Advances in Applied Mathematics 应用数学进展, 2022, 11(12), 8598-8604
Published Online December 2022 in Hans. http://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2022.1112906

Hans汉斯

基于文本挖掘的教学质量评价指标量化研究

赵宇昂*，周　胜，张泽文

南京信息工程大学，江苏　南京

收稿日期：2022年11月12日；录用日期：2022年12月6日；发布日期：2022年12月14日

摘　要

教学质量评价数据主要由专家给出的评判等第和综合评语构成，现存的教学质量评估方法只能根据专家给出的等第对教学质量进行粗略划分，缺少对于专家综合评语的具体量化。本文旨在通过文本挖掘方法实现对于专家评语的具体量化，然后结合专家给出的等第进行监督加权，得到一个结合专家等地和专家评分的综合量化得分，实现对于教学质量的具体评价。最后使用提出的量化指标对该校近三年的教学质量进行稀疏主成分分析，得到的三个主成分具有良好的解释性，可以被解释为“学期效应”、“理论课效应”和“实践课效应”，揭示出新冠疫情爆发导致高校教学质量急剧降低的规律。

关键词

文本挖掘，监督加权，教学质量评价，TF-IDF准则，稀疏主成分分析

Quantitative Research on Teaching Quality Evaluation Index Based on Text Mining

Yu'ang Zhao*, Sheng Zhou, Zewen Zhang

Nanjing University of Information Science and Technology, Nanjing Jiangsu

Received: Nov. 12th, 2022; accepted: Dec. 6th, 2022; published: Dec. 14th, 2022

Abstract

The teaching quality evaluation data are mainly composed of the evaluation grade and comprehensive comments given by experts. The existing teaching quality evaluation methods can only roughly divide the teaching quality according to the grade given by experts, and lack of specific quantification of experts' comprehensive comments. This paper aims to achieve the specific quantification of expert comments through text mining method, and then combine the grade given by experts to supervise and weight to obtain a comprehensive quantitative score combining expert

*通讯作者。

文章引用: 赵宇昂, 周胜, 张泽文. 基于文本挖掘的教学质量评价指标量化研究[J]. 应用数学进展, 2022, 11(12): 8598-8604. DOI: 10.12677/aam.2022.1112906

grade, so as to achieve the specific evaluation of teaching quality. Finally, the proposed quantitative indicators are used to conduct sparse principal component analysis on the teaching quality of the school in the past three years, and the three principal components obtained have good explanatory power, which can be interpreted as "semester effect", "theory course effect" and "practice course effect", revealing the law that the outbreak of the COVID-19 has led to a sharp decline in the teaching quality of colleges and universities.

Keywords

Text Mining, Supervisory Weighting, Teaching Quality Evaluation, TF-IDF, Sparse Principal Component Analysis

1. 引言

近年来，国家不断推进“双一流”建设，意图打造中国特色与世界一流有机融合的大学与学科。在这样的背景下，高校教学质量成为越来越受关注的话题，而对于教学质量的监控与评价也随之引起了许多学者的注意，一套行之有效的教学质量水平评价方法是目前所亟需的。

教学质量评价数据主要是由专家根据课堂表现给出的评价等第与描述课堂情况的综合评语两部分组成。现有的课堂教学质量评估方法[2]大多是根据专家给出的评价等第进行建模分析，比如丁等人[1]使用层次分析法或模糊综合评判等方法给出一个描述课堂教学质量的粗略得分，或是将专家等第视作教学质量评分直接进行统计分析[3]。

现有的教学质量评价方法显然过于依赖专家等第，这样的方法对不同教学质量的课堂评价划分比较粗略，最多将课堂表现分为几类，因此亟需一种对于教学质量量化的评价方法；并且现有方法没有在评价教学质量时考虑专家给出的综合评语，这样无疑会损失大量专家对于课堂教学质量的评价信息，因此本文旨在通过文本挖掘方法挖掘出隐藏在专家评语中的评价信息，并结合专家等第实现对于教学质量的监督加权量化。本文剩下的文章结构如下所示，第二节对本文所使用的教学质量评价数据进行简要介绍，第三节将对本文使用的监督加权量化方法和文本挖掘方法进行简要介绍，第四节是基于量化得分的实例分析，使用稀疏主成分分析对某高校所有教师近三年教学质量变化进行分析，第五节则是针对结果的讨论与展望。

2. 数据描述

2.1. 教学质量评价数据

教学质量评价数据共有 5821 条记录，每条记录代表着一次专家的听课记录，其包含的变量有听课专家、听课时间、授课教师、所听课程、教师教学质量评价等第、学生学习评价等第以及教师教学质量综合评语。评价等地可分为优秀、良好与一般，综合评语则是专家根据课堂教学评价重点内容和参考标准给出的标准化文本评价，一个良好的综合评语的行文内容与格式应该具有规范性，即与参考标准紧密相关。

2.2. 课堂教学评价重点内容和参考标准

课堂教学评价重点内容和参考标准是给予专家的一个打分标准，用以帮助专家撰写详细规范化的综

合评语，其内容可以细分为理论课与实践课两套打分标准，以理论课为例，评价内容被分为教学态度、教学内容、方法手段和教学成效四个评价方面，每个方面具有不同的打分权重和描述内容，具体内容如表 1 所示，一个良好的综合评语的行文内容与格式应该是与参考标准紧密相关的，也是本文可以对专家评语进行文本挖掘的基础。

Table 1. Key content and reference standard of classroom teaching evaluation (theory course)
表 1. 课堂教学评价重点内容和参考标准(理论课)

评价方面	评价内容与标准	评分权重
教学态度	精神饱满，教姿教态端正，严守纪律 课前准备充分，按时上下课	0.2
教学内容	讲授熟练，信息量丰富，重难点突出，逻辑性强 结合课程特点，合理补充学科前沿，有机结合应用	0.3
方法手段	教学方法灵活，鼓励学生质疑，注重启发思维，熟练驾驭课堂 恰当运用信息化教学手段，课件、板书清晰工整，课堂组织方式灵活	0.3
教学成效	学生注意力集中、思维活跃，有效促进知识内化 教学互动积极有效，课堂气氛活跃和谐，给予学生创新启迪	0.2

3. 模型介绍

3.1. 基于加权距离的文本信息量化模型

以理论课为例，本文分别对教学态度(权重 0.2)、教学内容(权重 0.3)、方法手段(权重 0.3)和教学成效(权重 0.2)四项的具体评价内容和标准依次进行等权重划分，将“精神饱满，教姿教态端正，严守纪律”记为 S1，赋权重 0.1；将“课前准备充分，按时上下课”记为 S2，赋权重 0.1；将“讲授熟练，信息量丰富，重难点突出，逻辑性强”记为 S3，赋权重 0.15；将“结合课程特点，合理补充学科前沿，有机结合应用”记为 S4，赋权重 0.15；将“教学方法灵活，鼓励学生质疑，注重启发思维，熟练驾驭课堂”记为 S5，赋权重 0.15；将“恰当运用信息化教学手段，课件、板书清晰工整，课堂组织方式灵活”记为 S6，赋权重 0.15；将“学生注意力集中、思维活跃，有效促进知识内化”记为 S7，赋权重 0.1；将“教学互动积极有效，课堂气氛活跃和谐，给予学生创新启迪”记为 S7，赋权重 0.1。

实践(实验)课的量化得分计算流程与理论课的思路一致，唯一区别在于评价重点内容和参考标准的文本信息不同，仅需调整语料库、修改权重即可。

本文首先对 S1~S8 这 8 条中文文本进行分词，根据上文筛选出来的特征关键词依次与 8 句文本做特征提取。以 S1 为例，从“精神饱满，教姿教态端正，严守纪律”中可以提取 $n_1 = 3$ 个相似关键词，分别为“精神饱满”，“端正”，“严守”，记为

$$Word_{11}, Word_{12}, Word_{13}。$$

类似地，我们对第 i 句文本进行特征提取，可提取出 n_i 个相似关键词，这样就可以得到 $N = \sum_{i=1}^{8} n_i$ 个特征相似关键词，即

$$Word_{i1}, Word_{i2}, \cdots, Word_{in_i}, i = 1, \cdots, 8。$$

接下来，就是对专家评语进行量化的第一步：向量赋值。搜索专家评语中是否有与 $Word_{i1}$ 意思高度

相近词语，若有，该特征向量的第一个分量赋值为1；若没有，第一个分量赋值为0；若出现带有消极情感色彩的负相关词语，第一个分量赋值为−1。

以此类推，我们可以得到某个待检测专家评语的8个特征向量$w_{s1}, w_{s2}, ..., w_{s8}$，此时，本文定义最顶尖教师的8个特征向量的分量全为1，即$ws_{\max i} = (1,1,\cdots,1)_{1*n_i}$。然后，计算某个待检测专家评语与最顶尖教师相对应的特征向量的平方距离，定义如下

$$d_{ws_i} = \left\| w_{si} - ws_{\max i} \right\|_2^2 = (s_{i1}-1)^2 + (s_{i2}-1)^2 + \cdots + (s_{in_i}-1)^2$$

根据上文划分的权重，可以计算出加权距离

$$d_{ws} = \sum_{i=1}^{8} d_{ws_i} \text{。}$$

与参考标准距离越小的教师授课水平越高，距离越大的教师授课水平越低。最后，将这组距离值取负号并映射到60~100的区间上，就可以得到一组与教师授课水平正相关的量化值，记为$Score_{PY}$。

3.2. 监督加权量化得分模型

上一节建立的基于加权距离的文本信息量化模型的确可以有效地将教师综合评语进行量化计算得到$Score_{PY}$，并且可以直接以该得分作为教师授课水平的量化指标进行排序。但是，$Score_{PY}$未考虑教师教学综合评价以及学生学习综合评价这两项定性指标的监督信息，这两项指标是对教师评语的不同角度概括，根据《课堂教学评价重点和参考标准》可知，教师教学综合评价权重占0.8，学生学习综合评价权重占0.2。

因此，本文对两项指标进行定量化，将教师教学综合评价的优秀、良好、一般分别赋值100、80、60 (合格与不合格的样本太少，已经剔除)，记为TE_{score}；将学生学习综合评价的优秀、良好、一般分别赋值100、80、60，记为STU_{score}；再结合上一节计算出的专家评语量化得分$Score_{PY}$，给出监督加权教师评语的量化得分公式如下：

$$Score = \theta \cdot Score_{PY} + (1-\theta) \cdot (0.8 \cdot TE_{score} + 0.2STU_{score})$$

其中，θ为设定的变加权系数，在使用时可以根据五折交叉验证对其进行选择。

3.3. TF-IDF 算法

TF-IDF算法是一种文本挖掘的常用加权技术，其基本思想是：如果某个单词在一篇文章中词频(TF)高，并且在其他文章中很少出现，则认为此词或者短语具有很好的类别区分能力，适合用来分类。IDF是逆向文件频率，某一特定词语的IDF，可以由总文件数目除以包含该词语的文件的数目，再将得到的商取对数得到。某一特定文件内的高词语频率，以及该词语在整个文件集合中的低文件频率，可以产生出高权重的TF-IDF。因此，TF-IDF倾向于过滤掉常见的词语，保留重要的词语。

4. 实例分析

4.1. 文本信息预处理

为实现综合评语的量化，本文根据课堂教学评价重点内容和参考标准，利用R软件分别建立理论课和实践(实验)课的小型语料库，首先进行分词，然后进行特征清洗–去除停留词(标点符号、虚词等)和去除词频过低的词，最后根据TF-IDF准则[4]，筛选出有效且高频的特征关键词，并查找其相似关键词。其中，“有效”主要体现在：一些词虽然高频，但它们并不是利于区分的特征词(文本信息特征提取流程如图1所示)。与此同时，与这些筛选出来的特征关键词负相关的词语也被查找出来，以便后续计算文本相似度。

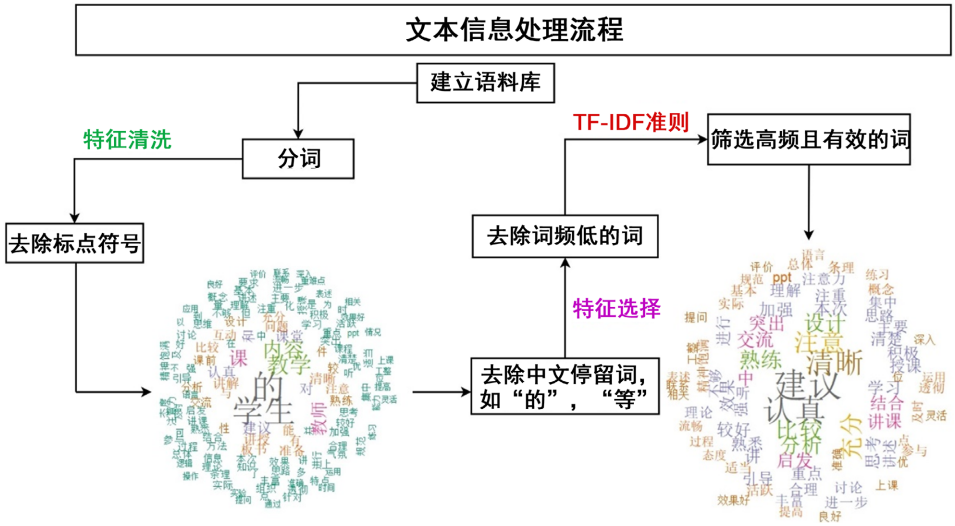

Figure 1. Flow chart of text information feature extraction
图 1. 文本信息特征提取流程图

4.2. 基于稀疏主成分的教学质量变化分析

在文本预处理后使用本文提出的监督加权量化得分模型，可以得到基于综合评语与评价等第的量化得分作为评判教学质量水平的标准，本节将在该量化得分的基础上对近年该校整体教学质量变化进行分。图 2 描绘了十名教师随时间量化得分的变化轨迹图，可以看出时间维度上对于教师的课堂得分记录较为稀疏，一些孤立的数据点则代表该名教师的记录只有一条，针对这种采集点稀疏的数据场景，本节选择稀疏主成分分析的方法对进行建模，稀疏主成分分析[5] [6] [7]是传统主成分分析(PCA)的改进方法，它可以很好的处理具有稀疏结构的数据。

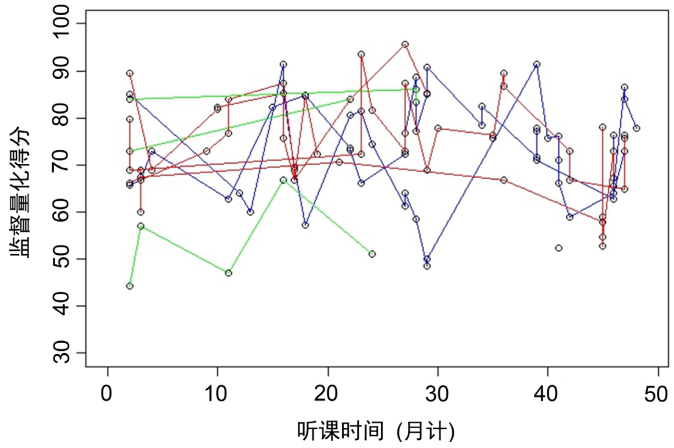

Figure 2. Track chart of ten teachers' quantitative scores
图 2. 十名教师的量化得分轨迹图

对教师得分进行稀疏 PCA，选取了三个稀疏主成分，其方差累计贡献率达到 85.63%，得到的主成分如图 3 所示，第一主成分随着时间向零均值线下周期波动，波动幅度也较为均匀，开始向下波动的时间往往是某一年的 9 月份或某一年的 2、3 月份，均为新学期开始的时间段，类似于时间序列分析中的季节效应，可以将这种教学质量随学期开学时间点波动的情况称之为“学期效应”。具体来说就是教师的教

学质量在学期开始时很好，但随着学期的推进而慢慢下降，经过一学期的推移，教学质量又会因为假期的接近而逐渐上升；因此本文将第一主成分称之为“学期效应主成分”。

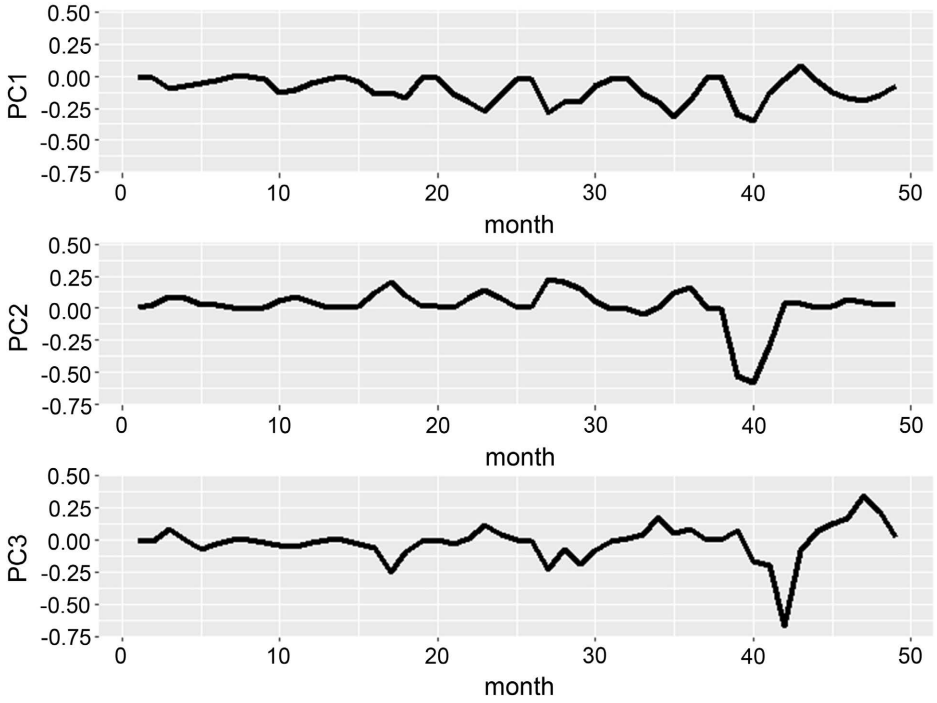

Figure 3. Sparse principal component diagram of quantitative score
图 3. 量化得分的稀疏主成分图

第二主成分第 36 个月前，即 2017 年 1 月到 2019 年 12 月期间，随着时间沿零均值线向上波动，但是波动幅度相比第一主成分较小，波峰却较第一主成分要尖，而从第 38 月开始，即 2020 年 2 月开始，第二主成分出现了一个和原先的周期性不符的向下波动，并且这个波峰的强度很大，说明在这一时刻开始教学质量出现了严重下滑，而结合时间可以知道是新冠疫情发生的时刻。因为新冠疫情的出现，学校的教学模式转变为线上教学，从而导致教师教学质量的下降，受疫情影响显著的是理论课，而第二主成分第 36 个月前的周期性也与理论课随着学习的深入教学质量会提高的特点相符合，因此本文将第二主成分称为“理论课效应主成分”。

与第一第二主成分不同，第三主成分没有明显的随学期变化的周期性趋势，在零均值线上下波动，这也与实践课的时间在一年中不固定的特点相统一，即实践课时间往往不受明确的学期时间限制，很多实践课都会发生在假期之中，因此本文将第三主成分称为“实践课效应主成分”。在疫情发生时期，第三主成分虽然受到影响，但没有第二主成分受到影响大也可以很好解释，因为实践课随着返校时间的延期大都推迟到下个假期，所以没有立刻受到线上教学形式的影响，第三主成分在七八月份的明显回升也可以很好的佐证上述观点。

综合上述三个主成分，可以看出该学校的教学质量在 2017 年到 2019 年年末间保持着稳定的趋势，而随着 2020 年新冠疫情的爆发，由于教学方式转变为线上，教学质量出现了断崖式的下滑，其中理论课教学质量受到的冲击最为严重。

5. 总结

基于文本挖掘对综合评语进行量化的方法在高校教学质量水平评价中有更为广泛的应用背景，可以

帮助教务部门实现对于教师教学水平更为有效的监控。本文的工作贡献主要有以下 3 点：

1) 借助文本挖掘方法实现了对专家评语这一文本数据的量化，为评判教学质量水平提供了更多维度上的标准。

2) 将我们提出的专家评语量化数据与专家等地通过监督加权的思想相结合，使得得到的监督加权量化得分具有更好的解释性和可信度。

3) 根据提出的监督加权量化得分对该校教师近五年的教学质量水平进行了稀疏主成分分析，得到的主成分具有良好的解释性和现实意义，成功捕捉到了新冠疫情对于高校教学质量水平的冲击与影响，也侧面印证了本文提出的监督加权量化得分的可靠性。

参考文献

[1] 丁家玲, 叶金华. 层次分析法和模糊综合评判在教师课堂教学质量评价中的应用[J]. 武汉大学学报(社会科学版), 2003(2): 241-245.

[2] 申卫星. 高等学校教学质量评价指标体系研究[D]: [硕士学位论文]. 上海: 东华大学, 2004.

[3] 吴培群, 陈小红. 大学生评教的统计分析及其改革途径探讨——基于北京一所高校学生评教分数的统计分析[J]. 高教探索, 2010(3): 78-81, 91.

[4] 张伟, 石倩, 何霄, 王晨, 李禾香, 李骥然. 改进的 TF-IDF 算法在文本分类中的研究[J]. 信息技术与网络安全, 2021, 40(7): 72-76+83.

[5] Shapiro, S.S. and Wilk, M.B. (1965) An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, **52**, 591-611. https://doi.org/10.1093/biomet/52.3-4.591

[6] Johnstone, I.M. and Lu, A.Y. (2009) On Consistency and Sparsity for Principal Components Analysis in High Dimensions. *Journal of the American Statistical Association*, **104**, 682-693. https://doi.org/10.1198/jasa.2009.0121

[7] Sigg, C.D. and Buhmann, J.M. (2008) Expectation-Maximization for Sparse and Non-Negative PCA. *Proceedings of the* 25*th International Conference on Machine Learning* (*ICML*'08), Association for Computing Machinery, New York, 960-967.