

科技教育评价中主客观赋权方法比较研究¹

俞立平 潘云涛 武夷山

中国科学技术信息研究所 北京 100038

摘要：本文利用英国《泰晤士报高等教育增刊》2007年世界大学排名的原始数据，分别采用主成分分析、因子分析、TOPSIS法、秩和比法、灰色关联法、熵权法六种客观赋权法进行排序，然后比较了各种评价结果，并将这六种客观赋权评价结果与泰晤士报主观赋权评价结果进行了两两比较。研究发现：客观赋权法的评价结果无法得到公认，比较适用宏观分级评价；评价对象自身的差异程度会影响不同评价方法评价结果之间的一致性；在评价数据呈正态分布情况下，客观赋权法评价结果与权威主观评价结果相比，处于数据中间段的评价结果相差较大，而处于数据系列两端的评价对象会得到较为一致的评价；主客观赋权方法相结合是综合评价的发展方向。

关键词：科技教育评价 客观赋权 主观赋权 大学排名

中图分类号：G31

1 引言

科技教育评价是科教管理工作的重要组成部分，是推动国家科教事业持续健康发展，促进科教资源优化配置，提高科教管理水平的重要手段和保障。除定性评价方法外，目前国内外综合评价方法有数十种之多。这些评价方法各有其特点，但大体上可分为两类，其主要区别在确定权重的方法上。一类是主观赋权，多数采取综合咨询评分确定权重，然后对无量纲的数据进行综合，如综合指数法、模糊综合评判法、层次分析法、功效系数法等；另一类是客观赋权，根据各指标间相关关系或各指标值变异程度来确定权数，如主成分分析法、因子分析法、TOPSIS法等等。

在科技教育评价领域，很多主客观赋权方法都得到了应用，目前学术界进一步的研究包括两个方面，一是利用几种赋权方法综合进行评价；二是各种赋权方法评价结果的比较。对于后者，刘占伟、邓四二、滕弘飞^[1]（2003）从理论上分析了常见的几种评价方法的特点及存在的问题，包括层次分析法、模糊评价法、灰色理论法、物元分析法、聚类分析法、价值工程法、神经网络法以及综合评价法。陈衍泰，陈国宏，李美娟^[2]（2004）将各学科领域的综合评价方法归纳、分类，讨论了各类方法的基本原理、优缺点及适用领域，指出目前综合评价存在着三大问题：多方法评价结论的非一致性问题；方法针对性不强；理论研究与实际应用的脱节问题。吴清平、张丹^[3]（2003）在医疗工作质量评价中对秩和比法、层次分析法、TOPSIS法、系统聚类法的评价结果进行比较，发现秩和比法与其他方法具有一定的一致性。总体上这方面的研究不多，实证研究更少，存在的主要问题是，针对同一评价对象，选取相同的指标，采取同样的数据，但不同指标赋权方法得出的评价结果不一致，这个问题已被广大学者所注意到。进一步的问题是：不同评价方法评价结果之间有何特征？它们之间的关系如何？评价对象数量的多少和数据特征对评价结果有何影响？弄清楚这些问题有利于客观评价不同的评价方法，从而进行评价方法的选取，并且有利于对评价结果进行总结和评估，从而做到评价的公平公正，调动广大科研人员的积极性，促进科技教育事业健康发展。

为了对这个问题进行深入研究，必须找到完备权威的评价数据和参照标准，本文拟选取世界大学排名数据进行分析。目前世界有影响的世界大学排名有《美国新闻与世界报道》排

¹国家十一五支撑计划项目（2006BAH03B05）基于海量信息的科技评价方法、技术研究与应用（2007-2009）；国家自然科学基金资助（70673019）强竞争与弱竞争环境下科技发展与科技评价的科学计量学研究。（2006-2008）

作者简介：俞立平（1967-）男，江苏泰州人，博士，扬州职业大学副教授，工商管理学院副院长。主要从事信息经济、科学计量领域的教学科研工作。

名、《泰晤士报高等教育增刊》排名^[4]、上海交通大学排名、瑞士科技委员会排名、《亚洲周刊》排名等，所有排名均采用指标体系评价，除了《亚洲周刊》仅对亚洲一流大学排名外，其它都是全球大学排名。这些世界大学排名中，采取大样本调查的只有《泰晤士报高等教育增刊》，该机构2007年在进行同行评议时聘请了全球5101个专家，在雇主评价时调查了全球1471个单位。此外，从2007年开始，评价所利用的论文数据库已经由Elsevier的Scopus数据库取代了美国的SCI数据库，Scopus覆盖面更广，包括了许多非英语优秀科技期刊，这样对发展中国家的大学而言更为公平。邱均平^[5]（2007）认为《泰晤士报高等教育增刊》大学排名除了兼顾教学和科研外，在6项指标中还包括了2项国际化指标，反映了一流大学的时代色彩。因此本文采用英国《泰晤士报高等教育增刊》2007年世界大学排名数据进行研究，其评价结果具有较高的权威性。

本文首先用常见的6种客观赋权法进行评价，得到每种方法的世界大学排序，然后比较各种排序结果之间的差异，最后将每种客观赋权法评价结果与《泰晤士报高等教育增刊》世界大学排名进行比较，在此基础上进行深入分析。

2 方法与数据

2.1 研究方法

2.1.1 主成分分析与因子分析

主成分分析是考察多个变量间相关性的一种多元统计方法，其产生的背景是，评价研究中经常牵涉到多项指标，这些指标间往往存在一定的相关，全部采用这些指标，不仅使计算过程复杂，而且可能因多重共线性而无法得出正确结论。主成分分析的目的就是通过线性变换，将原来的多个指标组合成相互独立的少数几个能充分反映总体信息的指标。它常被用来作为寻找判断某种事物或现象的综合指标，并且给综合指标所包含的信息以合适的解释，从而更加深刻的揭示事物的内在规律。

因子分析可以看成是主成分分析的一种推广，因子分析的基本目的是用少数几个变量去描述多个变量间的协方差关系。其思路是将观测变量分类，将相关性较高即联系比较紧密的变量分在同一类中，每一类的变量实际上就代表了一个本质因子，从而可将原观测变量表示为新因子的线性组合。

2.1.2 TOPSIS 法

TOPSIS 的全称是逼近理想解的排序法（Technique for Order Preference by Similarity to Ideal Solution），它是多目标决策分析中常用的一种方法。该方法的思路是根据各被评估对象与理想解和负理想解之间的距离来排列对象的优劣次序。所谓理想解是设想的最好对象，它的各属性值达到所有被评对象中的最优值；而负理想解则是所设想的最差对象，它的各属性值都是所有被评对象中的最差值。用欧几里德范数作为距离测度，计算各被评对象到理想解及到负理想解的距离，距理想解愈近且距负理想解愈远的对象越优。

TOPSIS 法也可以进行加权，即在计算各评价对象与最优方案及最劣方案距离时，都可以赋予一定的权重，为了保证评价方法的客观性，本文不进行加权处理。

2.1.3 秩和比法（Rank Sum Ratio, RSR）

秩和比法是一种全新的广谱的实用数量方法，是田凤调^[6]（1993）发明的一种统计学方法，该方法集中了古典参数统计和近代非参数统计各自优势，通过指标编秩来计算秩和的一个特殊平均数，进而进行综合评价。该方法在国内有较大的影响。

2.1.4 灰色关联法

灰色关联分析是灰色系统分析的主要内容之一，用来分析系统中因素之间的关系密切程度，从而判断引起该系统发展的主要因素和次要因素。灰色关联分析的实质，就是比较若干数列所构成的曲线与理想数列所构成的曲线几何形状的接近程度，从而进行排序，列出评价

对象的优劣次序，评价标准是灰色关联度，其值越大，评价结果越好。

2.1.5 熵权法

熵概念源于热力学，后由 Shannon 引入信息论。信息熵可用于反映指标的变异程度，从而可用于综合评价。设有 m 个待评对象， n 项评价指标，形成原始指标数据矩阵 $X = (X_{ij})_{m \times n}$ ，对于某项指标 X_j ，指标值 X_{ij} 的差距越大，该指标提供的信息量越大，其在综合评价中所起的作用越大，相应的信息熵越小，权重越大；反之，该指标的权重也越小；如果该项指标值全部相等，则该指标在综合评价中不起作用。

2.1.6 Friedman 检验

Friedman 检验也称 Friedman 卡方检验，是 1937 年 Friedman 提出的检验方法。它是检验 K 个总体的分布中心是否有差异的一种方法。Friedman 提出的检验方法是独立地在每一个区组内各自对数据进行排秩。其统计量计算公式如下：

$$Fridman = \frac{12n}{k(k+1)} \sum_{i=1}^k (\bar{R}_i - \frac{k+1}{2})^2 \quad (1)$$

其中， n 为评价对象数， k 为评价方法序号， R 为秩。

本文拟比较不同评价方法之间的关系。由于各种评价方法原理不一，因此只能比较排序结果，于是，必须采用非参数检验方法，而 Friedman 检验就是一种公认的排序结果检验方法。

2.1.7 Kappa 一致性检验

1960 年，Cohen^[7]等提出使用 Kappa 值作为判断评价一致性程度的指标，该方法目前得到了广泛的应用，已成为判断一致性和信度评价的一种常用的统计学指标。Kappa 计算方法如下：

$$Kappa = \frac{P_A - P_e}{1 - P_e} \quad (2)$$

其中， P_A 为观测一致率， P_e 为期望一致率，即两次检验结果由于偶然机会所造成的一致率。显然 Kappa 值在 0~1 之间，若 Kappa 值越大说明一致性越好。

Kappa 一致性检验主要用于主客观赋权法的比较，与 Friedman 检验相比，能提供更多信息。

2.2 数据

本文采用数据为英国《泰晤士报高等教育增刊》2007 年世界大学排名数据，共有 200 所大学，主要指标有 6 个：同行评议、雇主评价（他们愿意招募哪个学校的毕业生）、师生人数比、人均引文数、海外教授数、留学生人数比，这 6 个指标中，同行评议与雇主评价是主观指标，每个指标数据已经进行标准化处理，各指标分值最高均为 100。《泰晤士报高等教育增刊》采取主观赋权法，6 个指标权重分别赋值为 0.4、0.1、0.2、0.2、0.05、0.05，在此基础上计算指标总得分，所有数据的描述统计量如表 1 所示。需要说明的是，《泰晤士报高等教育增刊》做大学排名时，由于只精确到计算结果的整数位，因此出现了一些大学并列排名的现象。本研究保留了计算结果小数点后两位，因此排序结果与原《泰晤士报高等教育增刊》排序略有差异。

表 1 原始数据描述统计量

Table 1. Description of Data

指标	权重	最小值	最大值	平均值	标准差	偏度	峰度	Kolmogorov Smirnov 统计量	P 值
同行评议	0.4	31	100	72.895	18.677	-0.098	-1.086	0.082	0.002

雇主评价	0.1	5	100	70.205	24.820	-0.707	-0.568	0.133	0.000
师生人数比	0.2	11	100	62.020	25.902	-0.114	-1.189	0.093	0.000
人均引文数	0.2	1	100	75.145	15.617	-1.496	3.813	0.102	0.000
海外教授数	0.05	13	100	59.135	27.768	-0.065	-1.399	0.108	0.000
留学生人数比	0.05	11	100	61.605	26.659	-0.165	-1.273	0.090	0.000

从指标均值看，人均引文数和同行评议分值较高并且标准差较小，师生人数比、海外教授数及留学生人数比三项指标偏低并且标准差较大。所有指标的偏度为负，说明概率密度分布是左偏的，即高分值部分数据较少，中低分值数据较多，或者说，超一流大学较少，其它大学较多。从峰度看，除人均引文数指标是尖峰外，其它均是低峰。根据 Kolmogorov Smirnov 正态分布检验结果，所有指标数据都不服从正态分布。在数据量很大的情况下，数据不服从正态分布的情况一般是不多的，本项排名数据不符合正态分布的主要原因是：世界上的大学为数众多，排名前 200 名的大学实际上都可以归类为世界一流大学，这些大学的相关数据并不是随机抽样数据。

3 实证结果

3.1 各种评价方法结果的比较

利用 6 个指标 200 所大学数据，分别采用主成分分析、因子分析、TOPSIS 法、秩和比法、灰色关联法、熵权法进行评价。由于评价方法不一，不能用直接计算出的分值进行比较，因此，给出了各种评价方法的排序结果，如表 2 所示。由于结果数据太多，这里只列出前 30 所大学作为参考。

在进行主成分分析与因子分析时，必须首先进行 KMO 检验与 Bartlett 检验。KMO 是对样本充分度进行检验的指标，一般要大于 0.5。本文采用 SPSS 进行数据处理，KMO 值为 0.485，也就是说，不太适合进行主成分和因子分析；Bartlett 值为 178.135， $P < 0.000$ ，也就是说，相关矩阵不是一个单位矩阵，可以进行主成分和因子分析。换句话说，大学排名采用主成分和因子分析的条件并不全部具备。前 3 个主成分（因子）的累计贡献率为 71.35%，因此采用前 3 个主成分（因子）进行评价。TOPSIS 法、秩和比法、灰色关联法对数据并没有太高要求。

表 2 《泰晤士报高等教育增刊》世界大学排序结果与客观赋权法排序结果的比较（前 30 名）

Table 2. The Comparison of the Times Ranking and objective methods' Ranking

大学	泰晤士 报排序	主成分 排序	因子 排序	TOPSIS 排序	秩和比 排序	灰色关 联排序	熵权 排序
Harvard	1	1	4	1	2	1	1
Princeton University	2	2	2	2	3	3	4
University College London	2	7	8	7	18	7	3
Imperial College London	2	4	3	4	6	4	6
Columbia University	5	11	11	8	7	9	2
Massachusetts Institute of Technology	6	3	1	3	4	2	8
McGill University	7	23	25	13	1	6	7
Duke University	7	5	7	9	8	8	9
University of Pennsylvania	9	8	5	5	5	5	5
Johns Hopkins University	10	6	17	30	12	12	19
Australian National University	11	9	19	31	16	15	22
University of Tokyo	12	10	6	11	15	10	11
University of Hong Kong	13	12	35	57	30	24	36

Stanford University	14	15	26	12	10	18	13
University of California, Berkeley	15	18	42	46	26	34	39
University of Edinburgh	16	14	9	20	24	17	14
King' s College London	17	20	51	73	44	41	54
Kyoto University	18	22	10	6	11	13	10
Ecole Normale Supérieure, Paris	19	13	31	50	22	19	31
University of Melbourne	20	16	16	15	17	23	15
Ecole Polytechnique	21	17	38	48	41	37	42
Northwestern University	22	19	23	21	31	21	20
University of Manchester	23	21	21	19	50	31	23
University of Sydney	24	25	12	10	19	20	12
Brown University	25	34	79	92	61	57	80
Peking University (北京大学)	26	40	50	33	9	45	40
University of Bristol	27	24	18	26	65	28	28
University of California, Los Angeles	28	27	44	52	28	47	16
Tsinghua University (清华大学)	29	26	24	16	13	25	49
Monash University	30	29	15	23	20	14	24

数据来源: <http://www.timeshighereducation.co.uk>

很显然, 各种评价方法结果存在差异。《泰晤士报高等教育增刊》将哈佛大学列为全球第一, 但因子分析只将它列为第 4。7 种评价方法, 有 5 种将哈佛大学列为全球第一。《泰晤士报高等教育增刊》前 30 所世界大学排名中, 若采用主成分分析法, 则其中有 2 所会排在 30 名之后; 类似地, 原排名 30 名之内的大学按新方法排序“名落孙山”的, 因子分析法有 8 所, TOPSIS 法有 9 所, 秩和比法有 6 所, 灰色关联法有 7 所, 熵权法有 7 所。但各种客观赋权法评价结果的前 30 名与《泰晤士报高等教育增刊》排序前 30 名基本相同。在 30 所大学中, 排序结果差距最大的是布朗大学, TOPSIS 法将其排在 92 位, 而《泰晤士报高等教育增刊》将其排在第 25 位。

为了对 7 种评价方法的结果进行总体比较, 利用 SPSS 进行 Friedman 检验。Friedman 值为 2.816, p 值为 0.832, 不能拒绝 7 种评价方法无差异的原假设。也就是说, 7 种评价方法总体上是无差异的, 排序先后大致能反映出大学水平差异, 由于各种评价方法实际上相差较大, 因此说, Friedman 检验在这种情况下是相对粗糙的。

3.2 主客观赋权法的二维散点图分析

由于《泰晤士报高等教育增刊》世界大学排名是较权威的主观赋权法评价, 这里以该排名为基础, 分别绘出其与各种客观赋权法结果的二维散点图, 如图 1~图 6 所示。

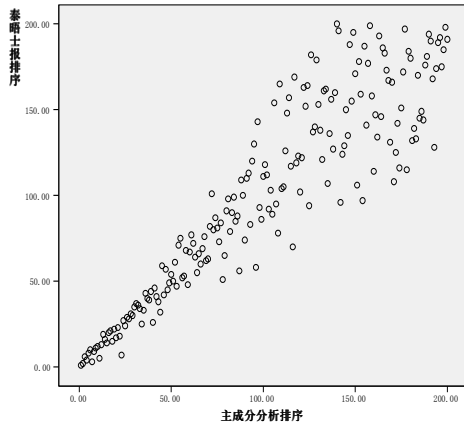


图1 泰晤士报评价与主成分分析评价散点图

Fig 1. The Dot plot of the Times Ranking and CFA

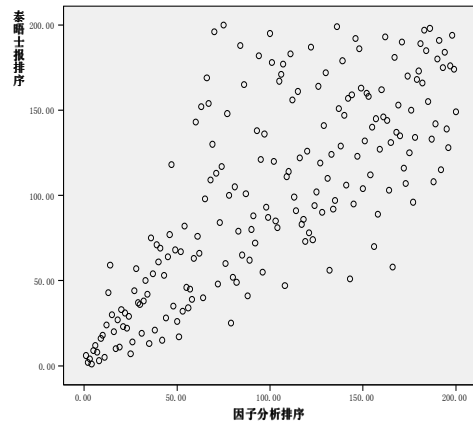


图2 泰晤士报评价与因子分析评价散点图

Fig 1. The Dot plot of the Times Ranking and FA

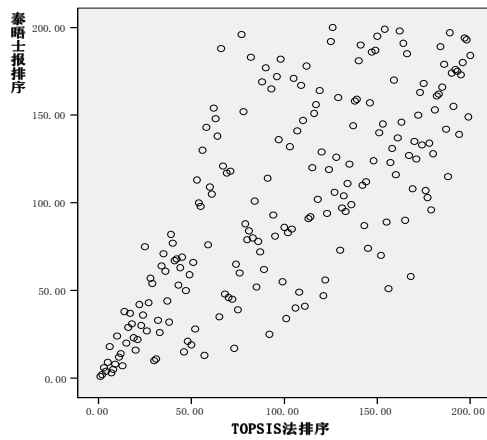


图3 泰晤士报评价与TOPSIS评价散点图

Fig 1. The Dot plot of the Times Ranking and TOPSIS

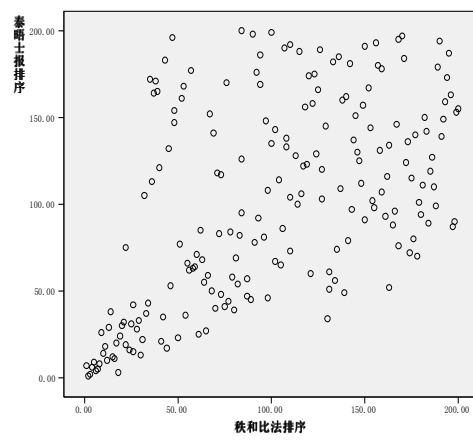


图4 泰晤士报评价与秩和比法评价散点图

Fig 1. The Dot plot of the Times Ranking and RSR

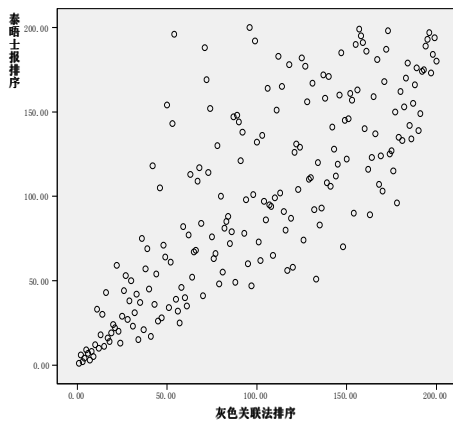


图5 泰晤士报评价与灰色关联评价散点图

Fig 1. The Dot plot of the Times Ranking and GCA

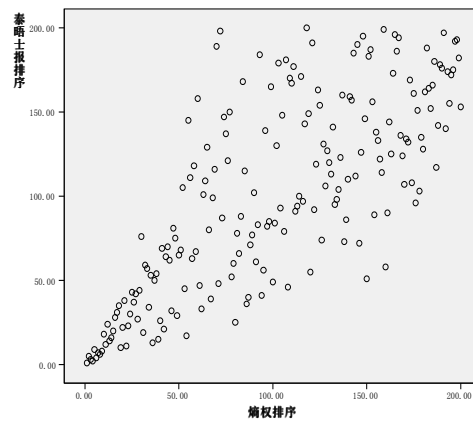


图6 泰晤士报评价与熵权评价散点图

Fig 1. The Dot plot of the Times Ranking and Entropy

如果各种评价方法与《泰晤士报高等教育增刊》评价结果基本一致，二维散点图应该是一条45度直线。从图1~图6可以看出，图1的散点图相对聚拢在45度线附近，说明《泰晤士报高等教育增刊》评价结果与主成分分析法基本一致；图4最为分散，说明《泰晤士报

高等教育增刊》评价结果与秩和比法排序差距最大。但是没有证据说明主成分分析就一定优于其它客观赋权法。

从所有图形的外形看，它们均呈现出一种不规则的纺锤形，左下角较细，右上角较粗，即各种评价方法与《泰晤士报高等教育增刊》评价结果在前部的差距较小，中间差距最大，尾部的差距又减小，但大于前部，这和原始指标的概率分布是有联系的。左下角的大学都是超一流大学，数量不多，各大学指标数据相差较大，容易区分，主客观赋权法结果相近；而由于概率密度分布是左偏的，右上角大学数量较多，各大学指标数据相差要小一些，主客观赋权法结果差异要大一些。

3.3 客观赋权法与泰晤士报排序结果的一致性检验

为了进一步比较各种客观赋权法与《泰晤士报高等教育增刊》排名的一致性程度，必须进行 Kappa 一致性检验。由于共有 200 所大学，按照 1: 2: 3: 4 的比例分为 4 个层次，即根据各种评价方法的排序结果，每种评价方法都分为 4 个档次，各档次大学数量分别为 20 所、40 所、60 所、80 所。需要说明的是，这 200 所大学实际上都可以将其归类为世界一流大学，因为原始数据不服从正态分布，如果采用通常的中间大两头小的分类方法不符合实际情况。

表 3-1 及表 3-2 给出了《泰晤士报高等教育增刊》评价结果与每种客观赋权法结果的二维列联表，同时给出了 Kappa 检验值和一致率，这里的一致率定义为两种评价结果完全一致的百分比，即列联表中对角线数据之和除以大学总数，如秩和比法的一致率为 $(15+13+21+47)/200=48\%$ 。

表 3-1 客观赋权法与《泰晤士报高等教育增刊》排序结果的比较

Table 3-1 The comparison of The Times Ranking and Objective methods' ranking

		主成分分析				因子分析				TOPSIS 法			
		1	2	3	4	1	2	3	4	1	2	3	4
泰	1	18	2			14	6			14	5	1	
晤	2	2	33	5		6	22	9	3	6	17	13	4
士	3		5	46	9		11	29	20		16	23	21
报	4			9	71		1	23	56		2	24	54
一致率		84%				60.5%				54%			
Kappa		0.771				0.436				0.343			
p		0.000				0.000				0.000			

表 3-2 客观赋权法与《泰晤士报高等教育增刊》排序结果的比较

Table 3-2 The comparison of The Times Ranking and Objective methods' ranking

		秩和比法				灰色关联法				熵权法			
		1	2	3	4	1	2	3	4	1	2	3	4
泰	1	15	5			16	4			15	5		
晤	2	5	13	16	6	4	15	10	1	5	20	13	2
士	3		9	21	30		8	32	20		13	25	22
报	4		11	22	47		3	17	60		2	22	56
一致率		48%				61.50%				58%			
Kappa		0.254				0.486				0.400			
p		0.000				0.000				0.000			

从表 3-1 及表 3-2 的结果可以看出，除了主成分分析与《泰晤士报高等教育增刊》评价

结果具有较高的一致性外 ($Kappa=0.771$), 其它评价方法与《泰晤士报高等教育增刊》评价结果一致性较低 ($Kappa<0.5$), 这与散点图分析的结果基本一致, 说明客观赋权法与主观赋权法所产生的评价结果差异较大。

4 结论与讨论

4.1 客观赋权法的评价结果无法得到公认

客观赋权法虽然排除了人为干扰, 但由于不同客观赋权法评价结果与人们比较认可的评价结果存在较大差异, 那么, 选取其中任何一种评价方法进行评价都不一定是合适的, 也就是说, 目前大量采用各种客观赋权方法进行的研究, 其可靠性是值得怀疑的, 结果也难以得到公认。

4.2 客观赋权法的结果适用于分级评价

对于科技教育评价对象, 采取不同的客观赋权法, 评价结果很难一致, 但如果将客观赋权法用来进行分级, 比如分为优、良、中、差 4 个档次, 却容易得到相对一致的结果, 因此, 客观赋权法可以作为一种等级分类方法, 用于宏观评价, 但不宜用于进行微观评比。在分级时, 既要考虑评价目的 (如评优有数量限制), 也要考虑不同级别断点之间的差距, 断点之间差距越大, 分级效果越好。

4.3 评价对象的数量及自身差距程度会影响客观赋权法评价结果

如果评价对象之间的差距较小, 那么客观赋权法评价结果与实际情况差距一般较大, 如果评价对象之间的差距较大, 即使采用不同的客观评价方法, 评价结果与实际情况可能也是一致的。比如在特别好与特别差的情况下, 评价对象间数据相差较大, 容易取得一致的评价结果。正因为如此, 在科技教育评价中, 会存在评价方法不当, 但评价结果却可以接受的情况, 或者说, 在评价数据自身差距较大的情况下, 会掩盖评价方法的不当, 这种情况比较普遍。

评价对象差距与其数量一般呈反比关系。如果评价对象众多, 评价对象的差距往往很小; 如果评价对象较少, 一般情况下评价对象的差距相对大一些。因此, 在评价对象较多的情况下, 建议以分级评价为主, 在评价对象较少的情况下, 无论排序和分级, 不同客观赋权方法容易取得相对一致的结果。

4.4 在评价对象数据服从正态分布情况下, 两极的排序结果更为一致

在评价对象原始数据服从正态分布的情况下, 处于中间段的评价对象的数据量较多, 数据间的差距较小, 不同客观赋权法产生的评价结果之差异较大; 处于数据系列两端的评价对象的数据较少, 数据间的差距较大, 不同客观赋权法的差异较小。如果某种客观赋权法对两端的评价拥有较好的效果, 并不一定说明该方法是最好的评价方法。

4.5 客观赋权方法有其适用范围

任何一种客观赋权方法都有其适用范围, 比如主成分分析与因子分析, 要求评价指标之间具有较高的相关性。在前提条件不具备的情况下, 不能滥用客观赋权法。

4.6 主客观赋权方法相结合是综合评价的发展方向

客观赋权法实施简单, 成本较低, 排除了人为干扰, 但结果无法得到公认。主观赋权能充分集中专家的智慧, 如果选取的专家合适, 赋权过程公正, 则主观赋权的评价结果具有权威性, 但该方法成本较高, 稳定性差。将来的发展趋势应该是客观赋权方法与主观赋权方法相结合, 在此基础上不断探索新的评价方法。

总之, 本研究在科技教育评价中不同评价方法评价结果的比较方面做了一些有益的探索, 由于客观事物的复杂性, 研究结论尚需继续进行验证和完善。进一步的研究需要对评价结果和权重进行灵敏度分析, 对评价价值体现出的评价对象之间的差距特征深入分析, 以期得到新的发现。

参考文献

- [1]刘占伟, 邓四二, 滕弘飞. 复杂工程系统设计方案评价方法综述[J]. 系统工程与电子技术, 2003 (12): 1488-1491
- [2]陈衍泰, 陈国宏, 李美娟. 综合评价方法分类及研究进展[J]. 管理科学学报, 2004 (4): 69-79
- [3]吴清平, 张丹. 秩和比法和几种常用评价方法在医疗质量评价中应用的比较[J]. 中国医院统计, 2003 (3): 3-5
- [4][HTTP://WWW.TIMESHIGHEREDUCATION.CO.UK](http://www.timeshighereducation.co.uk)
- [5]邱均平. 世界一流大学及学科竞争力评价研究报告[M]. 北京: 科学出版社, 2007: 6-7
- [6]田凤调. 秩和比法及其应用[M]. 北京: 中国统计出版社, 1993, 1-93
- [7]Cohen J. A coefficient of agreement for nominal scales[J]. Psychological Bulletin, 1960 (70) 213-220.

Comparison of objective weighting and subjective weighting in S & T and educational institutes assessment

Yu Liping, Pan Yuntao, Wu Yishan

Institute of Scientific & Technical Information of China, Beijing, China

Abstract: To see the difference among various objective weighting methods, this paper re-ranks the world 200 universities, as given by Times Higher Education Supplement, in terms of principal component analysis, factor analysis, TOPSIS, Rank Sum Ratio and grey system method based on 2007 Times world university ranking data, then compares the new rankings with Times rankings. The results show that objective evaluations can't be ensured entirely. Evaluation objects sitting in middle range of data series would get bigger ranking difference than those sitting on the two ends of data series when using different evaluation methods if the data observes the normal distribution. Objective evaluation methods seem more appropriate to attribute evaluated institutes into macro-scale rough categories. The gap between evaluated institutes affects the agreement among different objective evaluation results. The future trend seems to conduct more assessment exercises which combine subjective evaluation methods with objective evaluation methods.

Keywords: scientific and technology evaluation, educational institutes evaluation, objective evaluation, subjective evaluation, university ranking