

一种新的科技组合评价方法¹

——共性数据排序选择模型

俞立平 潘云涛 武夷山

摘要：为了解决科技评价中针对同一评价对象，选取相同的指标，采取同样的数据，但不同评价方法得出的评价结果不一致。本文以《泰晤士报高等教育副刊》世界大学排名为例，提出了一种新的组合评价方法——共性数据排序选择模型，其原理是，首先用各种可行的评价方法对科技评价对象进行评价，然后将评价结果排序后进行分级，筛选出各种评价方法公认的评价对象，接着采用排序多元选择模型进行回归，得到各指标的回归系数，将其标准化后作为组合评价的权重，最后进行加权汇总得到评价结果。共性数据排序选择模型克服了其他组合方法少数服从多数，对评价方法有限制的缺点。

关键词：科技评价 组合评价 共性数据排序选择模型 大学排名

1 引言

科技评价是科技管理工作的重要组成部分，是推动国家科技事业持续健康发展，促进科技资源优化配置，提高科技管理水平的重要手段和保障。评价方法与决策方法本质上是相通的，指标体系评价方法在决策中一般成为多准则决策 (MCDM) 或多属性决策 (MADM)，目前国内综合评价方法有数十种之多（当然这些方法并不仅仅应用于科技评价），这些评价方法各有特点，根据权重确定方式结合评价原理，本文将分为三大类（如图1所示）：第一类是主观赋权评价，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加权汇总，如综合指数法、专家会议法、德尔菲法等等；第二类是客观赋权评价，根据各指标间相关关系或各指标值变异程度来确定权数，然后进行加权汇总，如主成分分析法、因子分析法、熵权法等等。第三类是系统评价方法，其特点是综合运用运筹学、模糊数学、系统工程等领域的方法进行评价，评价结果不是根据原始指标值直接加权汇总。结果可以有具体的分值，也可能仅是排序。在评价过程中该类方法有些赋权，但与主客观赋权法相比，并不是简单的线性汇总，如ELECTRE法、模糊综合评价法、PROMETHEE等等；有些不赋权，如突变理论、数据包络分析等等。系统评价方法有些除了对数据标准化外，有些还进行进一步的变化处理。综合评价方法存在的主要问题是，针对同一评价对象，选取相同的指标，采取同样的数据，但不同评价方法得出的评价结果不一致，结果难以得到公认。

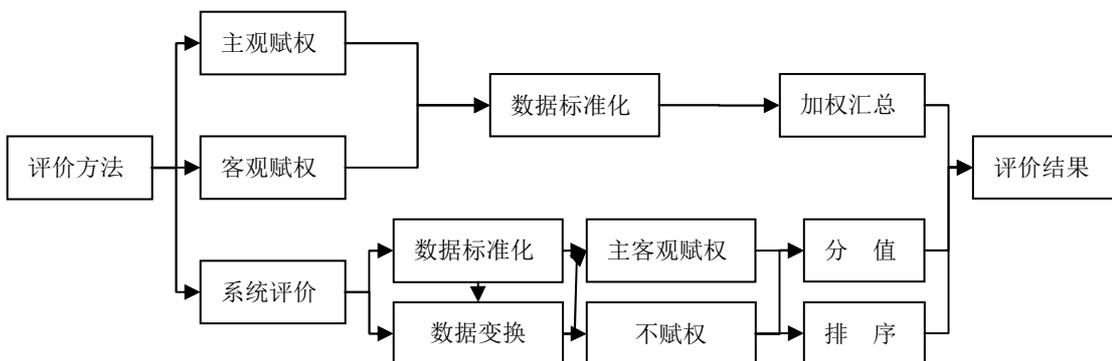


图1 评价方法体系结构

为了解决上述问题，学术界从不同角度进行了一些探索，首先是将不同评价方法结合起来使用，相互取长补短，这方面研究成果众多，但并没有找到最好的方法。其次是提出了“组合评价”的思想，即将各种评价方法的不同结果进行组合，从而给出最终评价结果。传统的

¹国家十一五支撑计划项目（2006BAH03B05）；国家自然科学基金资助（70673019）

组合方法有Borda法、Copeland法等，毛定祥^[1]（2002）提出了一种最小二乘意义下的组合方法，即最佳权重必须与各种主客观赋权评价所得权重的偏差平方和最小，用线性规划求解。徐泽水，达庆利^[2]（2002）提出的组合方法最终评价结果与各种评价方法差距最小。彭猛业、楼超华等^[3]（2004）提出了加权平均组合方法，权重采取各评价结果与平均评价价值的相关系数确定，最后计算组合评价结果。刘丽、张礼兵等^[4]（2004）利用遗传算法确定各评价方法的权重，然后计算组合评价结果。郭亚军、易平涛^[5]（2006）提出了一种基于整体差异的客观组合评价法，即存在一种评价结果（最佳组合），各评价方法投影到该评价结果后方差最大，体现了少数服从多数的思想。

从目前的组合方法看，都是一种求同存异的方法。而且新的问题出现了，和评价方法本身的多样性一样，各种评价结果的组合方法本质上也是一种“评价”，也就是说，从理论上讲，至少有数十上百种组合方法，那么哪种组合方法更好呢？这似乎进入了一个怪圈。其次，以上组合方法的前提是，各种评价结果必须有固定的评价价值，而某些系统评价方法评价结果只有排序，即使有评价价值，其含义与通常意义上主客观赋权法结果分值的含义是不一样的，如突变理论、ELECTRE法等，必须采用全新的思路进行评价方法的组合。

本文采取一种新的组合评价方法——共性数据排序选择模型 CDOC（Common Data Ordered Choice Model）。对于科技评价对象，采用不同评价方法进行评价，如果严格进行排序，结果很难得到公认，但是如果仅仅简单对评价对象进行分段评价，比如优、良、一般，却容易得到相对一致的结果。具体处理程序是：首先将其评价结果按优、良、一般进行分级（假设百分比分别为20%、60%、20%），然后找出各种方法公认的优、良、一般评价对象，在此基础上用排序选择模型进行回归，从而得出各评价指标的系数，经过标准化后即是各指标的权重，最后对所有评价对象进行加权得到评价结果。

为了进行深入研究，本文选取2007年《泰晤士报高等教育副刊》世界大学排名^[6]数据进行分析。大学排名是一项非常敏感的工作。目前世界有影响的世界大学排名有《美国新闻与世界报道》排名、《泰晤士报高等教育增刊》排名、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。《泰晤士报高等教育副刊》每年推出世界大学排名，将同行评议与指标体系结合起来进行大学排名综合评价。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，这是因为Scopus覆盖面更广，包括了许多非英语优秀科技期刊，这对发展中国家的大学而言相对公平。《泰晤士报高等教育副刊》除了公布主观赋权评价结果外，还公布了各大学各项指标的原始数据，这就为深入分析提供了可能。

本文的重点在于研究组合评价方法，因此选用一些代表性的评价方法进行分析。主观赋权评价采用《泰晤士报高等教育副刊》本身的评价结果；客观赋权法采用主成分分析、因子分析、熵权法3种评价方法；系统分析评价法采用TOPSIS、灰色关联2种方法。然后采取共性数据排序选择模型进行评价组合，最后得出组合评价结果。

2 研究方法

2.1 研究的基本前提

1、没有任何一种评价方法的结果能够得到公认

如果某种评价方法的结果能得到一致公认，比如《泰晤士报高等教育副刊》大学排名所选取的指标、同行评议的方法、指标权重的赋值、评价的结果能够得到公认，那么本文的工作是没有意义的。本文试图针对科技评价不同评价方法结果不一致问题，探索一般意义上的解决思路。

2、不存在评价方法的误用问题

所有采用的评价方法必须是合适的,不存在评价方法的误用问题。任何评价方法都有一定的适用范围,如果方法选取本身存在争议,那么在此基础上所从事的后续组合评价必然存在问题。

2.2 几种客观赋权评价方法

主成分分析(Principle Components Analysis)是考察多个变量间相关性的一种多元统计方法,它通过线性变换,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(Factor Analysis)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。

熵(Entropy)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。

以上 3 种客观赋权法评价时,根据不通的机理分别给指标赋权,然后进行加权平均,得到总体评价分值。主成分分析和因子分析在指标间相关度较高的情况下容易取得较好的评价结果。

2.3 几种系统分析评价方法

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 法可以赋权,但为了保证客观性,本文没有赋权。

灰色关联分析(Grey Relational Analysis)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。

2.4 排序选择模型

在多元回归中,通常情况下因变量是连续变量,当因变量只有两种选择时,称为二元选择模型,如购买或者放弃,赞成或者反对等。在现实生活中,当因变量不止两种选择时,就要用到多元选择模型(multiple choice model),多元选择模型又分为两种,一种是因变量之间存在优先顺序,如一个人购买汽车时,首先考虑价格、其次考虑安全、最后考虑节油。另一种情况是因变量有多种选择,但没有先后次序。前者称为排序选择模型或者排序多元离散选择模型,后者就是一般的多元选择模型。

在排序模型中,作为被解释变量的观察值 y 表示排序结果或者分类结果,其取值为整数,如 $0,1,2,3, \dots$ 。解释变量 x_i' 是可能影响被解释变量排序的各种因素, x_i' 可以是多个解释变量的集合,即向量。

排序模型的一般形式是: $y_i^* = x_i' \beta + \varepsilon$

式中, y_i^* 是隐变量 (latent variable) 或潜变量, 是不可观测的, x_i' 是解释变量集合, β 是待估计参数, ε 是随机扰动项。相对于显式变量 y 而言, 隐变量 y_i^* 没有观察值, 一个典型的解释变量是把隐变量理解为某种效用, 效用的大小可以用数值衡量。在估计排序模型时, 只需输入 y 的观察值和各解释变量 x_i' 的观察值。隐变量 y_i^* 由解释变量 x_i' 作线性解释后, 依据 y_i^* 所对应的如下规则, 对 y_i 进行排序分类:

$$y_i = \begin{cases} 0 & \text{if } y_i^* \leq \gamma_1 \\ 1 & \text{if } \gamma_1 \leq y_i^* \leq \gamma_2 \\ 2 & \text{if } \gamma_2 \leq y_i^* \leq \gamma_3 \\ \dots\dots \\ m & \text{if } \gamma_m \leq y_i^* \end{cases}$$

式中, 各 γ_j 是决定 y_i 排序的门限值 (threshold values) 或阈值。决定 y_i 排序的值是 0, 1, 2, $\dots\dots m$, 也可以是任意值。排序模型要求, 对于 y_i^* 而言, 较大的 y_i 对应于较大的隐变量 y_i^* 。所以, 当 $y_i < y_j$ 时, 就意味着 $y_i^* < y_j^*$ 。

各 y 观察值的概率由下式确定:

$$\begin{aligned} P(y_i = 0 | x_i, \beta, \gamma) &= F(\gamma_1 - x_i' \beta) \\ P(y_i = 1 | x_i, \beta, \gamma) &= F(\gamma_2 - x_i' \beta) - F(\gamma_1 - x_i' \beta) \\ P(y_i = 2 | x_i, \beta, \gamma) &= F(\gamma_3 - x_i' \beta) - F(\gamma_2 - x_i' \beta) \\ \dots\dots \\ P(y_i = m | x_i, \beta, \gamma) &= 1 - F(\gamma_m - x_i' \beta) \end{aligned}$$

式中 F 是 ε 的累积分布函数。如果选择 Probit 模型, F 就是标准的分布函数; 如果选择 Logit 模型, F 就是逻辑分布函数。由此可知, 排序模型估计得到的实际上由各观察值 y 落入到不同区间 (等级) 的概率。

γ 是与系数 β 一起估计的门限值 (阈值), 由极大化下列对数似然函数得到 β 和 γ :

$$\begin{aligned} L(\beta, \gamma) &= \sum_{i \ni y_i=0} \log(p(y_i = 0 | x_i, \beta, \gamma)) + \sum_{i \ni y_i=1} \log(p(y_i = 1 | x_i, \beta, \gamma)) \\ &+ \dots\dots + \sum_{i \ni y_i=m} \log(p(y_i = m | x_i, \beta, \gamma)) \end{aligned}$$

需要指出的是, m 个临界值 $\gamma_1, \gamma_2, \dots, \gamma_m$ 事先也是不确定的, 所以也作为参数和系数一起回归。利用经济计量软件 Eviews5.0 可以方便地进行估计。

3 变量与数据

《泰晤士报高等教育副刊》2007 年世界大学排名数据有 200 所大学, 共 6 个指标有:

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。《泰晤士报高等教育副刊》采取主观赋权法，6 个指标权重分别赋值为 0.40、0.10、0.20、0.20、0.05、0.05，在此基础上计算指标总得分。各指标数据已经标准化，描述统计量见表 1。

表 1 原始数据描述统计量

变量名称	指标	权重	最小值	最大值	平均值	标准差
Y	评价结果分级		优=3	良=2	一般=1	
X1	同行评议	0.40	31	100	72.895	18.677
X2	雇主评价	0.10	5	100	70.205	24.820
X3	师生人数比	0.20	11	100	62.020	25.902
X4	人均引文数	0.20	1	100	75.145	15.617
X5	海外教授数	0.05	13	100	59.135	27.768
X6	留学生人数比	0.05	11	100	61.605	26.659
N	大学数量			200		

4 实证结果

4.1 各种评价方法的使用及共性数据的获取

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、熵权法、TOPSIS 法和灰色关联法进行评价。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。各种评价方法评价结果如表 2 所示，由于数据较多，因此只公布了排名《泰晤士报高等教育副刊》主观赋权前 40 位大学的结果。

将《泰晤士报高等教育副刊》主观赋权评价结果与 5 种客观评价方法的结果分别进行降序排列，得到每种方法每个大学的排名。为了进行数据的分级，根据正态分布的中间多、两头少的原理，本文将 200 所大学按 20%、60%、20%的比例分为优、良、一般三个档次，然后分别进行共性数据的选取。在优档次中共有 40 所大学，取 6 种评价方法的交集后仅 19 所，即各种评价方法公认为优的大学有 19 所。同理，各种评价方法公认为良的有 64 所，各种方法公认为一般的大学为 12 所，共有 95 所大学，占整个大学数量的 47.5%。

表 2 各种评价方法及组合评价排序结果

大学	主观 赋权	主成分 分析	因子 分析	熵权法	TOPSIS	灰色 关联	组合 赋权
哈佛大学	1	1	4	1	1	3	2
牛津大学	2	2	2	3	2	2	3
耶鲁大学	3	7	8	6	7	9	8
剑桥大学	4	4	3	4	4	7	4
普林斯顿大学	5	11	11	8	8	24	10
伦敦大学帝国学院	6	3	1	2	3	6	1
加州理工学院	7	23	24	7	13	1	9
芝加哥大学	8	5	7	9	9	10	6

伦敦大学学院	9	9	5	5	5	18	5
麻省理工学院	10	6	17	19	30	15	12
哥伦比亚大学	11	8	19	22	31	27	18
麦吉尔大学	12	10	6	11	11	16	7
杜克大学	13	12	35	36	57	21	34
宾西法利亚大学	14	15	26	13	12	64	25
约翰霍普金斯大学	15	18	42	40	46	46	40
澳大利亚国立大学	16	14	9	14	20	12	17
东京大学	17	20	50	54	73	31	58
香港大学	18	22	10	10	6	28	11
斯坦福大学	19	13	31	31	50	8	20
卡内基梅隆大学	20	16	16	15	15	67	14
康奈尔大学	21	17	38	42	48	26	39
加州大学伯克利分校	22	19	25	20	21	25	15
爱丁堡大学	23	21	20	24	19	63	26
伦敦大学国王学院	24	25	12	12	10	73	21
京都大学	25	34	79	80	92	51	86
巴黎高等师范大学	26	40	51	39	33	69	41
墨尔本大学	27	24	18	28	26	20	22
东北大学	28	27	44	49	52	74	46
巴黎综合理工大学	29	26	23	16	16	32	23
悉尼大学	30	29	15	18	23	17	16
曼彻斯特大学	31	28	21	25	18	49	27
布朗大学	32	44	53	46	38	95	49
新加坡国立大学	33	35	22	17	32	4	13
北京大学	34	31	56	86	101	22	94
英属哥伦比亚大学	35	30	48	64	65	38	52
昆士兰大学	36	33	30	35	24	78	37
布里斯特大学	37	32	29	26	17	68	35
香港中文大学	38	43	32	21	14	36	29
密西根大学	39	38	58	67	75	48	60
清华大学	40	37	64	88	106	33	90

4.2 共性数据排序选择模型组合评价

数据筛选后，下面是采取排序选择模型进行回归，但本文数据在回归时出现了多重共线性问题，这是由于数据本身原因所造成的。为了解决这个问题，本文试着将大学分级由原来的3种改为4种、5种，但都没有得到改善。因此用大学分级分别与同行评议、雇主评价、师生人数比、人均引文数、海外教授数、留学生人数比进行回归，然后将每个变量的回归系数进行标准化处理，得到权重，结果如表3所示。

表3 共性数据排序选择模型及赋权结果

变量名称	指标	回归结果	组合权重	原主观权重
X1	同行评议	0.055*** (5.446)	0.22	0.40
X2	雇主评价	0.044*** (5.402)	0.17	0.10

X3	师生人数比	0.028*** (4.940)	0.11	0.20
X4	人均引文数	0.036*** (3.305)	0.14	0.20
X5	海外教授数	0.025*** (4.626)	0.10	0.05
X6	留学生人数比	0.066*** (5.289)	0.26	0.05

注：括号内为 Z 检验值，***表示在 1%的水平上统计检验显著

从回归结果看，所有变量的参数均比较显著，在 1%的水平上通过了 Z 检验，回归系数经标准化后得到各指标的权重，与《泰晤士报高等教育副刊》的主观权重相比，组合权重中留学生人数比变化最大，由最低权重变成最高权重，其他指标权重变化也较大，根据新的组合权重计算的新的排序结果见表 2（前 40 位）。

为了检验组合赋权后的排序结果和其他 6 种评价方法是否一致，本文进行了对照检验，将 7 种评价结果放在一起，参照共性数据的获取方法，进行分级并寻找 7 种方法的共性数据，发现也是同样的 95 所大学，说明采用排序选择模型进行评价方法的组合是有效的。

5 结论与讨论

5.1 共性数据排序选择模型的优点

1、与现有的其他组合方法相比，共性数据排序选择模型是一种求同不存异的方法，其选取的样本数据排序结果是所有评价方法公认的，克服了其他方法少数服从多数、寻求各种方法妥协的弊病。

2、传统的组合评价必须有评价价值，而除了主客观赋权评价是加权汇总外，很多系统评价方法结果值的原理是不一样的，甚至只有排序，这就限制了某些系统评价方法的选用，共性数据排序选择模型可以方便地处理排序结果。

3、其他组合方法所依托的数据是各种评价结果，这是一种中间数据，由于评价机理不一，中间数据原理各不相同，进行组合容易出现误差，共性数据排序选择模型关注的是各种评价方法结果排序的筛选，重新组合时采取的是原始评价数据，从而可靠性更高。

5.2 关于回归处理的讨论

在用排序选择模型进行处理时，最常见的问题是统计检验问题，如似然比统计量、拟 R^2 值、Z 检验值等，由于是一种共性数据的处理，因此，本文认为在统计检验没有通过时可以适当降低要求。对某些指标，如果系数符号出错，要慎重分析该指标的合理性，寻找更为合适的指标或删除一些不合适的指标。当然，一旦指标有所变化，所有的各种评价及组合评价必须重新进行。

共性数据排序选择模型对原始数据的数量有一定的要求，数据量过少，必然自由度不够，由于筛选各种评价结果的共同值，肯定会牺牲部分原始数据，这个问题在评价数据本身比较少时更为突出。因此共性数据排序选择模型是一种数据较多时的可靠的组合评价方法，其适用范围可以是各个领域的各种多指标综合评价和一些决策分析。

当然，共性数据排序选择模型还需要进一步从理论和实践上进行完善，比如分级多少、各级数量的比例？多重共线性问题的解决方法？变量不能通过统计检验如何处理等等。

参考文献

- [1]毛定祥. 一种最小二乘意义下主客观评价一致的组合评价方法[J]. 中国管理科学, 2002(5):95-97

- [2]徐泽水, 达庆利. 多属性决策的组合赋权方法研究[J]. 中国管理科学, 2002 (2): 84-86
- [3]彭猛业、楼超华等. 加权平均组合评价法及其应用[J]. 中国卫生统计, 2004 (3): 146-149
- [4]刘丽、张礼兵. 基于遗传算法的组合评价模型[J]. 合肥工业大学学报自然科学版, 2004 (8): 899-902
- [5]郭亚军、易平涛. 一种基于整体差异的客观组合评价法[J]. 中国管理科学, 2006 (3): 60-64
- [6][HTTP://WWW. TIMESHIGHEREDUCATION. CO. UK](http://www.timeshighereducation.co.uk)