

■ 高教治理

DOI:10.15998/j.cnki.issn1673-8012.2024.03.002

基于评教文本挖掘的高校教学质量评价指标构建与应用



李建龙¹, 李立国²

(1. 中国人民公安大学 教务处, 北京 100038; 2. 中国人民大学 教育学院, 北京 100872)

摘 要:学生评教既是保障教学质量的重要活动,又是高校进行教学质量评价的主要手段。为提高教学评价效果,为授课教师和教学管理部门提供更有参考性的评教结果,区别于主要基于学生结构化评分数据进行的定量研究和基于非结构化评教文本数据进行的定性研究,基于大数据评教文本挖掘的研究方式被提上研究议程。以学生主观评语和客观评分两方面的评教数据为研究对象,采用层次分析法构建基于评教文本挖掘的高校教学质量评价指标体系,并通过线性回归方法对指标体系的科学性进行实证检验和应用分析。研究发现,海量学生评教数据兼具学术研究价值和实践指导意义。通过对学生评教文本进行挖掘和分析,以有效获取评教数据中的关键信息,如教师的教学态度、教学内容、教学能力、教学方法和教学效果等。基于层次分析法构建的 5 个维度 23 个层级的高校教学质量评价指标体系,能够较为全面、准确地利用学生评教数据对本科教学质量进行评价和解释。多元线性回归分析结果显示,该指标体系不仅能够了解学生重点关注的维度、预测学生评价分数并揭示教师教学质量与各个评价维度之间的关系,还可以应用于各类教师的教学质量评价,从而为教师提供精细化的教学诊断,为学校进行科学的教学管理决策提供数据咨询与服务支持。该评价指标体系的实际应用评价结果显示,不同类别课程存在的问题不同。为此,一方面要通过大数据、人工智能等新兴技术手段提高对学生评教数据的利用效率,深度挖掘海量学生评教数据背后的数据价值和学术研究价值,并加强与同行评教、督导评教和 AI 评教等不同视角评教结果的协同运用,开展基于学生中心和产出导向的教学评价应用;另一方面需要针对不同类别课程和不同年龄段教师的教学薄弱环节进行针对性教学改进,从而更好地服务高校教学质量的持续提升和学生的全面发展。

修回日期:20240409

基金项目:北京市教育科学“十四五”规划青年专项课题“高等教育数字化转型背景下学生评教文本的实证研究”(CDCA23126)

作者简介:李建龙,男,安徽阜阳人,中国人民公安大学教务处助理研究员,北京理工大学博士生,主要从事高等教育管理与教学评价研究;

李立国,男,山东滨州人,中国人民大学教育学院教授,博士生导师,教育学博士,主要从事高等教育理论与管理和教育思想史研究。

引用格式:李建龙,李立国. 基于评教文本挖掘的高校教学质量评价指标构建与应用[J]. 重庆高教研究, 2024, 12(3): 14-24.

Citation format: LI Jianlong, LI Ligu. Construction and application of university teaching quality evaluation index based on teaching evaluation text mining[J]. Chongqing higher education research, 2024, 12(3): 14-24.

关键词: 学生评教文本; 高校教学质量评价指标体系; 教学诊断; 教学改进; 学生发展

[中图分类号] G642.0 **[文献标志码]** A **[文章编号]** 16738012(2024)030014-14

一、问题提出

教学质量评价一直是保障高校教学质量的关键手段,也是高等教育教学改革的重点关注对象。2019 年,教育部《关于深化本科教育教学改革 全面提高人才培养质量的意见》进一步强化学生评教的作用,强调构建以学生评价为核心的教学质量评价体系的重要性。2020 年,《深化新时代教育评价改革总体方案》又进一步明确了深化教育评价改革和提升教学质量的具体要求。目前,我国高校已普遍建立包括学生评教、同行评教、督导评教及 AI 评教等多方面的教学质量评价体系。同行评教通过教师之间的相互评估,增加了教学方法的专业性和多样性;督导评教从管理层面监督教学质量,保证教学标准的执行;AI 评教通过技术手段提供客观数据支持,促进教学方法的创新和优化;学生评教作为高校教学质量评价应用最广泛的一种方式,已成为高校评价教学质量的主要手段,在教育评价改革过程中扮演着重要角色。

学生评教最初主要依靠线下的纸质问卷进行,但随着教育数字化的发展,网络评教已成为高校学生评教的主流形式。学生评教引起国内外研究者的广泛关注,研究内容涵盖基础理论、指标设置、评教工具设计以及评教有效性分析等多个方面。随着学生评教数据的累积和“以学生为中心”教育改革要求的提出^[1],对学生评教数据的深度挖掘和有效利用逐渐成为研究的热点。事实上,学生评教数据是学生对教师教学最直接、最有效的反馈,蕴含着学生对教师教学多方面表现的具体态度^[2],真实、客观地反映了学生的基本诉求以及对任课教师教学情况的积极或消极情感倾向。深度挖掘这些文本信息有助于从学习者的视角识别教师授课中存在的具体问题,进而为教师的教学改进提供决策依据,保障高校教学质量的持续提高。

二、理论基础与研究回顾

学生评教是指学生根据特定的教学要求和评价标准,对教师的教学方法和自身学习过程进行评价的行为。学生评教作为高校教学评价和教育质量保障体系的重要组成部分,是促进本科教育高质量发展的重要手段^[3]。学生评教是“以学生为中心”教育理念在教学过程中的具体体现,强调学生在高等教育教学改革中的主体地位,关注学生的发展需求。因此,高校应主动从“课堂、教师、教材”三中心向“以学生为中心”转变,坚持“以学生发展为中心、以学生学习为中心、以学习效果为中心”。具体应用到教学评价中,就是要坚持以学生学习效果而不是以教师教学效果作为评价标准。在高等教育领域,学生评教数据通常包括结构化的评教数据和非结构化的文本评价数据两种形式。国内外关于学生评教的相关研究成果也主要集中在结构化的评教分数研究和非结构化的评教文本研究两个方面。

(一) 关于结构化学生评教分数的研究

目前,把结构化的学生评教分数作为对教学质量进行检测与管理的一种制度安排已被我国高校普遍采用^[4],即学生结合对教师授课的整体印象及收获给出一个具体分数。因此,我国关于大学生评教方面的研究主要集中在结构化的评分数据上,通常使用大数据分析工具对评分数据及其相关属性进行关系研究和信息挖掘。例如,刘忠轶等通过残差分析方法建立不同课程之间的学生评教比较

标准,为科学衡量不同课程教学质量和创新学生评教制度提供了学理支撑与经验证据^[5]。何喜军等以某高校 374 门课程的评教数据为例,利用统计分析方法研究学生评教数据中的关联规则^[6]。钱梦岑通过构建长面板数据样本库,运用固定效应模型控制课程与教师的异质性并处理序列相关问题^[7]。周继良基于制度分析的视角对高校学生评教行为偏差的影响因素进行了实证研究^[8]。程安林等分别测算课程内一致性、学生间可靠性和项目间可靠性指标,对某校学生评教结果信度进行了实证研究^[9]。马莉萍等利用某高校连续 6 年同一门课程由不同职称教师授课的学生评教数据,分析了不同职称教师教学质量的差异^[10]。刘玉静等以学生学习投入与收获理论为框架,结合中国大学生学情,构建以学生为中心的高校课程教学质量评价模型^[11]。戴璨等基于某高校近 5 年来 600 余门课程的学生评教数据分析了非教学因素(教师背景、班级属性和课程)对高校课堂效果的影响^[12]。同时,也有学者探讨了高校学生评教的效果和方法,如学生评教的信度、效度、影响因素及应用风险等^[13-14]。

尽管结构化评教分数在评估学生学习效果、课程评价以及满意度等方面具有实验验证和模型构建的实践意义,但是这种方法的局限性也很明显。具体来说,教学质量是一个多维度、复杂的概念,包括教学方法的多样性、个性化教学的实施、课程设计的有效性等,而结构化评教分数反映的仅是学生对教师教学表现的整体评价,难以支撑对教师教学的全面性评估和深入分析。因此,越来越多的研究者意识到单纯依赖评教分数无法完整地捕捉教学的丰富性和多样性,需要更多维度、更多层面的考量,以便更全面地理解教学质量和教师的教学效果。

(二)关于非结构化学生评教文本的研究

与结构化学生评教分数的研究相比,国内学者对非结构化学生评教文本的研究、利用和挖掘相对较少。然而,非结构化的学生评教文本实际上蕴含着大量的倾向性观点和语义信息,能够更为真实地反映学生本人的基本诉求以及对任课教师教学情况较为客观的描述。通过自然语言处理技术,分析大量文本数据并提取情感、主题等信息,可为教师提供更具体的反馈,促进个性化的教学生成。

国外关于学生评教的研究已经关注到学生评教的文本挖掘,学生评教文本的收集大部分依赖国家统一开发的教师评价网站,如 ratemyprofessors.com 等。例如,Rybink 等基于评教网站与评教问卷收集学生对教师、课程的文本评价,指出机器学习尤其是深度学习模型将很快成为学生评教数据处理的流行工具^[15]。Gottipati 等采用决策树算法从学生的定性评价中提取课程建议,使学生反馈的建议得以采纳并最终改善教学^[16]。Tashchian 等通过分析 21 个学期、240 门本科课程的学生评教数据发现:相比于兼职教师,学生普遍认为全职教师具有更深厚的知识背景和更高的教学热情^[17]。Stupans 等借助语义分析工具 Leximancer 进行评教文本挖掘,以分析纯定量评分数据识别不出的问题,从而对教学信息与教学效果进行更深入、细致的研究^[18]。

与国外评教文本的数据来源有所不同,国内一般通过相对独立的高校学生评教系统收集学生的评教文本并开展研究。例如,何喜军等基于贝叶斯算法对学生评教文本进行情感分析,通过不同课程文本的语义特征得出学生评教的影响因素^[6]。陈玉婵等采用机器学习与情感词典相结合的方式将学生评教文本分为积极与消极两类,对上述两类观点进行抽取与聚类,弥补了高校原始学生评教文本难以利用的不足^[19]。

综上,已有研究虽在学生评教文本分析方面取得一定进展,但大多数仍集中于评分数据的表层分析,而未能深度挖掘文本数据以全面理解教师的教学效果。目前的研究多依赖情感词典、机器学习或二者结合的方法,在深度学习等先进技术的应用上还有待改进。大多数研究聚焦于算法的准确性或

从单一视角进行文本分析,未能实现评教文本的多方位深度挖掘,从而限制了教学过程多维信息的全面展示。教育数字化转型为学生评教文本的深度挖掘提供了机遇,数字化转型不仅能够提供丰富的教育数据资源,还能够通过先进技术赋能学生评教文本的深度分析,为教学改进提供更精准的决策依据。因此,深度挖掘文本数据并构建全面、精确的教学质量评价指标体系,对准确理解学生真实感受和精确诊断教师教学质量意义重大。

三、研究方法 with 数据处理

文本挖掘(text mining)这一概念最早由 Feldman 等人在 1995 年进行非结构化数据研究时提出。文本挖掘方法涉及多个领域,包括文本分类、信息提取、概念和实体发现,以及情感分析等。然而,单纯使用词频进行教学文本分析与教学评估较为笼统,需要进一步细化相关教学评估指标体系。

(一)研究方法

1. 层次分析法

层次分析法由美国运筹学家 Saaty 于 20 世纪 70 年代提出,是一种将与决策问题有关的元素分解成目标、准则、方案等层次,并进行定量和定性分析的决策方法^[20]。层次分析法通过将定性的指标进行模糊量化来计算各个层次的单排序(权重数)和总排序,从而对一些多目标、多准则类型的方案进行优化^[21]。近年来,层次分析法在教学评价方面得到广泛应用。在教学评价中,层次分析法可以用于确定评价指标的权重和优先级,从而更全面、客观地评估教学效果。通过层次分析法,教学管理者可以将教学评价的指标进行层次化排列,确定每个指标在评价中的相对重要性。总之,层次分析法能够为教师教学提供一套科学、系统的评价框架,为管理者提供决策支持,促进教学质量的提升^[22-24]。本文基于层次分析法构建指标体系,将教学评语对应多个层次,构建一级、二级指标,形成树状多因子模型并确定指标体系各因子所占权重,以期更加科学地对教学效果和教师教学水平进行评估。

2. 多元线性回归分析

回归分析是用来确定两个或多个变量之间定量关系的一种统计方法。在回归分析中,如果自变量和因变量之间的关系可以用一条直线表示,那么可以称之为线性回归。如果在回归分析中有多个自变量,且自变量和因变量之间也为线性关系,则可称之为多元线性回归分析。线性回归的过程就是把各个自变量和因变量的个案值带入回归方程式,通过逐步迭代与拟合,最终找出回归方程式中的各个系数,构造出一个能够尽可能体现自变量与因变量关系的函数式。

在预测评教分数时,学生对教师的评语与分数之间具有相关性,因而评语转换的各个评价指标分数与评教分数也具有相关性。选取学生评语的 5 个维度分数作为自变量,所有学生评教的平均分作为教师评教分数,使用多元线性回归研究其关系,得到预测评教分数的回归方程。另外,通过最小二乘法寻找数据的最佳函数匹配^[25]。

(二)数据处理

1. 数据来源

本文数据主要来源于某大学本科教务系统中的学生评教数据,包括主观评语和客观评分。主观评语指的是学生对某教师某学期授课的实际表现发表合理化建议或评论;客观评分指学生对教师整体的授课表现进行的百分制打分。本文共收集近 8 个学期开设的全部课程关于学生对授课教师的主观评语与客观评分数据,共计 541 505 条记录。数据时间跨度大,覆盖范围广,内容丰富且具体,部分数据示例如表 1 所示。

表 1 评教数据示例

课程编码	评教分数	学生评语
081646	95	生动形象,态度严谨,与实践相结合,教学质量高
211606	85	老师课堂上积极引导学生向上,但常常不能理解学生的意思,有些自我式讲课
081636	90	老师从实际执法角度出发,结合自身经历,讲述现场执法的规范操作和要点
211603	95	思想与时俱进,思维严密,给人博学多识的感觉

2. 数据清洗

本文使用人文社科数据计算工具 ROST CM6 对原始的学生评教文本进行预处理,数据清洗主要包括删除简短评论、删除非汉字评论、删除符号类评论和人工文本清洗 4 个环节。一是删除简短评论。由于主观评语是必填项,“无”“好”“不错”等简短的评论往往是学生随意填写。这样的评语所表达的意见含量和价值较低,为保证实验效果,首先删除 3 个字以内的简短评论。二是删除非汉字评论。评教文本中存在英语或日语等非汉字评论,主要原因是该类课程为语言类课程。此类评论的数据量并不是很大,而且大多为“good”“nice”等低价值词汇。删除非汉字评论,避免文本分析产生错误,影响分析结果。三是删除符号类评论。评教文本经常存在“颜文字”,例如“666”“(⊙o⊙)”等内容。这些“颜文字”一般由标点符号和运算符号构成。为减少文本分词产生的错误,对符号类非规范用语的评论进行删除。四是人工文本清洗。在评语大于 3 个字的评教数据中也存在大量类似“讲得很好”“老师很好”等低价值评论,这类评语不能通过文本长度过滤直接删除。为保证实验质量,本文通过人工核验对这部分数据进行清洗。

3. 文本分词

文本分词指的是将每个汉字序列切分为一个个单独的词。目前已经有比较成熟的分词工具,如 ROST 系列人文社科研究数据计算工具、玻森的 Boson NLP 等。其中 ROST CM6 软件的界面简洁,具备易上手操作、分析效果好等特点,因此本文的学生评教文本分词由 ROST CM6 来完成。

首先删除停用词。在学生评教文本数据中存在很多介词、副词、连词等,它们对词语分析并没有实际意义,如“地”“和”“的”等,还包括一些“是”“就”“让”等没有明确意义的词语。这些停用词的词频过高会增加实验的工作量且对结果分析造成干扰,所以需要将其删除。随后通过 ROST CM6 对评教文本进行分词,部分学生评教文本的分词结果如表 2 所示。

表 2 分词结果

学生评语原文	分词后的评语(部分)
老师工作认真负责	工作 认真 负责
这门课程老师讲授得生动有趣、引人入胜	生动 有趣 引人入胜
希望老师可以增加一些更有实战挑战的课堂内容	希望 可以 更加 实战 课堂 内容
李老师知识渊博、博闻强识、讲课既幽默又风趣	知识 渊博 博闻强识 讲课 幽默 风趣

4. 词频统计

本文使用 ROST CM6 软件中的汉语词频统计功能对已经分词后的学生评语进行词频统计。在统计结果中存在“老师”“课堂”“大一”等非主观性词汇,词频占比大且对分析无意义,本文对该类词汇进行人工清除。最终得到有效高频词汇 636 个^①,如表 3 所示。

表 3 高频词汇统计表

① 高频词最高和最低的分别为 67 971 次和 31 次。经排查,该 636 个高频词汇均为正向的评价词汇,未发现恶意的评价词汇进入高频词汇范围。

排名	高频词语	词频
1	认真	61 971
2	内容	40 569
3	负责	40 269
4	丰富	20 690
5	生动	18 335
6	态度	17 962
7	知识	17 651
8	有趣	15 373
9	可以	13 686
10	讲解	10 540
.....		
636	立德树人	31

四、评价指标体系的构建、检验与应用

(一) 基于评教文本的教学质量评价指标体系构建

1. 基于层次分析法的评价指标体系构建

在使用层次分析法构建指标体系时,首先确认教师教学评价的总目标,从总目标出发构建准则层和方案层。根据对 20 所不同层次高校学生评教指标体系进行归纳统计,结合词频统计分析的结果,确认准则层的 5 个一级指标分别为教学方法、教学内容、教学能力、教学态度和教学效果。其中,教学方法指教师在教学过程中采用的帮助学生更有效地学习和掌握知识的一系列策略、技巧和活动;教学内容指在教学过程中传授给学生的知识、技能和价值观等;教学能力指教师的专业知识能力、语言表达能力等;教学态度指教师在教学过程中所表现出来的对待学生、教学工作的态度;教学效果指学生在接受教育后所得到的知识、技能和素质的提升程度^[26]。

方案层是准则层的具体内容,即一级指标下的二级指标。以教学内容为例,教师和学生较关注的维度包括教学内容是否全面、创新,重难点是否突出等方面。本文构建的评价指标体系由准则层 5 个一级指标,以及方案层 23 个二级指标组成。词频可以在一定程度上反映学生对于各个评教指标的关心程度,各个指标下的词频统计结果如表 4 所示。

表 4 指标体系与词频统计

一级指标	二级指标	词频统计
教学方法(A1)	案例分析讲解、例题讲解、讲解示范(B1)	26 984
	理论与实践相结合、让学生动手实操(B2)	23 038
	信息技术与教育教学深度融合(B3)	16 024
	与学生互动、引导式启发式教学(B4)	20 991
	寓教于乐、因材施教、活跃气氛(B5)	50 043
教学内容(A2)	内容前沿性强、内容创新(B6)	8 082
	内容全、内容丰富信息量大(B7)	39 030

续表

一级指标	二级指标	词频统计
	突出立德树人、注重思政元素融入(B8)	7 835
	延伸拓展资料丰富(B9)	13 858
	重难点突出、紧扣教学目标(B10)	35 211
	专业性强(B11)	11 029
教学能力(A3)	教学设计能力:逻辑性强、课件制作精良、板书(B12)	12 795
	课堂驾驭能力:主导课堂、引导学生、时间分配合理(B13)	18 021
	语言表达能力:滔滔不绝、语言幽默、层次感(B14)	51 216
	专业知识能力:学富五车、知识丰富、旁征博引(B15)	41 554
教学态度(A4)	备课充足、批改作业认真、课后辅导负责(B16)	51 448
	课上精神饱满声情并茂、仪表端庄、教态自然大方(B17)	18 328
	认真、严谨、细致(B18)	95 249
	态度和蔼、亲切(B19)	13 736
教学效果(A5)	动手实践操作运用能力(B20)	10 717
	理论专业知识(B21)	39 448
	情感态度价值观(B22)	48 109
	职业生涯、未来就业(B23)	25 338

2. 判别矩阵构建及权重的求解

结合上述学生评教结果对指标的关切程度,由本领域内专家组成专家组对指标的重要程度进行投票,根据投票统计结果构建一、二级指标的判断矩阵^[27]。具体来说,每一位专家要对每一个判断矩阵中的每一对元素的重要性进行判定并赋值,最后取其平均值。使用 Matlab 软件计算判断矩阵最大特征值以及对应的特征向量,通过特征向量法计算出权重向量。同时,还要计算每一个判断矩阵的一致性指标 CI,查询一致性指标 RI 表,得出一致性比例 CR。若 CR 小于 0.10,则认为判断矩阵通过一致性检验,否则需要重构判断矩阵,直到通过一致性检验。本文中各个判断矩阵均通过一致性检验,一级指标的判断矩阵结果如表 5 所示。

表 5 一级指标判断矩阵

	教学方法	教学内容	教学能力	教学态度	教学效果
教学方法	1	2	1	1/2	1
教学内容	1/2	1	1	1/2	1
教学能力	1	1	1	1/2	1
教学态度	2	2	2	1	2
教学效果	1	1	1	1/2	1

依据判断矩阵,得到一级指标 5 个维度的权重以及 5 个评价维度下的二级权重后,由一级指标与二级指标相乘,得到各个指标的综合权重。各指标如表 6 所示。

表 6 各级指标权重

一级指标	一级权重	二级指标	二级权重	综合权重
------	------	------	------	------

重庆高教研究 2024 年第 3 期		在线投稿: http://yxxy.cbpt.cnki.net/		
A1 (教学方法)	0.193 9	B1	0.20	0.039
		B2	0.14	0.026
		B3	0.08	0.016
		B4	0.10	0.019
		B5	0.48	0.094
A2 (教学内容)	0.145 9	B6	0.09	0.013
		B7	0.32	0.047
		B8	0.08	0.012
		B9	0.09	0.013
		B10	0.32	0.047
		B11	0.10	0.014
A3 (教学能力)	0.165 1	B12	0.09	0.016
		B13	0.11	0.018
		B14	0.42	0.070
		B15	0.37	0.062
A4 (教学态度)	0.330 1	B16	0.24	0.081
		B17	0.11	0.037
		B18	0.57	0.187
		B19	0.08	0.025
A5 (教学效果)	0.165 1	B20	0.10	0.016
		B21	0.28	0.046
		B22	0.47	0.078
		B23	0.16	0.026

由表 6 可以得出如下结论:

从 5 个维度看,教学态度对应的一级指标 A4 的权重最大,达到 0.3301,表明教师的教学态度是学生打分的重要参考。在评价教师的教学方法时,学生优先关注的教学方法是寓教于乐、因材施教和活跃气氛(B5)。在评价教师的教学内容时,学生最为关注的是教学内容是否充分(B7)和重难点是否突出(B10)。在评价教师的教学能力时,学生关注的重点是教师的语言表达能力(B14)和专业知识能力(B15)。在评价教师的教学态度时,学生重点关注教师在教学中是否认真、严谨、细致(B18)。在评价教师的教学效果时,学生更为注重自己在理论专业知识(B21)和情感态度价值观(B22)方面的收获。

(二)基于学生评教的教学质量评价指标体系检验

在对学生评教进行主观文本评价的同时,还需要对其进行量化打分,促使评分与评语在理论上具有一致性。通过分析学生的评语与评分之间的关系,将评教文本转化为 5 个维度的分数作为自变量,学生评分作为因变量,使用线性回归的方式进行拟合并对回归结果进行分析,以检验文本分析得到的教学评价指标体系的准确性与解释力。

1. 评教文本分数计算

将评教文本转化为 5 个维度的分数作为自变量。首先通过情感分析计算每位教师在 23 个二级

指标上的得分,并统计各二级指标中的正向情感词的词频和负向情感词的词频(正向词记为1分,负向词记为-1分),得出初步情感得分;再将情感得分除以有效评教人数并乘以1 000做归一化处理,得到每位教师在23个二级指标上的分数;最后将得到的各个二级指标分数与层次分析法得到的对应二级权重相乘,得到5个一级指标的分数,也就是5个自变量。该分数比文本数据更加直观地用于评估某位教师或某些教师群体的教学水平与质量。

2. 基于线性回归的教学评价指标体系检验

将5个一级指标的分数作为自变量,将该教师的学生直接打分作为因变量,使用最小二乘法进行多元回归分析,以验证指标体系评分的合理性和准确性。

$$y = 89.397 + 0.023x_1 - 0.031x_2 + 0.021x_3 + 0.019x_4 - 0.037x_5$$

假设评教分数 y 受到5个评价指标分数 x_1, x_2, x_3, x_4, x_5 ,以及常数和随机因素 δ 的影响,则 y 与 x_1, x_2, x_3, x_4, x_5 以及 δ 满足以下线性关系:

$$y = a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5 + \delta$$

本文使用SPSS 27.0分析得到回归方程,方差分析如表7所示。

表7 方差分析表

ANOVA					
	平方和	自由度	均方	F	显著性
SSR	54.750	5	10.950	4.695	0.027
SSE	18.658	8	2.332		
SST	73.408	13			

该指标体系回归平方和 SSR = 54.750、残差平方和 SSE = 18.658、总平方和 SST = 73.408。 R^2 用来度量指标体系的拟合度,其范围通常在0到1之间,表示指标体系解释的因变量的比例。该指标体系的 $R^2=0.746$,可见该回归方程的拟合程度较好,所选的自变量可以解释因变量的74.6%,因此教师在5个维度的得分数能在很大程度上体现学生对教师的评价,进而反映教师的教学水平。F统计量是用来测试指标体系中所有回归系数是否全部等于0(即指标体系没有预测能力),该指标体系的F值为4.695。显著性(P value)是基于F分布得出的F统计量的概率值,如果这个值小于某个预定的阈值(通常是0.05),可以认为指标体系具有统计显著性。该指标体系的显著性为0.027,表明5个自变量之间存在显著差异。

此外,本文还使用误差均方根、绝对误差平均、相对误差绝对值平均对预测指标体系进行评价,并进行实验验证。随机选取五分之一的学生评教文本,计算文本中关于教学质量评估的预测评价指标。结果显示,该评价指标体系的误差均方根为1.12,相对误差绝对值平均为0.012,均比较低;绝对误差平均为1.12,表明预测得出的学生评分分数与实际值的平均误差为1.12。由此可以从另一方面证明该教学评教分数预测评价指标体系相对准确,可以与学生评分相互印证。

(三) 基于学生评教文本的教学质量评价指标应用

在检验了教学质量评价指标分数与学生评教分数的一致性后,我们还分析了不同类型课程的评价指标分数差异和不同年龄段教师的评价指标分数差异。

1. 不同类型课程的评价指标分数差异

由于学科专业之间的差异,教师的授课方式有所不同,学生的接受效果和欢迎程度也不同。本部

分从5个一级指标对思政类课程、计算机类课程和专业类课程教师的教学质量评价结果进行横向比较。具体流程:对每个类别的课程随机选取40位教师,计算每位教师在5个维度的得分和每个类别课程教师的平均分(见表8),在此基础上分析3类课程在5个维度的差异。

表8 不同类别课程教师教学质量评分

	教学方法	教学内容	教学能力	教学态度	教学效果
思政类	179.8	136.9	228.3	227.1	196.9
计算机类	107.0	94.2	101.8	136.8	129.3
专业类	131.1	98.2	236.3	199.6	145.7

从5个维度看,思政类课程教师在教学方法、教学能力、教学态度和教学效果4个方面得分较高,在教学内容方面得分相对较低,可能存在内容陈旧、与实际情况联系不紧密等问题。所以,思政课教师在进行案例教学时应当尽量与当前社会热点、学生生活实际相结合。计算机类课程教师在教学态度方面得分相对较高,但其他方面的得分不高,表明计算机类课程教师的教学方式有待改进,教学能力有待提升。为此,教学管理部门应注重提升计算机等基础类课程教师的教学能力,鼓励教师更新教育观念、创新教学方法、提升教学质量。专业类课程教师在教学能力与教学态度方面得分较高,但在教学内容与教学方法方面得分不高,说明专业课教师在专业能力方面表现优秀,但在教学方法和课程内容呈现方面有待改进。因此,专业类课程教师应紧密结合行业需求,进一步提高课程内容的实用性和针对性,同时改进教学方法,采用案例分析等多元化的教学手段,提高学生的实践能力。

2. 不同年龄段教师的课程教学评价分数差异

本文还针对不同年龄段的教师进行挖掘剖析。具体来说,将40岁以下的教师界定为年轻教师,将40-50岁的教师界定为中年教师,将50岁以上的教师界定为年长教师,据此归类并随机选取100名教师,计算其在5个维度上的平均分,最终得分如表9所示。

表9 不同年龄段课程教师教学质量评分

	教学方法	教学内容	教学能力	教学态度	教学效果
年轻教师	122.5	78.3	143.5	243.1	135.7
中年教师	128.1	97.0	163.7	148.1	163.4
年长教师	103.3	97.1	160.8	94.4	111.9

总体来看,相比于具有活力的年轻教师,年长教师受学生欢迎度较低。分维度看,年长教师在教学方法、教学态度和教学效果3个方面的得分较低,在教学内容和教学能力方面的得分较高,反映了年长教师普遍具有丰富的教学经验,但在教学态度上可能随着年龄增长逐渐有些消极。年轻教师在教学态度上显得尤为认真,但由于教学经验不足,导致其教学能力有待进一步提升。中年教师既具备一定的教学经验和知识储备,又有精力去引导学生、调动课堂氛围,所以其在教学效果方面的表现尤为突出。因此,教学管理部门应当关注年轻教师的教学效果、教学内容和教学能力,通过开展优秀教师教学经验分享等活动促使年轻教师向中年教师和年长教师学习教学经验,提升教学质量。年长教师则应与时俱进,转变教学模式,改变教学态度,或者通过配备助教等方式缓解教学工作压力,提高学生的认可度。

五、结论与讨论

高校学生评教结果的应用既是学生评教作用得以实现的重要途径,又是学生评教管理功能有效发挥的基本前提^[28]。本研究不仅深化了对评教文本潜在价值的理解,也为构建全面、准确的教师教学质量评价指标体系提供了新视角,还揭示了不同类别教师在教学质量评价方面的差异性,为高校的教学质量提升和人才培养提供了参考。

(一)学生评教文本中蕴含着巨大的学术研究价值和教学改进意义

当前多数高校在对教师教学质量进行评价时,通常直接使用评教打分结果,对教师具体在哪方面教学能力的优劣缺乏精确刻画,包括学生对教师教学质量多维度评价的海量评教文本被“束之高阁”。评教文本不仅反映了学生对教学的直观感受,还包含着对教师教学方法、课程内容及其有效性的深层次评价。这些深层次的信息对于理解学生的真实需求、评价教学效果以及进一步提升教学质量具有重要价值。为此,实践中要通过大数据、人工智能等新兴技术手段提高对学生评教数据的利用效率,深度挖掘海量学生评教文本背后的数据潜能和学术研究价值。后续研究应集中于深度挖掘评教文本中的潜在价值,特别是在高等教育数字化转型向纵深发展的今天^[29],应充分利用数字技术赋能评教文本数据潜在价值的挖掘。

基于学生评教文本的教学评价只是改进教学质量的一个视角,想要更加全面、科学地窥探教学质量,还需要结合其他类型的评教手段和结果,如同行评教、督导评教、AI评教等,以更好地落实以学生为中心的教学理念。事实上,不同评教方式所提供的独特视角及其评教结果,有助于教学管理部门通过多个维度和视角交叉印证,更加全面地评估教学质量。当然,不同类型的评教方式在评估教师教学效果时都具有不可替代的作用,需要教学管理部门在实践中寻求平衡和综合运用。其中,学生评教重点在于收集学生的直接反馈,反映教师的教学方式和内容是否满足学生的发展需求;同行评教侧重同行的专业评价,包括教学内容的深度和广度、教学方法的创新性等方面^[30];督导评教更注重教学过程和结果的质量控制,确保教学活动符合教学大纲和学校的教育目标;AI评教通过算法分析大量的评教数据,提取关键信息,能有效弥补人工评教的不足。因此,未来应注重多种评教方式相结合,从而为高校教师教学提供更为全面和客观的教学质量反馈。

(二)基于文本挖掘技术的评价指标体系在解释教师教学质量方面更具全面性和准确性

相较于传统的评教方法,基于评教文本挖掘构建的高校教学质量评价指标体系以学生为中心,可从多个维度对教师的教学质量进行评估。与此同时,采用线性回归方法验证了评价指标体系与学生打分的一致性,结果表明该评价指标体系可以科学、准确地评价教师教学质量的优劣,即该指标体系不仅能够了解学生重点关注的维度、预测学生评价分数并揭示教师教学质量与各个评价维度之间的关系,还可以将评价指标应用于各类教师的教学质量评价,明确不同类别课程教师、不同年龄段教师在教学质量各维度方面存在的差异,从而为教师提供精细化的教学诊断,为学校教学管理部门进行更为科学的教学管理决策提供数据咨询与服务支持。

可见,运用数据分析和文本处理方法构建的教学质量评价指标体系,能够更全面地捕捉评教文本中的关键信息,不仅提高了评价的真实性和客观性,还有助于发掘改进教学的具体方向和策略。为此,后续应进一步应用和推广基于评教文本挖掘的高校教学质量评价指标体系。

(三)不同学科类别和年龄段的教师在教学质量评价结果上存在显著差异

该指标体系的实际应用评价结果显示,不同类别课程存在的问题不同。具体来看,思政课教师在

教学方法、能力和态度方面得分较高,但教学内容上需加强与学生实际生活的结合。计算机类课程教师虽在教学态度上表现良好,却在教学方法和能力上有所不足。专业类课程教师在教学能力和态度上得分高,但在教学内容和方法上需作出改进。在年龄段方面,年长教师在教学内容和能力上表现出色,但教学方法和态度略显落后,需要更新教学模式;年轻教师态度积极,但有待经验积累和能力提升;中年教师在教学效果上表现最好。为此,需要针对不同类别课程和不同年龄段教师的教学薄弱环节进行针对性教学改进,从而更好地服务于高校教学质量的持续提升和学生的全面发展。

六、结 语

教育数字化时代的到来,推动文本挖掘技术在教学质量评价方面得到广泛应用。通过对学生评教文本进行分析和挖掘,可以获取评教数据中的关键信息。本文基于文本挖掘分析与层次分析法构建了一个科学且富有解释力的教学质量评价指标体系,从教学态度、教学内容、教学能力、教学方法和教学效果 5 个维度对教师进行全方位评价。对教师教学质量的评价有很多维度,学生评教只是其中之一。这要求教学评价方法需要多角度反映教学效果,通过多种评教方法的综合运用共同构建一个全面、多维度的教学评价体系,确保每位学生都能获取既定的学习效果,提高教学质量和人才培养质量。未来应全面落实学生中心、产出导向、持续改进的学生评教理念,加快形成以学校为主体,教育部门为主导,行业部门、学术组织和社会机构共同参与的中国特色、世界水平的高校教学质量评价保障体系,推动以学生学习成果为导向的中国教育实践典范的建立。

当然,本研究也存在一些不足:一是研究对象仅基于某校学生评教数据,虽然评教数据也涉及 8 个学期 50 余万条数据,但是数据来源较为单一。未来研究需整合更多样本高校的数据,以获得更全面和更多元的评教信息,从而提高研究结果的普遍性和准确性。二是学生评教的文本数据质量问题。由于对学生提交的评教文本没有提出规范性要求,导致评教文本数据质量不高,大量无意义或过短的文本在数据清洗阶段被删除。未来研究应采用更有效的学生评教文本采集方法,使用引导性语言以获取更高质量的评教文本,如“请从教学态度、教学内容、教学能力、教学方法、教学效果 5 个方面对教师教学进行评价”。三是需要加强学生评教文本的横向推广应用和纵向深入研究。在横向推广应用方面,鉴于很多学校的评教系统中都存在学生评教文本数据,因此本文所构建的教学质量评价指标体系在一定程度可应用推广到其他高校;在纵向深入研究方面,未来可尝试采用行动研究和质性研究方法,将本研究构建的评价指标体系和结论应用于某个教师某个学科专业的教学质量评价与改进。

参考文献:

[1] 龙洋,王玲.论高校学生评教的“教学学术”价值之维[J].当代教育科学,2019(8):710,32.
[2] 张津.基于学生评教文本的细粒度情感分析研究[D].广州:华南理工大学,2020:43.
[3] 张雨晨,李勉.高校学生评教的理想和现实:基于利益相关者的视角[J].中国考试,2022(9):2938.
[4] 刘理,宁立伟.高校学生评教的制度正义及其实现[J].中国高教研究,2022(8):8489.
[5] 刘忠轶,于洪鉴.高校学生评教的实证研究[J].中国大学教学,2021(8):5456.
[6] 何喜军,朱相宇.高校学生评教数据深度挖掘的实证研究[J].黑龙江高教研究,2019,37(10):8588.
[7] 钱梦岑.学生评教对高校教师教学行为的影响:基于国内某高校学生评教结果的实证研究[J].中国高教研究,2018(10):8486.
[8] 周继良.高校学生评教行为偏差影响因素的实证研究:基于制度分析的视角[J].高等教育研究,2018,39(2):59-72.

- [9] 程安林,张亮. 高校学生评教结果信度实证研究[J]. 黑龙江高教研究,2016(7):2425.
- [10] 马莉萍,熊煜,董礼. 职称越高,教学质量越高?——高校教师职称与学生评教关系的实证研究[J]. 教师教育研究,2016,28(6):8389,121.
- [11] 刘玉静,杨洋. 基于学习投入的高校课程教学质量评价:基于华东地区 17 所本科高校的实证研究[J]. 教育发展研究,2019,38(9):2733.
- [12] 戴璨,苗璐,朱恒,等. 非教学因素对高校课堂效果的影响及其启示:基于学生评教数据的实证分析[J]. 高等教育研究,2017,38(5):7280.
- [13] BROWN R,BULL J,PENDLEBURY M. Towards a multi-dimensional approach to higher education assessment[J]. Higher education,2019,78(3):497513.
- [14] 包水梅,陈嘉诚. 高校学生评教的有效性及其影响因素:基于 L 大学 2139 门课程学生评教数据的分析[J]. 现代教育管理,2022(6):5463.
- [15] KRZYSZTOF R,ELZBIETA K. Will artificial intelligence revolutionise the student evaluation of teaching? a big data study of 1.6 million student reviews[J]. Assessment & evaluation in higher education,2021,46(7):11271139.
- [16] GOTTIPATI S,SHANKARARAMAN V,LIN J R. Text analytics approach to extract course improvement suggestions from students feedback[J]. Research and practice in technology enhanced learning,2018,13(1):419.
- [17] TASHCHIAN A,KALAMAS H M,FORRESTER W R. How faculty status impacts student evaluations of teaching: a study of full-versus part-time marketing faculty[J]. Journal of marketing education,2022,44(1):113126.
- [18] STUPANS I,MCGUREN T,BABEY A M. Student evaluation of teaching:a study exploring student rating instrument free-form text comments[J]. Innovative higher education,2015,41(1):3342.
- [19] 陈玉婵,刘威. 基于情感分析的学生评教文本观点抽取与聚类[J]. 计算机应用,2020,40(S1):113117.
- [20] 刘子建,李冉,陈富强,等. 基于 SEEQ 与层次分析法的认证型评教体系研究[J]. 河南师范大学学报(自然科学版),2019,47(5):3238.
- [21] 田贤鹏,姜淑杰. 高质量发展视域下高校学生评教:问题表征与制度变革[J]. 现代教育管理,2022(9):7483.
- [22] 王光明,张楠,李健,等. 教师核心素养和能力的结构体系及发展建议[J]. 中国教育科学,2019(3):8488.
- [23] 邱文教,赵光,雷威. 基于层次分析法的高校探究式课堂教学评价指标体系构建[J]. 高等工程教育研究,2016(6):138143.
- [24] 李毅,何莎薇,邱兰欢. 教育信息化 2.0 时代下师范生信息素养评价指标体系研究[J]. 中国电化教育,2020(6):104111.
- [25] 胡德,郭刚正. 最小二乘法、矩法和最大似然法的应用比较[J]. 统计与决策,2015(9):2024.
- [26] 何彦霏. 交互式教学:高校提升教学效果的重要方法[J]. 中国高等教育,2023(19):5255.
- [27] 王辞晓,徐珺岩,郭利明,等. 多场景人机协同在线教学评价框架研究:基于层次分析法和熵权法的分析[J]. 现代教育技术,2023,33(1):7482.
- [28] 宋旭红. 高校学生评教管理功能的价值回归[J]. 中国高等教育,2020(6):5759.
- [29] 钱明霞,赵磊磊. 大数据赋能本科教学质量评价:价值意蕴、现实困境与路径选择[J]. 重庆高教研究,2023,11(5):4048.
- [30] 郭丽君,蒋贵友. 高校教学同行评议的制度化困境研究:新制度主义视角的分析[J]. 湖南师范大学教育科学学报,2019,18(3):100104.

(责任编辑:杨慷慨 校对:张海生)

Construction and Application of University Teaching Quality Evaluation Index Based on Teaching Evaluation Text Mining

LI Jianlong¹, LI Ligu²

(1. Academic Affairs, People's Public Security University of China, Beijing 100038, China;

2. Education College, Renmin University of China, Beijing 100872, China)

Abstract: Student evaluation of teaching is not only an important activity to ensure teaching quality, but also the main means for universities to evaluate teaching quality. In order to improve the effectiveness of teaching evaluation and provide more reference teaching results for instructors and educational administration departments, the research approach based on big data teaching text mining should be put on the research agenda, which is different from quantitative research mainly based on structured student rating data and qualitative research based on unstructured teaching text data. Based on the subjective and objective evaluation data of students, the Analytic Hierarchy Process (AHP) was used to construct a university teaching quality evaluation index system based on evaluation text mining. The scientific validity of the index system was empirically tested and applied through linear regression analysis. Research has found that a massive amount of student evaluation data has both academic research value and practical guidance significance. By mining and analyzing student evaluation texts, the key information in evaluation data can be effectively obtained, such as the instructors' teaching attitude, content, ability, methods, and effectiveness. A university teaching quality evaluation index system with 5 dimensions and 23 levels constructed based on the AHP can comprehensively and accurately use student evaluation data to evaluate and elaborate the quality of undergraduate teaching. The results of multiple linear regression analysis show that the indicator system can not only understand the dimensions that students focus on, predict student evaluation scores, and reveal the relationship between instructors teaching quality and various evaluation dimensions, but can also be applied to the evaluation of teaching quality for various types of instructors, providing refined teaching diagnosis for instructors and data consulting and service support for scientific teaching management decisions in schools. The practical application evaluation results of the evaluation index system show that the problems existing in different categories of courses are significantly different. Therefore, on the one hand, it is necessary to improve the efficiency of utilizing student evaluation data through various emerging technologies such as big data and artificial intelligence, deeply explore the data value and academic research value, behind the massive amount of student evaluation data, and strengthen the collaborative application of evaluation results from different perspectives such as peer evaluation, supervisory evaluation, and AI evaluation, and carry out student-centered and output oriented teaching evaluation applications; on the other hand, targeted teaching improvements are needed to address the teaching weaknesses of instructors in different categories of courses and age groups, in order to better serve the continuous improvement of teaching quality in universities and the comprehensive development of students.

Key words: student evaluation texts; evaluation index system of teaching quality in colleges and universities; teaching diagnosis; teaching improvement; student's development