

基于多尺度聚类分析的学生评教大数据深度挖掘与趋势分析

杨安康

摘要: 学生评教已成为高校教学质量保障的重要内容,但数据在高校教学决策与管理中的可应用性仍被广泛质疑。评教的可信度不仅取决于原始数据,更取决于数据分析技术。通过引入人工智能、模式识别领域中常用的聚类分析算法,基于多年的教学管理工作实践,本文提出应用多尺度聚类分析方法,能够从低信噪比的学生评教历史数据中挖掘出高可信度信息,为高校科学决策与高效管理提供坚实的数据支撑。

关键词: 学生评教; 大数据挖掘; 聚类分析

一、高校学生评教现状与问题分析

学生评教是最基础的教学评估手段^[1],但运行多年后往往成为高校教学评估中为人诟病的部分。其原因主要包括三方面。一是学生评教数据严重失真,“分数虚高”,课程之间的差异度逐步变小^[2]。根据多年监测,学生评教的原始数据均分每年以0.2~0.3分的速度提升,部分院系全体教师的课程均分已经超过90分。这样的学生评教从分值上看已然失去了“可信力”,也会造成老师对于自己教学水平的误判。有些老师已被督导评价为“不合格”,仍会以历年学生评教80分以上的数据作为依据说明自己在学生中的反馈是良好的。二是学生评教模型复杂、低效。对于学生评教数据往往从均值、最高值、方差等方面进行统计,而这些统计方式最大的问题就是缺失了“典型性”。许多高校为了解决学生数据失真问题,建立更多的计算统计算法^[3]。但复杂的算法往往并不能得到更具可信度的结果,无法回答“是否高得分的院系,其教学水平就一定比低分的院系要好”

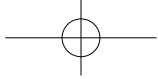
的问题。评教模型成为一种数字游戏,缺失学术与行政效能。三是学生评教重评分轻评语。目前大部分学校的学生评教分析指标只有评分指标,量化评价基本来自于学生的评教客观题评分。虽然是客观题评价,但从调研角度上看,评分基准存在巨大的个体差异。我们从历年教学检查与师生交流中发现,不论是老师还是学生,对于评语都缺乏足够的重视。学生评语的数量,学生评语的句长等指标逐年下降。其实,用自然语言表达对老师的教学反馈,是最直观也最有效的途径;而缺少评语分析模型,无法将学生评语量化并形成可视化的数据表格,是制约学生评语推广的最重要因素。

上述原因致使从教学管理层到一线教师都认为学生评教不具有“可信度”,对于学生评教工作存在诸多批评,而学生评教工作也无法发挥其在教学评估中的基础性作用^[4]。

二、高校学生评教工作的定位反思

虽然学生评教存在许多问题,但是否能够因

杨安康,东南大学教师教学发展中心研究员。



此取消或者削弱学生评教工作，以及如何界定学生在教学评估中的作用，需要我们进一步思考。

1. 学生评教是教学评估必备的一部分

学生评教在高校已经成为常规工作，工作量相当大。以东南大学本科评教为例，覆盖 1.6 万名本科生，涉及每学期 3000 多门课程。从教学管理上看，学生评教已经成为“唯一”的广样本全覆盖的评教手段，也是重要的工作环节。近年来，“以学生为中心”的教学模式得到广泛推广。重视学生对于教师教学的反馈，也是该教学模式的必备。

2. 学生不应该成为教学评估中的主体

学生是来学习的，对于老师教学能力的评价不可能做到全面。真正能够评价老师教学能力的应该是教学专家。如果完全依靠学生评教数据，必然会产生许多问题，最明显的是教师与学生互相“放水”^[5]。这与我们教学评估的初衷背道而驰，因此教学评估不应该以学生为主体。

3. 教学督导需要科学定位

一所大学教学评估的水平是由其教学督导与教学管理团队所决定的。学校、院系、教研组的专家是教学水平评估中最有发言权的。但教学专家也是大学教师，需要承担大量教学、科研与管理工作，完全依靠督导专家全面督导检查教学工作的方方面面，不论从学术角度还是从行政角度上，都是无法实现的。“好钢用在刀刃上”，教学专家的督导工作必须科学定位，让督导专家调查、分析与解决“典型性问题”，才能切实、有效、快速地提升教学水平。

基于上述分析，应利用技术手段，将学生评教数据与其他数据关联，形成一个数据网络，将评分与评语进行量化，从不同关联角度进行聚类分析，从不同尺度上定位所期望的典型。回到我们的问题：面上学生评教数据是否可信？从教学管理角度，将学生评教作为一个普查工具，从学生角度将老师分类，抓住典型问题，就能找到合适的方式进行分析与应对。同时，

专家专注在这些问题上，真正发挥督导专家的功能，才能真正做到将教学督导“好钢用在刀刃上”。

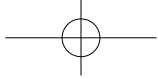
三、基于聚类分析的高校学生评教模型构建

通过对学生数据进行深度分析，挖掘到具有典型性的问题，再由专家逐个细致分析，这种混合评估模式中最主要的工作就是通过学生评教数据将教师进行科学分类，形成教师教学群数据。随着人工智能的发展，目前已经有大量的技术工具可以用来分析数据，解决数据集合的聚类问题^[6]。

1. 学生评分数据的聚类分析

(1) 学生评教数据的关联与网络化。以往学生评教数据之间是割裂、孤立的。没有关联的数据，再多也是“贫矿”。只有改变思维模式，在数据之间建立联系，形成高维度的数据网络，才能真正发挥数据的威力，也才能形成适合分析的数据资源。基于这种大数据思维，结合学校现有的信息化平台，将人事信息库、教务信息库、学工信息库与评教数据库进行关联，可以形成多维度的数据网络。

(2) 利用聚类分析算法进行分析。选择合适的算法对于多维度关联的学生评教原始数据集合进行计算，可以得到由计算机自动分类的子集合，这个过程就是“聚类分析”。本文选用经典的 K-means 算法，技术比较成熟，资源丰富；历史数据越多，可以分析的内容也越多。这种方法是多尺度的，可以在不同的层面上进行使用。在我们的研究中，曾以年龄为例，将课程评分与主讲教师进行关联，再用聚类分析的方法，可以得到“评分与年龄聚类分析图”（如图 1 所示）。图中横坐标是教师的年龄，按照天数换算成小数，纵坐标是教师本学期授课的学生评教均值，图的每一个点都代表一个老师。通过 K-means 聚类分析算法，就可以将这张图中的老师划分为多个类，并量化老师与类核心距离，从而得到该群中典型性老师与非典型老



师的集合。该数据可以直接应用到教学随堂听课工作中，能快速区分典型与非典型。

(3) 聚类后的局部曲线校正。聚类方法不仅仅可以快速定位典型性问题，还可用于解决学生评教中的“数据失真”问题。首先要说明的是，不论怎样改进调查方法，学生评教数据总存在偏差，且由于是一个跨学科大样本的采集，其原始数据是多个因素强耦合的，也就是说多种因素对于学生的评教存在复杂的影响，这些影响之间也有关联。比如学生的学科差异、同学科学术的学习习惯差异、老师的授课风格甚至老师的颜值，都会对学生评教有影响。这些影响无法用一个统一的方法进行校正，用标准分强行拉开过于密集的评教数据，或者用排序方法来解决学生评教分数虚高问题。而如果用聚类方式，找到典型老师群体，请专家对这些老师进行评测，就能够得到一个相对客观科学的评价。这样我们就可以在老师群体中找到“锚点老师”。有了“锚点老师”，针对这些

老师确定“局部校正因子”，再用插值的方式找到一个“校正曲线”（如图2所示），最后对所有老师的数据进行校正。这种“先聚类再校正”，以锚点老师作为人机结合点，充分发挥机器计算优势与专家诊断优势，从管理上其成本也比较低，效果也不错。这种方法可以跨院系、跨学科来做对比分析。

2. 学生评语的聚类分析

评分是学生评教中的冰山一角。我们与许多老师交流，骨干教师对于学生评分高低并不敏感，而对于学生评语非常关注。评语往往比评分更可靠。在人工智能飞速发展的今天，已经有很多工具可以帮助完成量化分析。

(1) 构建评语常态模型。量化首先需要指定数据标准，即构建一个常态模型。我们从5年10个学期的评教数据中手动筛选出35万条有效评语，用计算机拆分为268万个词，基于该库的词数和词频、构建了一个常态模型。该

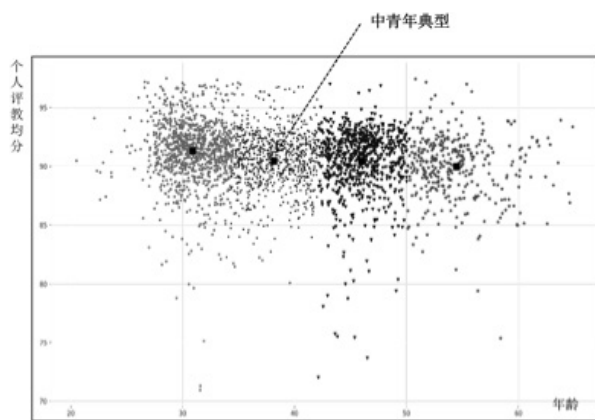


图1 单学期全校教师课程均分与年龄聚类分析图

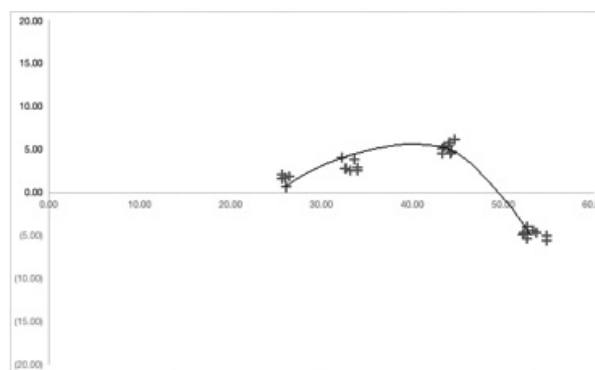
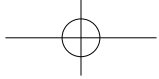


图2 聚类后局部校正曲线图



模型包含了词性、词频、年级、性别、学科等不同类型。有了这个分布，可以解决一部分学生评语失真的问题，不满足这个常态模型就应该判定为非正常的评语。

(2) 常态模型的新词修正。建模中发现每年学生评语中会出现大量的新词，因此这个常态模型需要不断改进。目前这部分工作还需要手工做。同时引申出许多新的研究方向，比如学生的新词趋势与行为文化，这是很有趣也很有意义的研究课题。

(3) 语句情感偏向度分析。我们用高频关键词来做常态模型，仅仅用于判定评语有效性，降低计算量。但学生的评语并不只有词，还有句式等语法结构。目前已经有人工智能可以分析语句的情感偏向度^[7]。我们采用腾讯公司提供的语义分析器来计算包含前 1000 个高频关键词的评语偏向度。不同学期存在两个峰值位置，一个是 0.682，另一个是 0.867。目前还不确定是学生评语内在的偏向度还是语义分析器的原因，但至少说明学生评语的偏向度是集中在某些点的。

(4) 基于评语偏向度的聚类分析。有了评语偏向度，就将评语量化为数值。沿用评分的聚类分析方式，我们构建一个评语的聚类值计算公式。

$$\delta = \frac{1}{M \cdot N} \sum_{j=1}^M \sum_{i=1}^N P_i \quad (1)$$

式(1)中 P_i 是某条评语的偏向度，某个关键词有 N 条评语就进行累加计算，然后将 M 个关键词再累加归一化，即可得到老师评语偏向度。图 3 就是部分老师的评语聚类图表，横轴是反向偏向度，纵轴是正向偏向度。

随着人工智能的发展，评语研究将更为广泛，在评教中也更有分量。从东南大学的实践中发现，学生评教指标体系对于最终聚类分析的影响有限。即在实践中可用更简单的方式来做大规模评教，鼓励学生多写评语，少做评分。2013 年我们在建筑学院设计课程中采用简化方式，只有两个评分项目，一个是老师教学对你的设计是否有效，另一个是老师的态度是否好，

但要求学生给出 50 字以上的评语。从实践上看，学生评教的效果比以往更好，老师阅读评教结果的比例也更高，院系使用结果的速度也更快。

四、结语

学生评教工作关系到学校的教学质量，本文建议确保学生作为大样本广覆盖的评教手段，利用学生评教数据将老师分类，由专家对各个教师群进行评价。学生的评教应用只将老师分到不同类里，然后专家对这个教师群做整体定性评价。这种方法可在校级、院系、学科等多层次多角度进行应用，协助教学管理层抓住典型问题与典型人群，再请专家去听“典型”老师的课，分析出这个教师群的问题，最后做专业的诊断，提出教学改革方案。而对于非典型性老师则可采用案例分析模式，或者做个案行政处置。

参考文献：

- [1] 江珊. 哈佛大学教学质量保障体系建设探析——基于学生评教的视角[J]. 高校教育管理, 2016, 10(2): 86-91.
- [2] 艾文国, 王桂伟, 关玉晶. 高校学生评教系统改进研究[J]. 中国大学教学, 2012(12): 76-79.
- [3] 林琳. 学生评教的困境、根源及其改进方向[J]. 当代教育科学, 2019(8): 11-14.
- [4] 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.
- [5] 李贞刚, 陈强, 孙婷婷. “以学生为中心”改进学生评教的思考与实践[J]. 现代教育管理, 2019(1): 62-66.
- [6] 马秀麟, 袁克定, 刘立超. 从大数据挖掘的视角分析学生评教的有效性[J]. 中国电化教育, 2014(10): 78-84.
- [7] 洪居兴, 王璐, 王虹, 祁俊, 陈秀明, 王世颖. 基于 Python 的情感数据分析[J]. 青海大学学报, 2019, 37(5): 97-104.

[责任编辑：周 杨]