

高校学生评教数据深度挖掘的实证研究

何喜军 朱相宇

(北京工业大学 北京 100124)

摘要: 利用统计分析及文本挖掘方法,研究高校学生评教数据中的关联规则,给出非结构化评教文本情感值的测算方法,挖掘优质和劣势课程的评教文本语义特征,并以某高校374门课程的评教数据为例进行实证研究,得出:第一,公共基础必修课和通识类课程评分普遍较低,学生对课程的期望水平以及认知水平是影响评教的关键要素;第二,课程性质和评教人数对评教分数和评语的情感值有重要影响,小班授课和必修类课程相比大班授课和选修课更易获得较高评价;第三,对评教文本情感值较高和较低的课程进行文本内容挖掘,得出:教师教学方法和教学内容是评教文本的关键特征,其中:教师备课认真程度、课程互动水平、课程内容逻辑性以及灵活的教学方式是影响学生评教的关键因素。文章为高校探索小班授课、提高公共基础必修课教学水平、引导学生对通识类课程的正确认知、以学生需求为导向改善教学方法并规范教学内容等改革提供决策支持。

关键词: 学生评教;数据挖掘;实证研究

中图分类号: G647

文献标志码: A

文章编号: 1003-2614(2019)10-0085-04

一、引言

自1915年美国出现第一张学生评教问卷以来,学生评教成为各国高校关注的重点。目前,虽然学生评教结果的有效性存在两种截然不同的观点,在学术界和教育界一直存在争议,但作为了解课程教学效果的重要依据,学生评教已成为教师进行反思、学校进一步监控并保障教学质量的重要参考^[1]。伴随网络化学生评教系统的广泛应用,积累了大量的数据,如何利用数据挖掘方法,揭示评教数据中蕴含的关联规则与内在规律,为评教数据的合理利用以及学校教学改革提供决策支持,具有重要的现实价值。

二、学生评教数据挖掘的研究现状

基于统计方法的评教数据挖掘,基于回归模型、因子分析、主成分分析等方法研究评教数据的成果主要侧重影响学生评教的因素^[2]以及评教数据中关联规则的挖掘。其中,影响因素包括:学生对于高分的期望^[3]、学生带有报复心理^[4]等主观因素;评教方式(在线、课堂)^[5];教师职称^[6]、年龄^[7]、性别^[8]等教师自身因素^[9];教师的教学态度和方法等^[10]教学因素等。同时,学科类别、学校类型与学业自评^[11]、课堂规模^[12]、课程类型与课程参评率^[13]等对评教偏差具有一定影响。

评教数据中关联规则挖掘的结果发现:评教分数与教师研究水平不存在显著相关性^[14]。李超锋和胡鹏^[15]采用

Apriori算法,针对某部属高校评教数据的实证研究发现:具有硕士学位的正高级年龄超过46岁的教师学生评教结果较差;具有副高级职称年龄在35岁及以下的教师的学生评教结果好,验证了职称、年龄等对学生评教的关联影响。李丽芳^[16]采用决策树C4.5算法和ID3算法并结合IF-THEN格式规则,对肇庆学院学生评教数据进行关联规则挖掘,得出“相对于体育与健康学院的男教师,女教师的评分要稍低”等7条因素间的关联规则。

学生评教文本挖掘研究,在学生评教结构化数据挖掘基础上,评教文本内容的挖掘受到学者关注,曾璇^[17]针对学生评教中主观评价文本的特点,抽取重要的主题词、情感词搭配词对,并以词云网络可视化的方式呈现,直观反映教学中的问题。桂红兵和张继美^[18]采用CUCsas倾向性分析系统,设计句子级的评价规则,通过分析学生在评教过程中的文字倾向,挖掘学生评教数据中的教师特征,识别包含观点的句子情感倾向。李效伟^[19]采用频率、信息增益、条件概率比、期望值差异等四种特征选取方法对学生评教中的留言进行分析,以预测学生留言的情感倾向性。目前,学生评教的挖掘多基于词频分析方法,在语义挖掘方面的成果还不多见。

综上,本文将综合利用相关性分析以及文本语义分析方法挖掘评教数据的关联规则,给出非结构化评教文本的情感值测算方法,并基于评教文本挖掘优质课程和劣势课程的语义特征,通过某高校374门课程评教数据的实证研究,为教育教学质量改革的政策制定提供决策支持。

收稿日期:2019-07-09

基金项目:北京工业大学教育教学研究课题重点项目“学生评教大数据深度分析方法及规避学生评教问题的对策研究”。

作者简介:何喜军,北京工业大学经管学院副研究员,研究方向:数据分析与建模;朱相宇,北京工业大学经管学院副研究员,研究方向:创新管理。

三、实证数据采集与处理

采集某高校部分学院 2017 - 2018 年两个学期的评教数据进行统计分析与深度挖掘,共包括 374 门课程(第一学期 193 门,第二学期 181 门);参与评教学生共 30500 人次(第一学期 17134 人次,第二学期 13366 人次);涉及评教文本 22402 条;课程性质包括公共基础必修课、实践环节(辅)、实践环节必修课、实践环节选修课、通识教育选修课、学科基础必修课、学科基础必修课(辅)、学科基础选修课、专业必修课(辅)、专业任选课、专业限选课、校选修课等 12 类。

表 1 不同性质课程评教分数统计结果

课程性质	开课门数	评教人数	均值	中位数	标准差
校选修课	5	44	89.41	90	2.81
专业限选课	62	3680	87.93	87.245	2.885
学科基础必修课	131	9284	87.84	85.895	2.833
实践环节必修课	43	2834	87.18	87.13	2.106
实践环节(辅)	6	1078	87.04	87.4	1.071
专业必修课(辅)	14	2164	87.03	86.815	1.377
实践环节选修课	5	265	86.96	86.11	3.546
学科基础选修课	17	1048	86.57	86.48	4.713
专业任选课	26	1590	86.49	86.585	2.426
学科基础必修课(辅)	16	2500	86.24	87.57	1.374
通识教育选修课	44	5527	85.53	85.54	2.222
公共基础必修课	5	486	82.82	83.27	8.366
总计	374	30500	86.75(平均)	-	-

通过表 1 发现:第一,大部分课程的标准差较小,在 1 - 3 分之间,差别不大,标准差较大的课程类型包括公共基础必修课、学科基础选修课、实践环节选修课,说明学生对这三类课程的评分存在明显差异性;第二,均值排名后两位的课程类型为公共基础必修课(评分方差最大)、通识教育选修课,其中:公共基础必修课中包含英语、高数、概率、线性代数等基础类课程,此类课程对学生的专业课程学习以及进修学习具有重要的支撑作用,学生期望水平较高,同时,此类课程通过的难度相对较大,会出现一定程度的挂科现象,因此评分较低既说明学生期望对评分具有较大影响,也说明此类课程的评分两极分化比较严重,因材施教的教学方式还有待进一步探索;通识教育选修课是通识教育课程体系中的重要组成部分,其目标是引导学生广泛涉猎不同学科领域知识,引导学生获得多种不同分析方法,注重能力、方法、性情和素质的培养,加强人文素养与科学素养的交融,增强学生对自身、社会、自然及其相互关系的理解,评分较低的原因包括:一是学生对课程地位的认知水平不高,大部分学生选择通识课程的动机为满足学分要求,认为课程对专业知识的提升作用不大,学习动力不足;二是通识课程的授课内容难易程度和教学方式等也影响学生评分。

五、学生评教分数关联规则挖掘

学生人数和评教分数的相关性,采用 SPSS 中 Pearson 相关系数研究 374 门课程评教学生人数和评教分数之间的相关性为 -0.138,说明二者之间呈现负相关,即:某一课程授

四、结构化评分统计分析

评教分数分布 将评教分数划分为五个区间并统计分布情况,得出:90 分及以上分数的课程占比 18%,85 - 90 分的课程占比 65%,80 - 85 分的课程占比 15%,75 - 80 分的课程占比 1%,75 分以下课程占比 1%,说明参评学生对开设课程的评分整体较高,评分呈现正态分布趋势。

不同课程性质评教分数统计分析,利用 SPSS 软件对两个学期不同性质的课程评分进行均值和标准差分析,统计结果如表 1 所示。

课人数越多(等于评教人数),评分低的可能性越大,为教学改革中小班授课提供决策支持。

课程性质与评分的相关性,将 12 类课程性质分为必修课和选修课(必修课 215 门,选修课 159 门),采用 SPSS 中的 Kendall 相关系数分析,得出课程类别与评教分数之间的相关性为 0.1,且必修课的评分高于选修课,说明课程性质与评分相关。

六、非结构化评教文本情感挖掘

非结构化评教文本的情感挖掘是文本量化分析、语义分析以及评教分数相关性研究的关键。采用能够综合考虑多种信息的情感分析方法 SnowNLP 进行学生评教情感值量化,考虑否定词的影响将评教文本进行积极情绪和消极情绪的分类。具体实现过程使用 Python 工具,步骤如下:

第 1 步:读取 22402 条课程评教文本信息并进行分句。

第 2 步:分词、去除停用词并使用 SnowNLP 自带的情感词库查找分句中的情感词,标记极性:积极为 1,消极为 -1。

第 3 步:利用 Python 自带情感分析语料库训练模型,并基于贝叶斯模型进行情感分类,对于两个情感正负类别 c_1 和 c_2 的分类问题,其特征为 w_1, \dots, w_n ,特征之间是相互独立的,属于类别 c_1 的贝叶斯模型的基本过程为:

$$P(c_1 | w_1, \dots, w_n) = \frac{P(w_1, \dots, w_n | c_1) \cdot P(c_1)}{P(w_1, \dots, w_n)} \quad (1)$$

其中 $P(w_1, \dots, w_n) = P(w_1, \dots, w_n | c_1) \cdot P(c_1) + P(w_1, \dots, w_n | c_2) \cdot P(c_2)$,在模型训练过程中,统计正负样本

类的词频,并返回最终的正负类之和作为该句的情感评分。

第4步:针对 c_1 类别的情感概率采用贝叶斯模型进行预测,公式如下:

$$P(c_1 | w_1, \dots, w_n) = \frac{1}{1 + \exp[\log(P(w_1, \dots, w_n | c_2) \cdot P(c_2)) - \log(P(w_1, \dots, w_n | c_1) \cdot P(c_1))]} \quad (2)$$

第5步:当 $P(c_1 | w_1, \dots, w_n)$ 接近1时为正向情感,当接近0时为消极情感,计算完一条评论所有分句的情感值,用数组 list 记录。

第6步:计算并记录所有评论的情感值。

第7步:合并每条评论信息中分句的情感值,计算每条评论的情感均值。

最后,计算一门课程的多条评论的情感值均值,支撑后续的相关性分析。

情感值与评教分数的相关性,采用 Pearson 相关系数对二者的相关性进行计算,得出相关性为 0.446,且显著性很高,说明情感值与结构化评教分数具有较强的相关性。

情感值与课程类别的相关性,采用 Kendall 相关系数对情感值与课程类别(必修和选修)的相关性进行计算,得出相关性为 0.13,且必修课的情感值稍高于选修课,与结构化数据分析结果具有一致性。

情感值与评价文本数量相关性,采用 Pearson 相关系数计算课程情感值和评教文本数量之间的相关性,得出相关性为 -0.163,说明参与评教学生的评教文本数越高,评分有可能越低,进一步验证了小班授课能提高评教效果的结论。

情感值较高和较低的课程特征挖掘,对情感值排名前 40 和后 40 的课程评教文本进行内容挖掘,并统计文本语义特征及词频分布如表 2 和表 3 所示。

表 2 情感值排名前 40 课程评教文本情感特征频次统计

评价项目	文本特征	频次
教师授课方法	讲课清晰流畅	8
	课堂氛围幽默生动	19
	备课认真	36
	对学生负责	21
教学内容	逻辑清晰	28
	重点突出	5
	具有创新性	5
	教学方式灵活	7
	内容生动丰富	24

从表 2 得出:情感值较高的课程中教师授课方法和教学内容是学生关注的重点,其中:备课认真、课程氛围幽默生动、对学生负责、逻辑清晰、内容生动丰富、教学方式灵活等是优秀课程的共性特征。

表 3 情感值排名后 40 课程评教文本情感特征频次统计

评价项目	文本特征	频次
教师授课方法	语速音量不当	10
	普通话不标准	5
	课堂枯燥缺乏互动	20
	备课不翔实	3

教学内容	逻辑性差	8
	重点不突出	8
	教学节奏不适	4
	课程难度大	5
	内容单一	10
	作业量大	5

从表 3 得出:课堂枯燥缺乏互动、语速音量不当、内容单一、逻辑性差、重点不突出等是学生评教情感值低的课程的共性特征。

七、研究结论及教学改进的思考

1. 研究结论

本文综合利用统计分析和文本挖掘方法对某高校 374 门课程的评分及评教文本进行深度挖掘,得出:

第一,公共基础必修课和通识类课程的评分较低,由于两类课程授课对象广泛,基础性较强,且承担着基础素质和人文素质的培养职能,因此,学生对课程的期望水平以及是否具有正确的认知等都将影响评教结果。

第二,分别对结构化评分以及非结构化评教文本的情感值与课程人数和课程性质进行相关性分析,得出一致性结论,即:小班授课能提高授课效果,学生评教分数较高,相比选修课,必修类课程的评分稍高。

第三,对情感值较高和较低的课程进行评教文本内容挖掘,从教师教学方法以及教学内容两个层面研究两类课程文本特征,得出:教师备课认真程度、课堂互动水平、课程内容逻辑性以及灵活的教学方式等是学生关注的重点,并直接影响学生评教结果。

2. 教学改进的思考

第一,倡导小班授课,提高学生与授课教师的互动。要兼顾教学效率和教学公平,对于重要的公共基础必修课、学科基础必修课、实践环节必修课等课程实施小班授课,探索以课堂讨论、案例分析、问题解决等创新型和研究型为主的教学方法的改革。一方面,提高学生在课堂教学中的参与度和积极性;另一方面,提高学生的自学能力和创新思考,通过因材施教提高学生满意度。

第二,完善通识类课程体系,引导学生选课的正确认知。改善通识类课程独立设置的流程体系,探索以面带点的模块化课程体系建设,例如,进行文学与艺术、自然科学、社会分析、经济管理、创新创业等不同模块的课程建设,切实引导学生基于兴趣、以需求为导向的选课行为和对课程的正确认知,避免盲目选课、为满足学分选课带来的学习效率低下、学习效果不佳的后果。

第三,改善学生评教体系,深入挖掘非结构化评教信息。基于实证研究发现学生评教量表相对简单,一方面,缺乏学生对课程的投入和认知信息,导致评教结果的可信度受到影响;另一方面,评教系统难以分析汇总评教文本的共性信息,导致基于评教的教学改革缺乏针对性依据。因此,在评教量

表中增加学生学习投入情况、选课动机与倾向、学习感受、建议等,并开发评教文本深度挖掘与分析功能,对于有效地利用评教数据、指导教学改革具有重要意义。

参考文献:

- [1]余海波.基于学生评教的高校教师教学能力提升[J].国家教育行政学院学报,2017(6):77-81.
- [2]戴璨,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(5):72-80.
- [3]ISELY P, SINGH H. Do Higher Grades Lead to Favorable Student Evaluations? [J]. Journal of Economic Education, 2005(1):29-42.
- [4]MAURER T W. Cognitive Dissonance or Revenge? Student Grades and Course Evaluations [J]. Teaching of Psychology, 1983(3):176-179.
- [5]NOWELL C, Gale L R, KERKVLIT J. Non-response bias in student evaluations of teaching [J]. International Review of Economics Education, 2014(17):30-38.
- [6]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(6):83-89.
- [7]FELTON J, et al. Attractiveness, Easiness and Other Issues: Student Evaluations of Professors on Rate-my-professors.com [J]. Assessment & Evaluation in Higher Education, 2008(1):45-61.
- [8]MACNELL L, DRISCOLL A, HUNT A N. What's in a Name: Exposing Gender Bias in Student Ratings of Teaching [J]. Innovative Higher Education, 2015(4):291-303.
- [9]董桂才.学生评教的非教师因素研究[J].高教探索,2014(2):104-106.
- [10]李香林.基于因子分析与层次聚类的学生评教数据挖掘[J].吕梁学院学报,2014(2):1-4.
- [11]周继良,龚放,秦雍.高校学生评教行为偏差及其与学科类别、学校类型和学业自评的关系——基于南京和常州十所高校的实证调查[J].高等教育研究,2017(10):64-74.
- [12]陈晓琳,刘微.学生群体属性对高校学生评教的影响及对策探析[J].黑龙江教育:高教研究与评估,2008(1):181-183.
- [13]龚辉锋.课程因素影响学生评教的实证研究[J].黑龙江高教研究,2017(10):1-6.
- [14]PALALI A, ELK R V, BOLHAAR J, et al. Are good researchers also good teachers? The relationship between research quality and teaching quality [M]. Cpb Discussion Paper, 2017.
- [15]李超锋,胡鹏.基于关联规则挖掘的学生评教结果分析[J].海峡科技与产业,2017(6):206-208.
- [16]李丽芳.学生评教数据知识挖掘方法应用研究[D].广州:华南理工大学硕士学位论文,2015.
- [17]曾璇.基于句法模式的评教信息挖掘[J].电脑编程技巧与维护,2016(16):57-58.
- [18]桂红兵,张继美.句子级倾向性分析在学生评教系统设计中的应用[J].宜春学院学报,2017(6):119-122.
- [19]李效伟.学生评教留言的倾向性分析[J].微型机与应用,2016(3):24-27.

An Empirical Study on Deep Mining of College Students' Teaching Evaluation Data

HE Xi-jun, ZHU Xiang-yu

(Beijing University of Technology, Beijing 100124, China)

Abstract: By using statistical analysis and text mining methods, this paper studies the association rules in college students' teaching evaluation data, gives the method of measuring the emotional value of unstructured teaching evaluation texts, excavates the semantic characteristics of teaching evaluation texts of high-quality and inferior courses and takes the teaching evaluation data of 374 courses in a university as an example to carry out empirical research. The study gives results as follows: Firstly, the scores of compulsory courses and general courses are generally low. Students' expectation level and cognitive level on courses are the key factors affecting the evaluation of teaching; Secondly, the nature of curriculum and the number of teachers have an important impact on teaching evaluation scores and emotional value of comments, smaller and compulsory classes are easier to obtain higher evaluation than large and elective classes; Thirdly, textual content of courses with higher and lower emotional value of evaluation texts is excavated. It is concluded that teachers' teaching methods and contents are the key features of teaching evaluation texts. The key factors affecting students' teaching evaluation are teachers' earnestness in preparing lessons, the level of curriculum interaction, the logic of curriculum content and flexible teaching methods. This study provides decision support for exploring small-class teaching, improving the teaching level of public basic compulsory courses, guiding students to understand general courses correctly and improving teaching methods and standardizing teaching content based on students' needs.

Key words: student evaluation; data mining; empirical research