

# 基于综合排序的高校学生评教思想、方法与实践

吴俊<sup>1</sup>, 肖宇<sup>1</sup>, 耿辉<sup>2</sup>

(国防科技大学 1. 系统工程学院; 2. 教学考评中心, 湖南 长沙 410073)

**摘要:** 作为高等教育内部质量保障体系的重要组成部分, 学生评教制度对于提升高等教育教学质量具有重要的积极作用。针对当前高校学生评教实践中存在的问题, 提出“分散指标体系下的综合排序”这一新的学生评教思想, 即每个学生不再对统一指标体系中的具体指标进行逐项打分, 而是根据个人心中特有的指标体系对评教周期内给他上过课的教师进行综合排序, 然后再通过构造竞争图将所有学生给出的大规模、不等长、只包含部分教师的排名列表加以聚合, 最后通过计算加权出度入度比来得到教师评分。实证分析结果表明该方法科学有效、可操作性强。

**关键词:** 高等院校; 学生评教; 综合排序; 排名聚合

**中图分类号:** G647 **文献标识码:** A **文章编号:** 1672-8874 (2018) 04-0078-06

## Method and Practice of Ranking-based Student Evaluation of Teachers

WU Jun<sup>1</sup>, XIAO Yu<sup>1</sup>, GENG Hui<sup>2</sup>

(1. College of Systems Engineering, 2. Teaching Evaluation Center, National University of Defense Technology, Changsha 410073, China)

**Abstract:** As an important measure to ensure the quality of higher education, student evaluation of teachers plays an important role in improving the quality of higher education. To overcome the drawbacks of the existing student evaluation of teachers, we propose the idea of “comprehensive ranking under decentralized index system”. Specifically, each student no longer scores the specific indicators in the unified index system item by item, but according to the invisible index system in his mind to rank comprehensively the teachers who have taught him in the last semester, and then a competition graph will be constructed, with which we can aggregate these massive, partial teacher rankings given by students, and then calculate the ratio of weighted out-degree and in-degree to score the teachers. Empirical analysis shows that the method is scientific, effective and operable.

**Key words:** university; student evaluation of teachers; ranking; rank aggregation

## 一、引言

高等教育的教学质量一直是一个备受关注的问  
题, 随着新时期教育教学改革不断深入推进,

高等教育质量保障体系建设正受到越来越多的关  
注<sup>[1-3]</sup>。如何构建合理有效的高等教育质量保障体  
系, 为提升高等教育质量提供有力支撑, 成为教  
育学研究的一个热门问题<sup>[4-5]</sup>。从20世纪90年代  
中期以来, 原国家教委开始实施普通高校本科教

学工作评估,特别是2003年开始实行的五年一轮的高等院校教学质量评估制度的建立,形成了一套教学工作评估的指标体系和评估方法,逐步建立了高等教育外部质量保障体系。在政府组织的高校教学工作评估得到充分发展的同时,高等教育内部质量保障体系也逐步得到完善,在促进高等院校教学改革、提升教学质量方面发挥着重要作用,与高等教育外部质量保障体系相互配合共同完成高等院校教育质量保障的任务。众所周知,教师在整个高等教育系统中扮演着一个非常重要的角色<sup>[6]</sup>,而教学质量的好坏很大程度上取决于教师在执行教学功能时的经验、知识和动机<sup>[7]</sup>。因此,教师教学评价作为高等院校内部质量保障体系的重要组成部分,也日益受到重视,得到了系统的发展<sup>[8-10]</sup>。

教师教学评价一般由学生评教、同行评价、教师自评和领导评价等构成。其中,学生作为最重要的利益相关者整个学期全程参与教师的教学过程,对教师教学的教学水平和教学状态最为了解。所以,与教师自评、领导评价和同行评价等其他教学评价方法相比,学生评教在衡量教师教学方面更具有可靠性和有效性<sup>[11-12]</sup>。目前,很多高等院校已经将学生评教作为一项制度确定下来,使之成为高等教育内部质量保障制度的重要组成部分。很多大学开发了学生网上评教系统,使学生评教更便捷、更有利于综合统计分析。同时,许多大学以学生评教结果为参考,建立了教师评价与奖励机制,将学生评教与教师职务晋升、津贴补助等联系起来。此外,部分大学还针对学生评教发现的问题,组织办学讨论和经验交流,优化师资配置,营造更有利的教学环境。这些反映出学生评教已经不仅仅是一项辅助性质量保障措施,而成为了高等院校教学质量保障的一项重要制度,对教学工作发挥着重要作用。

## 二、当前高校学生评教方法存在的问题

当前,作为评价高等教育质量的重要手段,学生评教已经渗透到高等教育研究与发展的方方面面。虽然不同高校的学员评教方法千差万别,

但是这些评教方法背后的核心思想都可以归纳为“统一指标体系下的逐项打分”,即先制定一个统一的评价指标体系,其中含有若干大项,每个大项由若干小项组成,参与评教的学生对每一个小项进行评分,最后依据一定的加权方案综合量化得到评教结果<sup>[13]</sup>。这种传统的评教方法经过几十年的积累已经形成了较为完善的理论体系,但在评教实践中很多弊端也逐渐凸现出来,主要有以下三个方面的问题:

1. 评价指标的模糊性。当前大学学生评教指标体系由若干评分项目组成,在大学学生评教实践过程中,指标体系内容多种多样,其中很多指标定义模糊,很难被参与评教的学生所理解,如某评教指标体系中有“教学理念是否先进”、“教学效果是否显著”等,而很多参与评教的学生很难理解“教学理念”、“教学效果”的具体意义,使得评教结果不能达到预期。

2. 指标体系的复杂性。随着对大学学生评教研究的不断深入以及学生评教实践活动的广泛开展,越来越多的指标被提出,涵盖了整个教学活动的方方面面,这也使得指标体系变得越来越复杂,进行一次评教操作甚至要完成几十上百个项目的量化打分,需要花费较长时间,极易使学生对评教活动产生抵触心理,使最终评教结果难以反映真实情况。

3. 指标权重的主观性。由于整个学生评教指标体系内容复杂多样,而且各个评分项目类别不同,在对各个教师进行综合量化评分时,需要对各个项目类别赋予不同的权重进行计算。由于不同的评价决策者对各个权重安排不同,亦即不同的决策者对各个项目类别的重要性认识不同,导致指标权重设计具有很强的主观性和不确定性,使得最终的评教结果很难被所有人接受。

## 三、基于综合排序的高校学生评教思想

随着信息化时代的大数据特征越来越明显<sup>[14]</sup>,针对各种各样的科学问题,传统的研究思路正在被更加新颖的“大数据”等新思维所超越。受益于“大数据”思维的启发,我们提出“分散指标体系下的综合排序”这一新的高校学生评教思想,用“分散指标体系”代替“统一指标体系”,用

“综合排序”代替“逐项打分”。

所谓“分散指标体系”是指不再制定统一的指标体系,每个学生在他自己心里建立一个“无形”的指标体系,然后依据这个特有的指标体系对老师进行评价。虽然,每个学生的指标体系可能有失偏颇,例如,有的学生注重课堂气氛的活跃程度,有的学生注重教学内容的新颖性,甚至有的学生主要关注教师的外貌,但是当把大量学生的指标体系叠加在一起的时候,我们就可能得到一个“最客观、最全面”的指标体系,亦即“没有指标的指标体系可能是最好的指标体系”。而且,这种“无形的指标体系”很好地解决了前面提到的“评价指标模糊性”以及“指标权重主观性”问题。

所谓“综合排序”是指不再对指标体系中的具体指标进行逐项打分,而是每个学生根据自己心中的指标体系对评教周期内给他上过课的教师进行综合排序。这种综合排序思想的好处首先在于大大简化了学生评教过程,缩短了评教所花费的时间,提升了学生评教的体验,解决了前面提到的“指标体系复杂性”问题。其次,这种综合排序思想可以有效避免学生在评教过程中给每个教师都打相同分数的“消极评教”现象,从而有效提高学生评教结果的区分度。另外,这种“强制排序思想”可以消除学员不敢给教师打低分的思想顾虑,从而提高了评教结果的客观性。

根据以上思路,针对现有学生评教方法存在的问题,本文首先提出基于综合排序的高校学生评教思想,建立基于教师排序的学生评教模型,并应用排名聚合方法将学生给出的大规模、只包含部分任课教师的排名列表加以聚合,得到教师的整体水平评价及排名。

#### 四、基于竞争图的高校学生评教方法

##### (一) 初始教师排名列表

令教师的个数为  $N$ , 学生的个数为  $M$ 。令  $R = (r_{ij})_{M \times N}$  表示排序矩阵, 亦即排名列表的矩阵表示, 其中矩阵的第  $i$  行  $R(i, :)$  表示学生  $b_i$  给出的排序列表, 在位置  $(i, j)$  上的元素  $r_{ij}$  表示教师  $a_j$  在学生  $b_i$  心中的排名。由于每个学生只能对部分曾经给他上过课的教师进行排序, 因此单个排名列表

通常都是不完全、不等长列表。若学生  $b_i$  没有给教师  $a_j$  排名, 则令  $r_{ij} = 0$ 。在表 1 的例子中, 我们给出了 8 名学生对 5 个教师的排名列表, 其对应的排序矩阵如下:

$$R = \begin{bmatrix} 5 & 4 & 3 & 2 & 1 \\ 0 & 4 & 2 & 3 & 1 \\ 3 & 4 & 1 & 2 & 0 \\ 0 & 3 & 2 & 0 & 1 \\ 4 & 3 & 2 & 1 & 0 \\ 4 & 3 & 0 & 2 & 1 \\ 3 & 4 & 2 & 0 & 1 \\ 4 & 3 & 0 & 1 & 2 \end{bmatrix}$$

表 1 8 名学生给出的关于 5 个教师的排名列表

学生	排名列表
$b_1$	$[a_5 \ a_4 \ a_3 \ a_2 \ a_1]$
$b_2$	$[a_5 \ a_3 \ a_4 \ a_2]$
$b_3$	$[a_3 \ a_4 \ a_1 \ a_2]$
$b_4$	$[a_5 \ a_3 \ a_2]$
$b_5$	$[a_4 \ a_3 \ a_2 \ a_1]$
$b_6$	$[a_5 \ a_4 \ a_2 \ a_1]$
$b_7$	$[a_5 \ a_3 \ a_1 \ a_2]$
$b_8$	$[a_4 \ a_5 \ a_2 \ a_1]$

##### (二) 教师之间的竞争图

为了描述教师教学水平之间的差异, 令  $P^i = (p_{st}^i)_{N \times N}$  为学生  $b_i$  的转移矩阵, 其中  $p_{st}^i = 1$  表示学生  $b_i$  认为教师  $a_s$  优于教师  $a_t$ , 反之,  $p_{st}^i = 0$ 。则, 学生  $b_8$  的转移矩阵为:

$$P^8 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

令  $A = (a_{st})_{N \times N}$  为竞争矩阵, 其中  $a_{st} = \sum_i^M p_{st}^i$  表示有多少学生认为教师  $a_s$  优于教师  $a_t$ 。因此, 上述例子中的竞争矩阵  $A$  为:

$$A = \begin{bmatrix} 0 & 2 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 0 \\ 4 & 6 & 0 & 2 & 0 \\ 5 & 6 & 2 & 0 & 1 \\ 4 & 6 & 4 & 3 & 0 \end{bmatrix}$$

为了更直观的反应教师教学水平之间的优劣关系, 我们基于竞争矩阵定义竞争图  $G = (V, E)$ , 其中  $V$  表示节点集合,  $E \subseteq V \times V$  表示边集合,  $N = |V|$  表示节点数量,  $W = |E|$  表示边数量, 竞争图的邻接矩阵即为  $A$ 。如果  $a_{st} > 0$ , 将会有一条有向边从节点  $V_i$  出发, 指向节点  $V_j$ , 且该边权重为  $a_{st}$ 。图 1 给出了上述示例对应的竞争图。

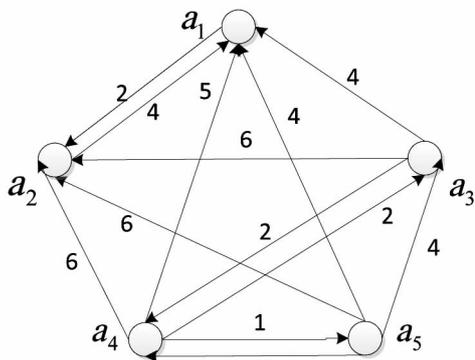


图 1 5 名教师的竞争图

(三) 加权出度入度比

为了得到最终学生评教结果, 在竞争图中我们定义  $d_j^+$  为节点的加权出度, 即所有从节点  $v_j$  出发的边上的权重总和, 定义  $d_j^-$  表示节点的加权入度, 即所有指向节点  $v_j$  的边上的权重总和。计算方式如下:

$$d_j^+ = \sum_{i=1}^N a_{ji};$$

$$d_j^- = \sum_{s=1}^N a_{sj}.$$

令  $\eta_j$  表示节点的加权出度入度比, 即

$$\eta_j = \frac{d_j^+}{d_j^-}$$

$\eta_j$  越大表示该教师教学水平越高。若  $d_j^+$  为 0, 即该教师在所有学生中均排名最后, 则  $\eta_j$  等于 0。若  $d_j^-$  为 0, 即该教师在所有学生中均排名第一, 则  $\eta_j$  等于无穷。若  $\eta_j$  等于 1, 则表明该教师的教学水平为“中等水平”。根据上述定义, 我们很容易计算得到表 1 示例中 8 个教师的加权出度入度比, 如表 2 所示。可以看出, 教师  $a_5$  的教学水平明显高于其他教师, 教师  $a_1$  的教学水平最差。

表 2 学生评教结果示例

教师	$d_j^+$	$d_j^-$	$\eta_j$	排名
$a_1$	2	17	0.12	5
$a_2$	4	20	0.20	4
$a_3$	12	6	2.00	3
$a_4$	14	5	2.80	2
$a_5$	17	1	17.00	1

五、实证分析

我们将上述学生评教方法在某高校本科学生评教中进行了实践应用, 得到了 2014 年—2017 年共计四年的本科学生评教数据。在具体操作中, 若一名教师有多门课程, 则将教师 + 课程组合 (以下统一简称教师) 作为一个被评价对象分别评价。本文中, 我们选取 2014—2015 学年秋季学期评教数据作为实证分析对象。该数据集中共有 7199 名学生以及 1139 名教师, 排名列表的平均长度为 13, 相较于 1139 名教师, 每个排名列表只包含了很少的一部分 (每个排名列表长度的大小反映了该学生本学期所修课程的数量)。

通过将 7199 个初始排名列表矩阵化可以得到排序矩阵  $R = (r_{ij})_{M \times N}$ , 进而得到竞争矩阵  $A = (a_{st})_{N \times N}$  以及竞争图, 最后计算得出各个节点的加权出度、加权入度及其比值, 得到每个教师的评分。图 2 给出了 1139 名教师的得分频率分布直方图。可以看出, 有 60% 左右的教师得分分布在 0.5 至 1.5 间, 表示 60% 的教师教学水平处于中等水平; 有 12% 左右的教师得分小于 0.5, 表示这些教

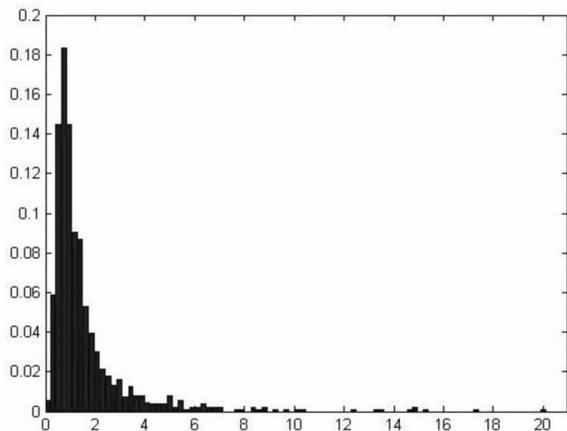


图 2 分布直方图

员的教學能力需要改进;有10%左右的教师得分超过了3,表示这些教员的教學能力得到了学生广泛的认可。上述结果表明,本文评教方法能显著区分教师教學水平之间的差异,具有很好的区分度。

基于该方法得到的学生评教结果,我们对不同属性、不同类别的课程授课教师进行了对比分析。首先将教师按照所授课程属性分为必修和选修两组,其中必修组有711名教师,选修组有428名教师,分析结果如表3所示。

表3 不同属性课程之间得分比较结果

组	课程数量	求和	平均值	标准差
必修	711	1528.203	2.14937	9.0825
选修	428	457.4934	1.06891	1.0171

从表3可以看出,必修课程的平均得分是2.14937,明显高于选修课程的平均得分1.06891。从整体上看,必修课程较选修课程更受学生的喜爱,在学生心目中更受欢迎。

其次,我们又将教师按照所授课程类别分为自修、专业基础、学科基础以及公共基础这四个类别,其中50门课程属于自修,103门课程属于专业基础,397门课程属于学科基础,589门课程属于公共基础。分析结果如表4所示。从表4可以看出,学科基础类型的课程较其他课程更受学生喜爱,而自修类的课程在排名中比较靠后。

表4 不同课程类别之间评分比较结果

组	课程数量	求和	平均值	标准差
自修	50	35.02216	0.700443	0.415683
专业基础	103	163.7056	1.589375	1.120617
学科基础	397	1008.222	2.539602	12.04268
公共基础	589	778.7468	1.322151	1.511843

## 六、结束语

作为高等教育教学质量评估体系的重要组成部分,学生评教已经在我国高等院校中得到了广

泛开展和应用。学生评教结果在教师绩效考评、教学方式改善、教育体系改革等方面发挥着越来越重要的作用。但由于当前学生评教中存在着诸多问题,学生评教也饱受诟病。

针对当前学生评教中存在的问题,本文对学生评教思想和方法进行了创新研究。首先提出了“分散指标体系下的综合排序”这一新的高校学生评教思想,用“分散指标体系”代替“统一指标体系”,用“综合排序”代替“逐项打分”,即不再对指标体系中的具体指标进行逐项打分,而是每个学生根据自己心中的指标体系对评教周期内给他上过课的教师进行综合排序,然后再通过构造竞争图计算加权出度入度比来得到教师教學水平的评价值。该方法克服了指标体系的复杂性,提升了学生评教的积极性,增强了评教结果的区分度。在高校学生评教中的实践也验证了本文方法的有效性和可行性。

此外,本文方法还有以下几个优点:(1)容错性好。在保证评教学生一定规模的情况下,个别学生随意打分不会对最终的评教结果产生显著影响。(2)可拓展性好。基于竞争图的学生评教方法能在原有的竞争图上加入新的评教数据,即补充新的加权有向连边,可以得到动态的教师教學水平评价结果。通过对该结果的时间序列分析,可以得出教师教學水平发展趋势,为学校提供更有价值的参考信息。(3)可推广性好。排名机制可应用于生活中的各个领域,如体育竞技、人员选拔、商品采购等,本文提出的学生评教方法可以推广到很多领域。

## 参考文献:

- [1] Marsh H W. Students' Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases and Usefulness [J]. Journal of Educational Psychology, 2007(5): 707-754.
- [2] 别敦荣,孟凡.论学生评教及高校教学质量保障体系的改善[J].高等教育研究,2007(12):77-83.
- [3] 王玲玉,朱景坤.学生评教基本理论探析[J].黑龙江教育:高教研究与评估版,2017(8):70-72.
- [4] Marsh H W. Students' Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases and Usefulness [J]. Journal of Educational Psychology, 2007(5): 707-754.

- [5] Costa C A B E, Oliveira M D. A multicriteria decision analysis model for faculty evaluation [J]. *Omega*, 2012 (4):424-436.
- [6] Marsh H W. Students' Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases and Usefulness [J]. *Journal of Educational Psychology*, 2007(5):707-754.
- [7] Rothstein J. Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement [J]. *National Bureau of Economic Research*, 2008(1):175-214.
- [8] 别敦荣,孟凡.论学生评教及高校教学质量保障体系的改善[J].*高等教育研究*,2007(12):77-83.
- [9] 王玲玉,朱景坤.学生评教基本理论探析[J].*黑龙江教育:高教研究与评估版*,2017(8):70-72.
- [10] Hativa N, Raviv A. Using a single score for summative teacher evaluation by students [J]. *Research in Higher Education*, 1993(5):625-646.
- [11] Marsh H W. Students' Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases and Usefulness [J]. *Journal of Educational Psychology*, 2007(5):707-754.
- [12] Hativa N, Raviv A. Using a single score for summative teacher evaluation by students [J]. *Research in Higher Education*, 1993(5):625-646.
- [13] 别敦荣,孟凡.论学生评教及高校教学质量保障体系的改善[J].*高等教育研究*,2007(12):77-83.
- [14] Caragiannis I, Krimpas G A, Voudouris A A. Aggregating Partial Rankings with Applications to Peer Grading in Massive Online Open Courses [C]//*International Conference on Autonomous Agents and Multiagent Systems*, Istanbul, Turkey:IFAAMAS, 2015:675-683.

(责任编辑:王新峰)

(上接第33页)

我校心理教育为军服务的作用。二是适时编著配套教材。尽管地方有不少高水平的有关“心理健康”、“幸福课”、“积极心理学”的教材和著作,但引进地方教材并不能很好适应本门课教学的需要。教学团队要在教学内容基本成熟的基础上,及时编写出版适应军校研究生需要的教材。

#### 参考文献:

- [1] 刘芳.积极心理学视野下我国大学生心理健康教育研

究[D].郑州:郑州大学,2014.

- [2] 李海云.90后大学生幸福感及其提升研究——基于积极心理学视角[D].武汉:华中师范大学,2014.
- [3] 韩立敏,文斌.军校学员心理服务需求分析与教育对策[J].*解放军理工大学学报:军事科学版*,2015(6):56-60.
- [4] 韩立敏,兰芬.努力提高军校学员心理服务质量[J].*军队政工理论*研究,2013(4):105-107.

(责任编辑:邢云燕)