

# 科技评价指标相关消除方法——相关系数调整法

俞立平 潘云涛 武夷山

**摘要：**为了消除科技评价中指标相关导致重复计算，影响评价精度和效果的问题。本文以科技期刊评价为例，采用相关系数调整法解决这个问题，取得了良好的效果。其原理是：首先判定指标间是否具有两两相关关系，如果有，用较小的指标乘以拟合优度作为相关部分的估算值，然后采用客观相关系数调整法或权重相关系数调整法对指标相关部分进行修正，并可以采用指标相关优化系数检验调整效果，从而降低指标间的相关性，最后将调整后的指标数据再一次标准化后加权汇总进行评价。该方法一定程度上可以淘汰旨在消除指标相关的某些客观评价法。

**关键词：**科技评价 指标相关 相关系数调整法 指标相关优化系数

**中图分类号：**G301

## 1 引言

科技评价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，调动科技工作者的积极性，提高科技管理水平的重要手段和保障。在科技评价中，多指标综合评价是一种主流的评价方法，主要通过主观和客观赋权法确定权重，然后进行加权汇总。但是由于科技评价工作的复杂性，指标间往往存在相关关系，如总被引频次、影响因子、即年指标三个指标明显存在相关关系，导致评价时重复计算，影响评价的精度和效果，这方面的问题至今没有得到较好的解决。

王伟夏<sup>[1]</sup>（1988）从集合论的角度分析了指标之间存在重迭性而引起的重复计算及其对评价造成的不利影响，将存在重复相关关系的指标分别进行分解，分析各个指标重复的因素，把重复的因素分离出去，得到一组相互独立的指标，但该方法在具体处理时没有唯一解，而且无法考虑主观因素。黄擎明<sup>[2]</sup>（1991）提出采用指标间的两两相关系数来修正权重，从而消除指标相关的影响。但是没有考虑指标间不相关的情况，并且计算时权重调节系数只能根据经验确定。王庆石<sup>[3]</sup>（1994）提出消除重叠指标的复相关系数和多元回归法，问题是剔除相关部分必然损失重要信息。彭勇行、张茂祥<sup>[4]</sup>（1994）认为应该精选指标，尽量减少指标的总量，从而减少指标间的相关性，当然这种处理方法的作用是有限的。徐祥发、肖人彬<sup>[5]</sup>（2002）在评价指标相关矩阵的基础上，首先对指标集合进行等价类划分，然后利用随机向量线性相关性，对指标等价类加以优化求解，使得指标等价类之间的相关性最小，进而得到等价类的指标综合函数，最后根据等价类的指标综合函数进行系统评价，这是一种完全依赖数据的客观赋权评价方法。此外主成分分析法和因子分析法被公认为消除指标间相关关系的一种较好的解决方法。这些完全依赖数据说话的客观评价方法是不全面的，忽视了评价时的主观因素。

本文利用中国科学技术信息研究所的科技期刊原始数据，首先从理论上分析指标相关对评价结果的影响，然后通过一种新的相关系数调整法（CRA, Correlation Ratio Adjustment）计算出指标间的重叠部分，在此基础上重新修正指标，将由于指标相关造成的冗余信息去除，最后再对指标二次标准化后进行加权汇总，从而得出更为科学合理的评价结果。

## 2 研究方法——相关系数调整法

### 2.1 两个指标相关对评价结果影响分析

为了简化计算，首先分析仅有两个评价指标相关时对评价结果的影响。

假设有  $n$  个评价对象，有两个评价指标  $X$ ,  $Y$ ，权重分别为  $a$ 、 $b$ ，显然有  $a+b=1$ ，若不

考虑 X 和 Y 之间的相关性，则评价值为：

$$P = aX + bY \quad (1)$$

如果将评价结果按照升序排列，相邻两个评价对象的差为：

$$\Delta = (ax_{i+1} + by_{i+1}) - (ax_i + by_i) \geq 0 \quad (2)$$

#### 情况一：

现在考虑 X 和 Y 之间存在相关关系的情况，假设相关部分为 S（superposition），若将相关部分剔除，则评价值为：

$$p_i = a(x_i - s_i) + b(y_i - s_i) = ax_i + by_i - (a+b)s_i = ax_i + by_i - s_i \quad (3)$$

相邻两个评价对象之间的关系为：

$$\Delta' = [ax_{i+1} + by_{i+1} - s_{i+1}] - [ax_i + by_i - s_i] = \Delta + (s_i - s_{i+1}) \quad (4)$$

由于  $\Delta \geq 0$ ，若  $s_i - s_{i+1} \geq 0$ ，则考虑变量相关关系后，排序不变。然而通常情况下，X 和 Y 越大，S 越大；X 和 Y 越小，S 越小。考虑到升序排列，因此  $s_i \leq s_{i+1}$  的概率很大，即  $s_i - s_{i+1} \leq 0$  的概率很大，导致  $\Delta'$  可能为正值、负值或 0。也就是说，考虑指标间相关以后，总体上评价值变小，评价排序有可能发生变化。

#### 情况二：

考虑 X 和 Y 之间存在相关关系的情况，相关部分为 S，若将相关部分完全剔除会丢失重要信息，尤其在相关程度较高的情况下，最好的方法是避免重复计算。比如将相关部分从 X 上剔除，全部放在 Y 上计算，则评价值为：

$$p_i = a(x_i - s_i) + by_i = ax_i + by_i - as_i \quad (5)$$

相邻两个评价对象之间的关系为：

$$\Delta' = [ax_{i+1} + by_{i+1} - as_{i+1}] - [ax_i + by_i - as_i] = \Delta + a(s_i - s_{i+1}) \quad (6)$$

同样有  $\Delta \geq 0$ ，并且  $a(s_i - s_{i+1}) \leq 0$  的概率很大，因此  $\Delta'$  可能为正值、负值或 0，与不考虑指标间相关关系相比，同样很大程度上会导致评价值变小和排序发生变化。

由于  $a < 1$ ，因此当  $s_i > s_{i+1}$  时， $a(s_i - s_{i+1}) < s_i - s_{i+1}$ ；当  $s_i < s_{i+1}$  时， $a(s_i - s_{i+1}) > s_i - s_{i+1}$  因此与情况一相比，情况二对排序的影响要比情况一要小一些。即将相关部分做部分剔除对不考虑相关排序的影响要小于将相关部分完全剔除后对不考虑相关排序的影响。因此情况二是一种比较温和的处理方式，而且不会丢失重要信息，这是一种较好的处理办法。

如果相关部分 S 的值可以计算，那么就可以进行精确计算和排序，从而可以对评价进行修正，得到更为客观公正的结果。

本文采取一种近似算法，假设以 Y 为因变量，X 为自变量进行回归，r 为调整后的  $R^2$ ，则相关部分 S 为

$$m_i = \min(x_i, y_i)r \quad (7)$$

这种近似算法有如下特点：

1、由于 X 和 Y 是相关的，一般  $x_i$  和  $y_i$  是同步变化的， $x_i$  和  $y_i$  越大， $s_i$  也越大； $x_i$  和  $y_i$  越小， $s_i$  也越小，这和实际情况是相符的。

2、调整后的  $R^2$  是指标相关程度的一个重要衡量指标，调整后的  $R^2$  越大，相关部分 S 越大，符合规律。

3、无论是自变量 Y，因变量 X，还是自变量 X，因变量 Y，所计算得到的相关部分 S 的值是相同的，因为调整后的  $R^2$  值相等。

4、之所以依据 X 和 Y 中较小的进行修正，是为了防止极端情况下剔除相关部分后指标出现负值。

## 2.2 多个评价指标相关情况下的评价修正

现在考虑多个评价指标的情况，假设有  $X_1, X_2, \dots, X_n$  个指标， $c_1, c_2, \dots, c_n$  为权重，显然有  $c_1 + c_2 + \dots + c_n = 1$ ，则评价值为：

$$P = c_1 X_1 + c_2 X_2 + \dots + c_n X_n \quad (8)$$

多指标评价时指标间的相关关系相当复杂，一般一个指标与一个或数个其他指标相关，也有可能某个指标与其他所有指标都不相关。如图 1 所示。假设有 6 个指标， $X_1$  与  $X_3$ 、 $X_4$  相关， $X_3$  与  $X_1$ 、 $X_2$ 、 $X_5$  相关， $X_5$  与  $X_3$  相关，而  $X_6$  独立，与其他所有指标都不相关。

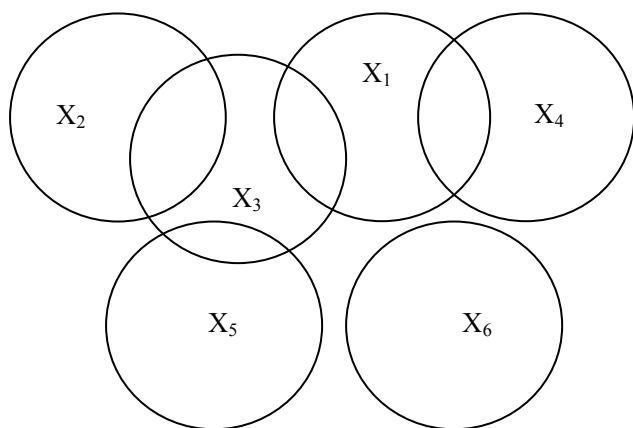


图 1 多指标相关

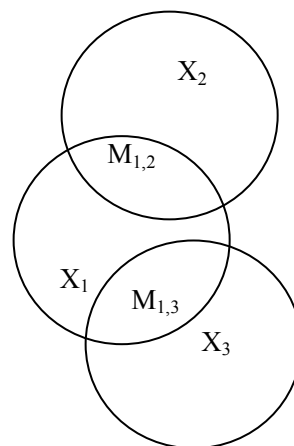


图 2 三指标相关

可以很容易将 2 个指标相关的情况推广到多个指标。在多指标的情况下，如果考虑变量间的相关，则总评分值会降低，评价排序结果与不考虑相关相比也有可能发生变化。

但是在估算相关部分时出现了新的问题，比如  $X_1$  与  $X_3$ 、 $X_4$  相关，相关部分就有两块，即使可以估算出总相关值，那么这两部分如何进行分配？又如  $X_3$  与  $X_1$ 、 $X_2$ 、 $X_5$  相关，相关部分就有 3 块，即使估算出相关部分，那么这三部分如何分配？增加了计算的复杂性与不确定性。

如图 2 所示，假设有 3 个指标， $X_1$  和  $X_2$ 、 $X_3$  相关，但  $X_2$ 、 $X_3$  不相关。 $X_1$  与  $X_3$ 、 $X_4$  相关部分的近似值为：

$$S_{1,23} = r_{1,23} \min(X_1, X_2, X_3) \quad (9)$$

若以  $X_3$  为自变量， $X_1$  为因变量进行回归，则  $X_3$  和  $X_1$  相关部分为：

$$S_{1,3} = r_{1,3} \min(X_1, X_3) \quad (10)$$

同理，则  $X_2$  和  $X_1$  相关部分为：

$$S_{1,2} = r_{1,2} \min(X_1, X_2) \quad (11)$$

理论上, 应该有:

$$S_{1,23} = S_{1,2} + S_{1,3} \quad (12)$$

由于公式(7)仅是一种估算方法, 因此公式(12)是不成立的。也就是说, 我们无法在多指标相关的情况下计算出详尽的“重叠块”。

必须另辟蹊径解决这个问题, 只能通过变量之间的两两相关来进行解决, 当然变量是否两两相关必须通过统计检验。假设有  $n$  个指标, 则最大有  $n(n-1)/2$  个“重叠块”。相关部分采用公式(7)估算出来以后, 有两种方法进行指标相关修正:

$$x_i' = x_i - 0.5s_{i,j} \quad x_j' = x_j - 0.5s_{i,j} \quad (13)$$

$$x_i' = x_i \quad x_j' = x_j - s_{i,j} \quad (\omega_i > \omega_j) \quad (14)$$

$x_i'$ 和  $x_j'$ 表示调整后的指标,  $\omega_i$ 、 $\omega_j$ 表示权重。公式(13)是在不知道指标权重的情况下, 可以用两个指标各减去二分之一重叠部分进行修正, 称为客观相关系数调整法; 第二种是在已经确定指标权重的情况下, 权重较大指标不剔除重叠部分, 而在权重较小的指标中减去重叠部分, 因为根据常识, 在权重较大的指标中剔除重叠部分相当于降低了权重较大指标的权重, 显然不符合常识, 该方法称为权重相关系数调整法。

客观相关系数调整法和权重相关系数调整法统称为相关系数调整法。数据修正完毕后在进行加权汇总前还必须再做一次标准化。

为了衡量相关系数调整法的调整结果, 定义指标相关优化系数:

$$I = 1 - \frac{\sum R'}{\sum R} \quad (15)$$

其中 $\sum R'$ 为调整后的相关系数和,  $\sum R$ 为调整前的相关系数和, 指标相关优化系数  $I$  表示指标相关的改善情况。

### 3 数据

本文数据来自于中国科学技术信息研究所 CSTPC 数据库, 本文以医学类期刊为例进行分析。中国科学技术信息研究所从 1987 年开始对中国科技人员在国内外发表论文数量和被引情况进行统计分析, 并利用统计数据建立了中国科技论文与引文数据库, 同时出版《中国科技期刊引证报告》。本文数据是 2006 年的医学类数据, 来自 136 个医学期刊和医学院校学报, 选取的指标有总被引频次、被引半衰期、影响因子、即年指标、基金论文比 5 个指标。本文采取专家打分法确定权重。原始数据必需进行标准化处理, 本文设定每项指标最大值为 100, 其他数据分别与最大值相除后得到各自标准化后的结果。数据描述统计量及权重如表 1 所示。

表 1 评价指标及描述统计量

变量名称	变量含义	权重	均值	最大值	最小值	标准差
$X_1$	总被引频次	0.50	13.06	100.00	1.07	12.81
$X_2$	被引半衰期	0.05	69.65	100.00	31.07	15.81
$X_3$	影响因子	0.30	23.74	100.00	5.69	15.16
$X_4$	即年指标	0.05	14.48	100.00	0.30	14.48
$X_5$	基金论文比	0.10	36.87	100.00	2.06	24.04

n	期刊数量	136
---	------	-----

## 4 实证结果

首先画出变量二维表，计算两两相关系数，注意必须在 5%的水平上通过统计检验，否则不能认为两个变量间存在相关关系。计算时采用 EVIEWS5.0 软件，结果如表 2 所示。

表 2 原始指标两两相关系数

	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>
X <sub>1</sub>	--		0.42	0.15	--
X <sub>2</sub>		--		0.02	0.18
X <sub>3</sub>			--	0.58	0.07
X <sub>4</sub>				--	0.03
X <sub>5</sub>					--

根据表 2 的结果，X<sub>1</sub> 和 X<sub>3</sub>、X<sub>4</sub> 相关，X<sub>2</sub> 和 X<sub>4</sub>、X<sub>5</sub> 相关，X<sub>3</sub> 和 X<sub>4</sub>、X<sub>5</sub> 相关，X<sub>4</sub> 和 X<sub>5</sub> 相关，也就是说，对于每种期刊而言，由于指标间相关导致的重叠块共有 7 块，分别为 S<sub>1,3</sub>、S<sub>1,4</sub>、S<sub>2,4</sub>、S<sub>2,5</sub>、S<sub>3,4</sub>、S<sub>3,5</sub>、S<sub>4,5</sub>。根据公式 (7) 进行计算。

下面结合权重来对指标值进行修正：

1. X<sub>1</sub> 和 X<sub>3</sub>、X<sub>4</sub> 相关，权重分别为 0.5、0.3、0.05，根据在权重较低指标中剔除重叠块的原则，S<sub>1,3</sub> 应该在 X<sub>3</sub> 中消去，S<sub>1,4</sub> 应该在 X<sub>4</sub> 中消去。

2. X<sub>2</sub> 和 X<sub>4</sub>、X<sub>5</sub> 相关，权重分别为 0.05、0.05、0.1，由于 X<sub>2</sub> 和 X<sub>4</sub> 权重相等，因此放在哪边消去都可以，但只能减一次。S<sub>2,4</sub> 可在 X<sub>2</sub> 中消去，S<sub>2,5</sub> 应在 X<sub>2</sub> 中消去。

3. X<sub>3</sub> 和 X<sub>4</sub>、X<sub>5</sub> 相关，权重分别为 0.3、0.05、0.1，S<sub>3,4</sub> 应该在 X<sub>4</sub> 中消去、S<sub>3,5</sub> 应在 X<sub>5</sub> 中消去。

4. X<sub>4</sub> 和 X<sub>5</sub> 相关，权重分别为 0.05、0.1，S<sub>4,5</sub> 应该在 X<sub>4</sub> 中消去。

综上所述，对原指标值做如下调整：

X<sub>1</sub> 不做任何调整，其权重最大；X<sub>2</sub> 减去 S<sub>2,5</sub>、S<sub>2,4</sub>；X<sub>3</sub> 减去 S<sub>1,3</sub>；X<sub>4</sub> 减去 S<sub>1,4</sub>、S<sub>3,4</sub>、S<sub>4,5</sub>；X<sub>5</sub> 减去 S<sub>3,5</sub>。为了检验相关关系消除结果，再计算一次相关系数，如表 3 所示。

表 3 权重相关系数调整前后的相关系数

	x <sub>1</sub> '	x <sub>2</sub> '	x <sub>3</sub> ' 及减少值	x <sub>4</sub> ' 及减少值	x <sub>5</sub> ' 及减少值
x <sub>1</sub> '	--		0.21 -0.21	0.04 -0.11	-- --
x <sub>2</sub> '		--	0.02 0.02	0.06 0.04	0.03 -0.15
x <sub>3</sub> ' 减少值			--	0.36 -0.22	0.07 --
x <sub>4</sub> ' 减少值				--	0.02 -0.01
x <sub>5</sub> ' 减少值					--

对比表 2 和表 3 的结果，可以看出，相关系数比较大指标，的经过调整后其相关系数显著降低，X<sub>2</sub> 和 X<sub>3</sub> 本来不相关，经调整后反而轻微相关，X<sub>2</sub> 和 X<sub>4</sub> 本来相关系数为 0.02，调整后略有升高，变成 0.06，仍然属于轻微相关的范畴。调整前总相关系数为 1.45，调整后总相关系数为 0.81，指标相关优化系数为 44%，因此采用相关系数调整法的效果是相当显著的。

最后再用调整后的各指标值经二次标准化后进行加权汇总，消去相关因素后结果如表 4 所示（前 40 期刊）。从表 4 的结果看，消去相关因素后排序有所变化。

表 4 消去相关因素前后得分与排序对比

刊名	调整前 得分	调整前 排序	调整后 得分	调整后 排序
中华医学杂志	78.56	1	72.66	1
中华医院管理杂志	63.91	2	63.97	2
中国危重病急救医学	62.58	3	60.87	3
中国现代医学杂志	46.64	4	43.56	4
中华麻醉学杂志	42.94	5	41.28	5
中国病理生理杂志	42.25	6	38.66	8
CHINESE MEDICAL JOURNAL	41.92	7	38.79	7
中华血液学杂志	41.11	8	39.33	6
第四军医大学学报	40.25	9	37.57	9
第三军医大学学报	35.15	10	32.40	17
中华病理学杂志	34.76	11	34.46	12
细胞与分子免疫学杂志	34.49	12	35.33	11
北京大学学报医学版	34.20	13	34.05	14
中华医学遗传学杂志	34.12	14	34.24	13
中国疫苗和免疫	33.73	15	37.03	10
生理学报	32.79	16	33.13	16
中国寄生虫学与寄生虫病杂志	32.13	17	33.63	15
医学研究生学报	30.84	18	31.03	19
解剖学报	30.81	19	30.85	21
吉林大学学报医学版	30.58	20	29.29	27
中华高血压杂志	30.41	21	31.72	18
中华男科学杂志	30.39	22	30.66	23
中华微生物学和免疫学杂志	30.36	23	29.27	28
中山大学学报医学科学版	30.24	24	30.76	22
南方医科大学学报	29.84	25	27.80	32
解放军医学杂志	29.79	26	27.79	33
免疫学杂志	29.55	27	30.05	26
第二军医大学学报	29.46	28	27.66	34
中国医学科学院学报	28.73	29	28.46	30
中国人兽共患病学报	28.67	30	27.13	35
航天医学与医学工程	28.55	31	30.27	25
浙江大学学报医学版	28.33	32	30.86	20
生理科学进展	28.28	33	28.56	29
中国医院管理	28.01	34	27.89	31
中国急救医学	27.90	35	25.97	40
中国免疫学杂志	27.26	36	26.19	38
ASIAN JOURNAL OF ANDROLOGY	27.03	37	30.36	24
基础医学与临床	26.87	38	26.93	36
医疗设备信息	26.62	39	26.61	37
现代免疫学	26.02	40	26.03	39

为了便于比较，表 6 同时给出了不根据权重调整后的指标两两相关系数。对比表 2 和表 6 的结果，可以看出，相关系数比较大的指标，经过调整后相关系数显著降低， $X_2$  和

X<sub>4</sub> 本来相关系数为 0.02，调整后略有升高，变成 0.04，仍然属于轻微相关的范畴。调整前总相关系数为 1.45，调整后总相关系数为 0.92，指标相关优化系数优化率为 37%，因此采用客观相关系数调整法的效果同样是相当显著的。

表 6 客观相关系数调整前后的相关系数

	x <sub>1</sub> '	x <sub>2</sub> '	x <sub>3</sub> ' 及减少值	x <sub>4</sub> ' 及减少值	x <sub>5</sub> ' 及减少值			
x <sub>1</sub> '		--	0.27	-0.15	0.08	-0.07	--	--
x <sub>2</sub> '					0.04	0.02	0.09	-0.09
x <sub>3</sub> ' 减少值					0.34	-0.24	0.07	--
x <sub>4</sub> ' 减少值							0.03	--
x <sub>5</sub> ' 减少值								

## 5 结论

指标间相关问题不仅仅存在于科技评价中，而且广泛存在于一切多指标综合评价中，采用相关系数调整法可以对这个问题进行优化，但是完全消除是十分困难的。

1、作为一种近似算法，相关系数调整法可以显著降低由于指标相关产生的重复计算问题，使评价结果更贴近客观实际，同时修正了由于数据冗余导致的排序错误。采用本方法的前提是，各项指标都是正向指标。

2、采用相关系数调整法前必须对数据进行标准化处理，指标重复部分剔除后还要再进行一次数据标准化处理，然后才能进行加权汇总。

3、经过调整后，指标相关情况虽然有所降低，由于这是一种近似算法，仍然会存在相关，不建议继续循环采用本方法降低相关系数，这样会加大误差。

4、使用相关系数调整法进行评价数据相关优化，即可以在指标确定权重的情况下进行，也可以直接对原始数据进行客观公正的相关修正。

5、由于科技评价的复杂性，一般采用多级指标进行评价，在指标众多的情况下，最好逐步进行调整。

6、采用相关系数调整法优化后，如果最大指标间的两两相关系数降到 0.5 以下，在进行进一步评价时，建议取消旨在消除指标相关性的客观赋权法，如主成分分析，因子分析等。

## 参考文献

- [1]王伟夏. 消除评价中重复计算的方法研究[J]. 系统工程, 1988, 6(4):24-29
- [2]黄擎明. 技术经济分析[M]. 北京: 科学出版社, 1991. 12
- [3]王庆石. 统计指标间信息重叠的消减方法[J]. 财经问题研究, 1994. 1
- [4]彭勇行, 张茂祥. 系统评价理论与方法[M]. 北京: 中国经济出版社, 1994
- [5]徐祥发, 肖人彬. 评价指标相关性的消除方法研究, 系统工程理论与实践, 2002 (12) :1-5