掘在大学英语教学的应用是教育领域的革命, 它为大学英语质量分析指标体系的构建提供了更全面、科学、客观的技术路线, 有效地弥补了传统教学质量指标体系的不足。应用数据挖掘关键性数据有助于直观观察不同因素和环境下英语教学情况, 帮助决策者寻求规律和预测未来, 以制定更有针对性的教学改进策略指导教学与学习。

《孙子兵法》有言, 多算胜, 少算不胜。数据挖掘作为一种数据实证研究方法, 其在教育领域的突破和创新将产生更深远的影响。事实上, 我国当前大数据相关技术研究还不成熟, 大学教师现代教育技术素养不高, 区域间教育数据资源分布不均, 教育信息系统平台构建与发达国家相比也不完善。因此, 我们一方面要利用数据挖掘等前沿技术提升对教育数据资源的获取、管理、分析和应用; 另一方面要积极构建全面系统教育数据库来推动资源整合与共享, 为大数据在大学英语教育中的应用奠定技术和平台基础。

参考文献:

- [1][美]维克托·迈尔·舍恩伯格·盛杨燕, 周涛, 译. 大数据时代[M]. 杭州: 浙江人民出版社, 2013.
- [2]吉根林, 赵斌. 面向大数据的时空数据挖掘综述[J]. 南京师大学报(自然科学版), 2014, (3).
- [3] Fayyad U M, Piatetsky-Shapiro G, Smyth P. International Conference on Knowledge Discovery and Data Mining[C]. Portland, Oregon: AAAI press, 1996:82-88.
- [4]骆兰, 薛艳, 唐国强. 论高校课堂教学质量评估体系的构建[J]. 高教探索, 2006, (6):55-56.
- [5]Marsh H W. Students evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility [J]. Journal of Educational Psychology, 1984, (76):707-754.
- [6]郭茹. 数据挖掘技术在教育领域的应用[J]. 科技和产业, 2014, (9):122.

The Research about Quality Analysis on Project Teaching via Data Mining

Li Can

(Guangzhou Huali Science & Technology Vocational College, Guangzhou/Guangdong, 511325)

Abstract: 《The Ministry of Education document about suggestion on promoting higher occupational education reform & innovation as well as leading occupational education development》 (occupational (adult) education (2011) No. 12) indicates that international communication and cooperation should be encouraged and supported among higher educational school so as to take part in setting its international standards & rules, increase its international influence, introduce high-quality educational resources, and improve education quality. The thesis is based on *the Project of Sino-foreign Certificate Award*, utilizing data mining to discover and display the elements groups that affect English teaching quality, further providing a new pattern in quality analysis & teaching improvement to construct a more scientific, objective & specific quality evaluation system under the background of “Big Data Era”.

Key words: Data Mining; College English Teaching; Quality Evaluation

作者简介

李灿 (1973-)，男，汉族，广州，硕士，讲师，国际合作与交流中心主任，研究方向：外语教学和管理、国际合作办学；本文为广东省教育厅基金项目2013年度广东省教育厅高等学校质量工程项目——跨境“2+2”双文凭项目（会计）的建立与认证（20130201070）研究成果；也是2014年度广东省教育厅教学成果奖培育项目。