

学术期刊多属性评价中指标相关关系修正研究¹

俞立平 潘云涛 武夷山

摘要：为了克服学术期刊评价中指标相关带来的重复计算问题，本文提出了一种新的降低指标相关关系的近似算法——综合回归调整法，其原理是首先全面分析所有的指标间复杂的相关关系，绘制出指标相关图，然后利用回归系数和拟合优度估算出相关部分，最后进行指标调整，在此基础上再进行评价。为了减少误差，综合回归调整法只能使用一次，并且在指标较多的情况下要结合指标分级使用。

关键词：学术期刊 多属性评价 指标相关 综合回归调整法

中图分类号：G204

1 引言

期刊评价是文献计量学研究的重要组成部分，它通过对学术期刊的发展规律和增长趋势进行量化分析，揭示学科文献数量在期刊中的分布规律，为优化学术期刊的使用提供重要参考，同时可以提高学术期刊的内在质量，促进学术期刊的健康成长和发展。美国著名情报学家加菲尔德博士在 20 世纪 60 年代对期刊文献的引文进行了大规模统计分析，得到了大量被引用文献集中在少数期刊上，而少量被引用文献散布在大量期刊中的结论，这可以被认为是国外期刊评价理论的起源。

学术期刊评价主要有两大类方法，一种是直接根据期刊文献计量学指标进行对比，二是采用多属性综合评价方法，该方法克服了单指标评价容易带来的片面性，比较适合期刊的综合评价，因此得到了较广泛的应用。但是由于科技评价工作的复杂性，学术期刊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。

指标相关问题是统计学和多属性评价（决策）的基本问题，不仅仅是学术期刊评价中存在的问题。一些学者在指标相关关系调整方面做了一些研究。王伟夏^[1]（1988）从集合论的角度分析了指标之间存在重迭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。黄擎明^[2]（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。王庆石^[3]（1994）提出消除重叠指标的复相关系数法和多元回归法。彭勇行、张茂祥^[4]（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。徐祥发、肖人彬^[5]（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。俞立平（2006）^[6]在通过统计检验后的两两相关系数的基础上提出了一种新的评价指标相关性消除方法——相关系数调整法。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。从目前的研究来看，尚有以下问题没有得到解决：

1、采用指标间两两相关系数来进行调整是不可靠的。原因有两个：第一，某些指标之间即使相关系数较高，但不一定能够通过统计检验，没有通过统计检验的相关其实是不相关

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，汉族，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。

的。第二：既然是针对某个领域的评价，评价指标间往往存在错综复杂的相关关系，也就是说存在多指标之间的复杂的相关关系，这已经超越了指标两两相关的范畴，必须从系统的角度全面进行分析。正如天文学中存在的三体问题，并非分别解决两个双体问题就能解决三体问题，局部问题的解决并不意味着整体问题的解决。

2、采取删除部分相关程度较高指标的做法容易丢失重要信息。尤其在评价对象较多的情况下，删除指标对评价结果的影响很大。

3、主成分分析和因子分析能够在若干相关指标中筛选出主成分或者关键因子，有其优越性，但它们是完全客观的评价方法，没有考虑评价者的主观权重，不够全面。

本文利用中国科学技术信息研究所的医学院校学报数据，提出一种新的指标相关关系调整方法——综合回归调整法（IRA, Integrative Regression Adjustment）。其原理是，首先分析指标间错综复杂的关系，绘出指标相关关系图，在此基础上，根据回归系数和拟合优度估算出指标间的相关部分并进行消减，最后验证指标相关关系调整后的结果。本文同时给出指标相关关系调整前后评价结果的比较，最后对多属性评价指标相关问题进行总结和讨论。

2 综合回归调整法原理

综合回归调整法的本质在于找出指标体系中所有指标之间的相关关系，经估算后进行调整，其步骤如下：

1、首先搜集整理数据，并且对数据进行标准化。

2、假设有 n 个指标，分别为 X_1 、 X_2 …… X_n ，分别做 X_i 与其它所有指标的回归，然后进行统计检验，依次删掉统计检验最不显著的指标。考虑到多重共线性问题，如果某些指标的回归系数为负，无法从理论上加以解释，因此，也删掉回归系数为负的指标。

3、根据指标间的相关关系，对其进行总结，画出指标体系的完整相关图，在指标较多的情况下，这项工作可能比较复杂。

4、根据指标回归方程，用拟合值乘以拟合优度 R^2 得到相关部分的近似值，因为拟合值本身代表了相关部分的总体效果，拟合优度代表的拟合的“质量”。极端情况下，当拟合优度为 0 时，相关部分为 0，说明指标间不相关；当拟合优度为 1 时，相关部分等于拟合值，说明完全相关。

对于任一比较显著的回归方程，接下来再根据回归系数的大小按比例进行分配，得到相关部分的重叠值。因为回归方程中回归系数的大小反映了不同自变量指标对因变量指标的解释程度，类似于“权重”。

需要注意的是，如果某个重叠部分牵涉到两个或两个以上的方程，那么该重叠部分的近似值是两个或两个以上方程估算结果的平均值。

5、对指标数据进行调整，以降低相关性。方法是用各指标值减去重叠值的一半，即重叠部分均匀分配到两个相关指标中，这也符合相关的本质含义。之所以用该方法进行调整，因为这是一种“中庸”的处理方法，如果将重叠部分从一个指标中去除，另外一个指标保持不变，在拟合优度和回归系数较小的情况下（即重叠部分较小的情况下），这是没有问题的。然而在拟合优度和回归系数较大时（即重叠部分较大的情况下），如果仅从一个指标中减去重叠信息就会失去该指标中包含的许多重要信息。

此外，还有一种极端情况，在数据标准化后，由于不同评价对象之间指标值相差较大，在重叠部分较高的情况下，如果仅从一个指标中去除重叠部分，某些评价对象该指标的值可能为负，这是要尽力避免的。

6、比较指标值调整前后所有回归方程拟合优度的变化，作为衡量指标相关水平有没有降低的标准。

3、变量和数据

本文数据来自于中国科学技术信息研究所 CSTPC 数据库，以医学院校学报为例进行分析，共 60 种。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《中国学术期刊引证报告》。作为一种算例，考虑到简捷性，本文选取总被引频次、扩散因子、影响因子、基金论文比、即年指标、被引半衰期 6 个指标，表 1 为数据的描述统计量。

表 1 变量描述统计量

变量名	内容	均值	最大值	最小值	标准差
X1	总被引频次	500.433	2257.000	114.000	389.021
X2	扩散因子	48.578	75.440	19.370	14.009
X3	影响因子	0.266	0.640	0.076	0.142
X4	基金论文比	0.381	0.970	0.060	0.184
X5	即年指标	0.037	0.163	0.004	0.031
X6	被引半衰期	4.701	6.560	2.630	0.785
n	期刊数量		60		

数据来源：中国科学技术信息研究所 CSTPC
数据库 2006 年数据

由于要对期刊进行评价，所以必需对数据进行标准化处理，每项指标最大值设为 100，然后按比例进行调整。此外，被引半衰期是反向指标，必须进行适当处理，方法是用 100 减去其标准化后的结果后再做标准化，这种处理方式是线性处理方式，比反向指标取倒数的非线性处理方式更为科学。

4 实证结果

4.1 指标间相关关系的调整

首先利用 Eviews6.0 采用多元回归对学术期刊评价指标间的错综复杂关系分别进行回归，结果如下：

$$X1 = 2.617 + 0.470X3 \quad (1)$$

(5.811) $R^2=0.368$

$$X3 = 5.482 + 0.502X1 + 0.423X4 + 0.295X5 \quad (2)$$

(4.802) (3.902) (3.580) $R^2=0.682$

$$X4 = 15.610 + 0.569X3 \quad (3)$$

(6.849) $R^2=0.447$

$$X5 = -34.877 + 0.615X3 + 0.656X6 \quad (4)$$

(5.620) (2.842) $R^2=0.475$

$$X6 = 52.335 + 0.185X5 \quad (5)$$

(3.625) $R^2=0.185$

根据以上回归结果，绘出指标相关图，如图 1 所示。扩散因子 X2 相对独立，与所有指标都不相关；影响因子 X3 同时和总被引频次 X1、基金论文比 X4、即年指标 X5 相关，其相关部分分别为 S1、S3、S2；即年指标 X5 与影响因子 X3 和被引半衰期 X6 相关，相关部分分别为 S2、S4。也就是说，相关部分 S2 在两个方程中同时出现。

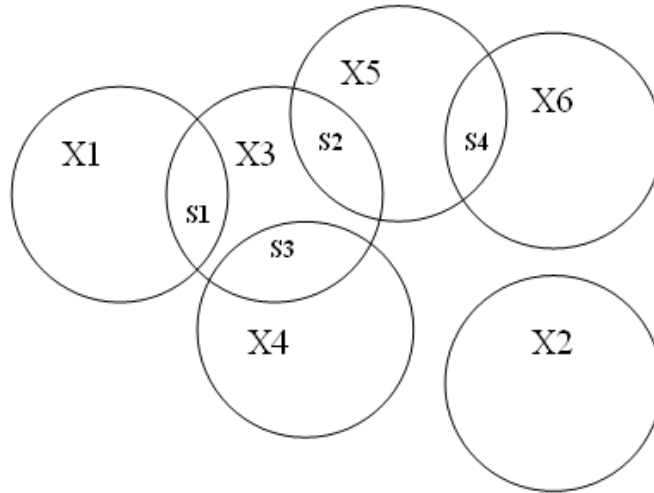


图 1 期刊评价指标相关图

利用方程 2 的拟合值乘以拟合优度 R 得到相关部分 S1+S2+S3 的汇总值，然后根据 X1、X5、X4 回归系数（0.502、0.295、0.423）按比例进行分配，得到重叠部分 S1、S2、S3 的估算值。

同理，利用方程 4 的拟合值乘以拟合优度 R 得到相关部分 S2+S4 的汇总值，然后根据 X3、X6 回归系数（0.615、0.656）按比例进行分配，得到重叠部分 S2、S4 的估算值。

由于 S2 牵涉到两个方程，因此，必须对 S2 进行调整，方法是将以上两个 S2 的计算结果相加后求平均值，得到 S2 的最终估算值 S2'。

对各指标进行调整，分别减掉相关部分的一半，过程如下：

$$X_{11} = X_1 - 0.5S_1$$

$$X_{33} = X_3 - 0.5(S_1 + S_2' + 0.5S_3)$$

$$X_{44} = X_4 - 0.5S_3$$

$$X_{55} = X_5 - 0.5(S_2' + S_4)$$

$$X_{66} = X_6 - 0.5S_4$$

只有 X2 保持不变，调整后重新进行回归分析，相关结果如下：

$$X_{11} = 4.752 + 0.421X_{33} \quad (4.038) \quad R^2=0.219 \quad (1)$$

$$X_{33} = 4.442 + 0.406X_{11} + 0.367X_{44} + 0.188X_{55} \quad (3.719) \quad (3.368) \quad (2.211) \quad R^2=0.482 \quad (2)$$

$$X_{44} = 14.209 + 0.550X_{33} \quad (4.989) \quad R^2=0.300 \quad (3)$$

$$X_{55} = -29.104 + 0.367X_{33} + 0.647X_{66} \quad (3.970) \quad (2.467) \quad R^2=0.284 \quad (4)$$

$$X_{66} = 51.330 + 0.126X_{55} \quad (2.346) \quad R^2=0.087 \quad (5)$$

指标相关部分调整后所有变量都通过了统计检验，并且所有拟合优度都得到了降低。调整前，拟合优度汇总为 0.368+0.682+0.447+0.475+0.185=2.157，调整后拟合优度汇总为 0.219+0.482+0.300+0.284+0.087=1.372，降低了 36.4%，说明指标相关问题调整效果是比较显著的。

4.2 指标相关关系调整前后评价结果的比较

为了比较指标相关关系调整前后评价结果及其排序变化，采用专家赋权法，依次给总

被引频次、扩散因子、影响因子、基金论文比、即年指标、被引半衰期 6 个指标赋权为 0.35、0.1、0.25、0、1、0.1、0.1。由于指标相关关系调整后指标值发生了变化，因此需要重新进行标准化，然后再进行加权汇总。

表 2 为指标调整前后的评价及排序情况。相关系数调整前后评价总得分之间存在极高的相关关系， R^2 为 0.991，说明二者是一致的，即指标相关系数调整后并没有改变评价得分的总体状况，因此采用综合回归调整法来降低指标间的相关关系是可行的。从排序看，部分期刊的排序位次发生了变化，这就是降低指标间相关关系后带来的结果。

表 2 指标相关系数调整前后评价结果的比较

期刊名称	总被引 频 次	扩 散 因 子	影 响 因 子	基 金 论 文 比	即 年 指 标	被 引 半 衰 期	调 整 前 得 分 及 排 序	调 整 后 得 分 排 序		
第四军医大学学报	428	46.96	0.25	0.48	0.024	4.91	67.64	1	62.67	1
医学研究生学报	335	36.72	0.187	0.34	0.02	5.62	62.68	2	57.20	2
第三军医大学学报	284	44.72	0.192	0.12	0.036	3.72	61.72	3	56.90	3
北京大学学报医学版	954	40.57	0.64	0.56	0.042	5.58	58.23	5	55.59	4
中国药科大学学报	924	19.37	0.406	0.64	0.069	6.12	55.55	6	52.64	5
浙江大学学报医学版	245	47.76	0.205	0.56	0.058	5.58	59.35	4	51.87	6
中山大学学报医学科学版	182	71.43	0.125	0.06	0.014	5.17	54.04	7	49.87	7
中国医学科学院学报	1347	34.89	0.286	0.45	0.048	4.92	53.95	9	49.63	8
南方医科大学学报	1767	25.58	0.428	0.41	0.042	3.55	53.96	8	49.35	9
第二军医大学学报	2257	23.35	0.438	0.33	0.045	4.19	49.61	11	44.67	10
吉林大学学报医学版	213	70.42	0.24	0.32	0.029	4.06	51.24	10	44.44	11
北京中医药大学学报	322	56.83	0.303	0.44	0.036	4.21	47.88	12	42.79	12
沈阳药科大学学报	650	50.77	0.316	0.48	0.036	5.8	44.55	13	42.68	13
西安交通大学学报医学版	316	52.85	0.137	0.21	0.018	4.57	43.81	14	38.90	14
四川大学学报医学版	509	47.35	0.107	0.32	0.004	4.95	42.45	16	38.78	15
南京医科大学学报自然科学版	178	67.98	0.161	0.28	0.017	4.75	41.57	17	38.61	16
复旦学报医学版	420	30.48	0.341	0.37	0.038	4.93	40.84	19	37.56	17
中南大学学报医学版	239	60.25	0.122	0.28	0.029	4.64	39.81	20	36.73	18
郑州大学学报医学版	142	54.23	0.076	0.07	0.01	5.86	43.65	15	36.45	19
上海中医药大学学报	340	60	0.212	0.28	0.016	4.85	41.03	18	35.45	20
华中科技大学学报医学版	244	70.9	0.167	0.16	0.017	4.82	38.49	21	35.38	21
福建医科大学学报	301	37.54	0.244	0.45	0.019	4.64	37.35	22	34.46	22
上海交通大学学报医学版	598	49.67	0.289	0.47	0.028	5.5	36.99	23	33.68	23
广州中医药大学学报	849	34.75	0.371	0.97	0.042	4.11	35.35	25	33.06	24
军医进修学院学报	226	59.73	0.141	0.33	0.011	4.06	35.98	24	32.61	25
武汉大学学报医学版	267	62.17	0.114	0.15	0.017	3.91	34.17	28	32.14	26
东南大学学报医学版	303	59.08	0.23	0.54	0.012	4.74	33.57	29	31.38	27
新乡医学院学报	364	48.08	0.264	0.32	0.069	4.3	34.31	27	31.36	28
安徽医科大学学报	167	69.46	0.13	0.28	0.023	4.81	34.40	26	31.30	29
重庆医科大学学报	1139	34.86	0.417	0.55	0.041	3.65	32.90	32	30.62	30
中国医科大学学报	635	39.21	0.352	0.37	0.045	4.24	33.29	31	30.41	31
军事医学科学院院刊	420	33.33	0.23	0.46	0.045	6.1	33.37	30	30.30	32
首都医科大学学报	236	62.29	0.178	0.23	0.007	4.67	31.16	35	29.47	33
哈尔滨医科大学学报	236	62.29	0.178	0.23	0.007	4.67	30.80	36	29.05	34

南京中医药大学学报	349	61.6	0.16	0.4	0.006	4.62	31.73	34	28.25	35
湖南中医药大学学报	387	31.01	0.185	0.3	0.03	6.01	30.65	37	27.99	36
成都中医药大学学报	369	58.27	0.163	0.23	0.008	4.84	32.18	33	27.81	37
山东大学学报医学版	545	52.29	0.238	0.38	0.039	4.38	30.01	38	27.46	38
苏州大学学报医学版	219	38.36	0.351	0.73	0.086	4.36	28.28	41	27.04	39
河北医科大学学报	685	30.51	0.474	0.22	0.069	5.4	27.85	43	26.71	40
蚌埠医学院学报	297	61.62	0.23	0.32	0.012	4.9	28.35	40	26.67	41
山西医科大学学报	716	45.67	0.317	0.5	0.034	4.64	28.11	42	26.57	42
温州医学院学报	418	62.92	0.144	0.11	0.013	4.46	29.26	39	26.22	43
青岛大学医学院学报	213	61.03	0.152	0.41	0.024	4.22	27.30	46	26.06	44
青岛大学医学院学报	331	58.01	0.263	0.37	0.009	3.92	27.30	47	26.06	45
广州医学院学报	114	75.44	0.083	0.1	0.008	4.08	27.53	44	25.72	46
江苏大学学报医学版	522	45.02	0.34	0.56	0.08	4.55	27.48	45	25.17	47
广西医科大学学报	342	30.12	0.276	0.24	0.063	3.33	27.21	48	24.70	48
安徽中医学院学报	207	66.67	0.099	0.22	0.012	5.23	26.78	50	24.57	49
昆明医学院学报	791	24.53	0.626	0.56	0.113	2.63	26.64	52	24.51	50
江西医学院学报	135	54.81	0.134	0.27	0.017	6.56	26.15	54	24.48	51
广东药学院学报	368	49.46	0.547	0.77	0.163	3.38	26.28	53	24.46	52
山东中医药大学学报	398	38.44	0.11	0.15	0.019	5.4	26.65	51	24.35	53
贵阳医学院学报	698	33.24	0.208	0.49	0.128	3.56	26.79	49	24.29	54
大连医科大学学报	987	27.56	0.603	0.55	0.045	5.71	23.74	56	23.22	55
徐州医学院学报	487	52.77	0.217	0.48	0.011	5.66	23.95	55	22.34	56
武警医学院学报	782	45.91	0.498	0.48	0.094	4.72	22.98	58	22.00	57
浙江中医学院学报	658	47.26	0.302	0.43	0.032	5.11	23.15	57	21.53	58
云南中医学院学报	581	44.23	0.568	0.79	0.046	4.37	22.43	59	20.89	59
贵阳中医学院学报	390	52.05	0.224	0.31	0.018	3.81	18.18	60	17.62	60

5 结论与讨论

5.1 综合回归调整法是一种有效的降低指标相关关系的方法

综合回归调整法是一种有效的指标相关关系降低方法,能够较大限度地降低多属性评价中指标间相关带来的重复计算问题,使评价结果更加客观和接近真实。实际应用过程中,为了简化计算,对于回归拟合优度较低的指标间相关关系可以不调整,一般而言, R^2 的阈值可以设定为0.4,即当 R^2 低于0.4时可以不调整,因为在指标众多的情况下,相关关系可能极为复杂,靠人工绘制相关图进而进行调整的工作量巨大。

5.2 指标分级是综合回归调整法的重要辅助手段

在指标众多的情况下,某类指标可能较多,由于多重共线性,会导致某些指标的回归系数为负,从而无法进行调整,在这种情况下,建议将指标分级后分批进行处理。比如分成两级指标,对于二级指标,当其数量较少时,可以采用综合回归调整法降低指标相关性后再进行评价,当二级指标数量较多时,建议直接采取指标赋权评价,不需要考虑相关关系的调整,因为此时多重共线性问题会很严重,回归系数为负的指标较多,采用综合回归调整法反而误差较大。对于一级指标,完全可以采取综合回归调整法来降低指标间的相关关系后再进行评价。

5.3 综合回归调整法只能应用一次

由于综合回归调整法只是一种近似算法,虽然有效,但一般只能用一次,因为循环使

用虽然可以减少指标间的相关关系，但是容易带来误差的放大，反而使评价结果出现失真。

5.4 综合回归调整法是一种数据预处理技术

综合回归调整法是一种数据预处理技术，在此基础上可以用其他一些多属性评价方法进行评价。与旨在消除指标相关关系的一些评价方法如主成分分析、因子分析相比，综合回归调整法可以结合评价者主观意愿进行评价，当然其处理结果就不能再用主成分分析、因子分析进行评价了，因为相关关系大大降低后不符合主成分分析和因子分析的条件了。当然，主成分分析和因子分析也有其自身的优势，它们可以在分析关键因子中发挥积极的作用。

综合回归调整法不仅仅适用于学术期刊评价，也适用于所有的其它领域的多属性评价。

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): 57-61
- [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
- [6]俞立平, 潘云涛, 武夷山. 科技评价指标相关消除方法——相关系数调整法[J]. 情报杂志, 2009 (2): 3-6