

科技评价中效用函数合成方法的比较研究¹

俞立平 潘云涛 武夷山

摘要：本文以科技期刊评价为例，采用专家打分法确定权重，重点分析加法合成、乘法合成、对数合成、代换合成方法的特点及适用范围。首先比较不同合成方法结果的相关性及排序情况，然后利用离散系数、极差比率、高端区分度、低端区分度、高低端汇总比率指标比较各种合成方法的区分度、对较优期刊及较差期刊打分的强化情况，最后分析不同乘法合成方法与科技期刊协调发展程度的关系。认为不同合成方法排序结果不尽相同，应根据需要选择不同的合成方法，各种合成方法综合运用是较好的解决问题的方法。

摘要：评价 科技期刊 效用函数 合成方法

1 引言

科技评价包括科技项目评价、科研机构评价、科技期刊评价、科技人员评价等诸多方面。科技评价是科技管理工作的重要组成部分，国家对科技评价工作十分重视，科技部联合五部委，以国科发基字[2003]142号文的名义发布了《关于改进科学技术评价工作的决定》，认为“要优化评价程序，改进评价方法，注重评价实效”。目前国内外多指标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，本文将其分为两大类（如图1所示）：第一类是效用函数合成类评价，一般采取主观或客观评价法确定权重，然后对无量纲的数据进行加权汇总，如专家会议法、德尔菲法、熵权法等等；第二类是系统评价方法，其特点是综合运用运筹学、模糊数学、多元统计、系统工程等领域的方法进行评价，评价结果不是根据标准化指标值进行加权汇总，而是采取各种不同的算法，评价中可能用到权重，也可能不用权重。如 ELECTRE 法、模糊综合评价法、突变理论、数据包络分析等等。总体上，效用函数综合评价方法仍然是现实生活中用得较多的评价方法。

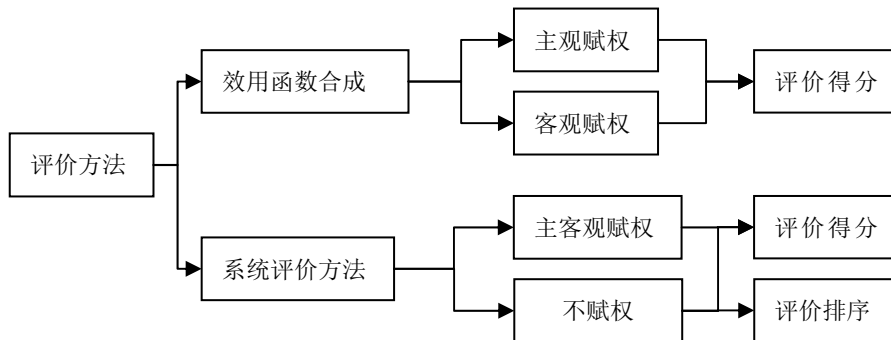


图1 评价方法分类体系

关于效用函数的合成方法，统计学上有加法合成、乘法合成、代换合成等等。美国密特公司在美国环境质量委员会委托下研究密特大气质量指数（MAQI），它以五项污染物为参数，采用简单平方平均合成公式。邱东（1991）^[1]从方法论上分析无量纲法与合成方法的特点，并比较了效用函数综合评价与其他系统评价方法的关系，认为乘法合成的区分度要大于加法合成。苏为华（2002）^[2]系统研究了幂函数合成中不同幂次情况下评价结果的平均值的变化以及补偿关系，认为幂次在-1~3之间比较合适，幂次越低，越是惩罚落后指标，幂次越高，越是奖励先进指标。张田玉（1987）^[3]在学校教育评价中结合加法合成与乘法合成的

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

作者简介：俞立平（1967-）男，江苏泰州人，博士，中国科学技术信息研究所博士后，扬州职业大学教授，主要从事信息经济、科学计量领域的教学科研工作。Email: chinayangzhou@yahoo.com.cn

特点,进行综合评价。田赤勇、陈进殿(2006)^[4]在对埋地输油管道的评价中用乘法合成对可靠性指标与估价指标进行组合。商红岩、宁宣熙(2005)^[5]对第三方物流企业绩效评价时采用了乘法合成。总体上,对各种合成方法的基本特点除了少数理论研究外,缺乏深入分析;从应用看,国内基本上以简单加权合成为主,对于其他形式合成方法的应用极少。

本文利用中国科学技术信息研究所的科技期刊原始数据,采用专家打分法确定权重,然后分别用不同合成方法进行合成,对结果进行比较,试图寻找不同合成方法的规律及适用范围,在此基础上进行进一步讨论。

2 研究方法

2.1 常见的几种效用函数合成方法

1. 加法类合成的几种函数形式

加法类合成的函数形式可以写成如下形式:

$$I = \left(\sum_{i=1}^n \omega_i x_i^k \right)^{\frac{1}{k}} \quad (1)$$

公式(1)中, $\omega_1, \omega_2, \dots, \omega_n$ 表示权重,为了研究方便,假设权重已经进行了标准化处理(下同), $\omega_1 + \omega_2 + \dots + \omega_n = 1$ 。 x_1, x_2, \dots, x_n 表示标准化处理后的指标值, k 表示幂次,取值一般为正整数或负整数。 I 为指标总得分。当 $k=1$ 时,公式(1)变为加权算术平均合成模型(简称加法合成):

$$I = \omega_1 x_1 + \omega_2 x_2 + \dots + \omega_n x_n \quad (2)$$

当 $k=-1$ 时,公式(1)变为加权调和平均合成模型(简称调和平均):

$$I = \frac{1}{\omega_1 \frac{1}{x_1} + \omega_2 \frac{1}{x_2} + \dots + \omega_n \frac{1}{x_n}} \quad (3)$$

当 $k=2$ 时,公式(1)变为加权平方平均合成模型(简称平方平均)。

$$I = \sqrt{\omega_1 x_1^2 + \omega_2 x_2^2 + \dots + \omega_n x_n^2} \quad (4)$$

理论上, k 可以取值为任意正整数和负整数,甚至小数,但考虑到人们的思维习惯和实际评价中的需要,一般以上三种函数形式就足以满足评价需要了。

2. 乘法合成模型

所谓乘法合成,其一般形式为:

$$I = x_1^{\omega_1} x_2^{\omega_2} \dots x_n^{\omega_n} \quad (5)$$

乘法合成比较注重指标全面发展,如果一个指标偏低,即使其他指标再高,总体评分值也不会太高。

3. 对数合成

将公式(5)取对数,变为:

$$\log(I) = \omega_1 \log(x_1) + \omega_2 \log(x_2) + \dots + \omega_n \log(x_n) \quad (6)$$

对数合成是乘法合成的基础上进行的转换,权重系数 ω_i 是弹性系数,表示当指标 x_i 提高 1% 将会导致总指标值提高 $\omega_i\%$,非常易于理解和运用。对数缩小了指标数据的大小,在一定程度上可以缩小异常值的影响。对数合成兼有乘法合成与加法合成的一些特点,但是在评

价中极少遇到，在经济计量学中有很广泛的用途。

很显然，对数合成和乘法合成的排序是完全一致的。

4. 代换合成

代换合成和乘法合成非常像，常见的代换合成一般形式（简称高端代换）为：

$$I = U - (U - x_1)^{\omega_1} (U - x_2)^{\omega_2} \cdots \cdots (U - x_n)^{\omega_n} \quad (7)$$

公式(7)中，U为单项指标的理论最高值（通常为100）。代换合成实际上是将指标进行简单处理后再进行乘法合成，很显然，只要任何一项指标达到最高值，评价总指标就为满分，因此，代换合成是一种“一美遮百丑”的合成方法，少数指标优异会大大弥补其他指标的不足。在现实科技评价中，这种情况极为罕见。

另外一种代换合成方法称为低端代换，其函数形式为：

$$I = (x_1 - L_1)^{\omega_1} (x_2 - L_2)^{\omega_2} \cdots \cdots (x_n - L_n)^{\omega_n} \quad (8)$$

公式(8)中， L_1, L_2, \dots, L_n 是各指标允许的取值下限。很显然，一旦某项指标达到最低下限，总指标值为0，因此，低端代换合成是一种“一丑遮百美”的合成方法，类似于一票否决制，在实践中有一定的意义，比如环境评价中，一旦某个指标超标，则总体环境质量很差。由于每个指标的最低容许值不确定，本文设定 L_1, L_2, \dots, L_n 为每个指标的极小值进行处理。

5. 几种合成方法的总结

本质上，调和平均、加法合成、平方平均都属于加法类合成，代换合成也是乘法合成，对数合成兼顾了乘法合成与加法合成的特点。加法合成通常被认为是指标间可以互相取长补短的一种合成方法，如果一个指标偏低，只要其他指标值较高，可以弥补单个指标偏低的差距。乘法合成通常被认为是要求各指标全面发展的一种合成方法，因为一个指标偏低，它和其他指标相乘后必然导致总指标值变小，当然，一个指标偏高，也会带动总指标值提高。它们的具体特性需要做进一步的研究。

2.2 几种合成方法的比较指标

1. 评价结果的相关系数

采用的不同的合成方法，有可能导致排序结果不一致，为此，我们采用各种合成方法评价结果的两两相关系数进行比较。

2. 离散系数与极差比率

离散系数是标准差与平均值的比值，借以反映不同合成方法的离散程度，离散程度越大，某种程度上说明评价结果数据的区分度较大，也就是说，由于评价对象比较接近而导致误判的可能性越小。在评价实践中，当评价指标和评价对象较多时，数据间距往往较小，导致评价结果之间的间距也很小，尤其是相邻的评价对象，其优势和弱势都不明显，而不同合成方法处理时这种间距是不同的，间距越大，说明区分度越好。

极差比率是最大值与最小值的差除以极大值，反映了评价结果的相对取值范围，因此也是区分度的一个指标。

3. 高低端区分度

分为高端区分度与低端区分度两个指标，用来进一步评价较好期刊和较差期刊的区分度。所谓高端区分度，指某种合成方法得分降序排列后最高20%评价对象的极差与总极差的比值，低端区分度是后20%评价对象的极差与总极差的比值，借以反映合成方法对较好评价对象的区分度以及较差评价对象的区分度。之所以用20%的评价对象，因为一般评价对象的分布呈正态分布，较好期刊与较差期刊都是少数，具体在计算时可根据评价对象的多少和数据分布情况进行适当调整。

4. 高低端汇总比率

由于不同合成方法的函数特点不一样,有些合成方法可能对较好的期刊打分较高,属于鼓励先进型;而有些合成方法可能对较差的期刊打分较高,属于鼓励落后型。例如在评价对象总体水平偏高时可能不需要鼓励先进,要鼓励落后。为了对这种情况进行综合衡量,本文采用高低端汇总比率指标。

本文采用前 20%期刊的评分之和与所有期刊评分总和的比值来衡量对较好期刊的强化(弱化)程度,采用后 20%期刊的评分之和与所有期刊评分总和的比值来衡量对较差期刊的强化(弱化)程度。二者之比称为高低端汇总比率,用来衡量总体强化(弱化)效果。

2.3 指标的协调性与合成方法的关系

一种科技期刊,如果其各项指标都能全面协调发展,固然较好。如果少数指标比较突出,可能也不见得是一件坏事。如果少数指标偏低,在某些情况下可能是不可以的。加法合成类方法认为指标间可以无限替代,较好指标的优势可以弥补较差指标的不足。乘法合成类方法强调指标协调发展,因为某项指标偏低会导致总体指标值降低。为了对乘法合成的这种特性进行检验,并比较不同乘法合成方式对指标协调性的强化程度,本文用某种期刊各指标的离散系数来衡量该期刊的协调性,离散系数越大,说明期刊各指标越不均衡。然后利用离散系数与乘法类的合成方法进行回归,分析各种不同类型的乘法合成方法对科技期刊协调发展的反映程度。

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,共 136 个医学期刊和大学医院学报,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。由于在科技期刊评价中指标选取和权重赋值存在一些争议,而本文重点研究科技期刊评价中的效用函数合成方法问题,因此指标选取暂且采用以上 5 个指标,采用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。

4、实证结果

4.1 指标权重的确定

本文重点是研究效用函数的合成方式,因此必须通过某种方法确定各指标的初始权重以便于深入研究,本文采取专家打分法,得到各指标的权重,依次为:总被引频次(0.18)、被引半衰期(0.17)、影响因子(0.23)、即年指标(0.18)、基金论文比(.24)。

4.2 不同合成方法的相关性及排序

将 7 种合成方法的结果计算出来后,计算它们之间的两两相关系数,结果如表 1 所示。总体上,各种合成方法之间的相关度较高,说明都能反映评价对象质量的差异,高端代换与调和平均的相关度最低,仅为 0.679,高端代换与乘法合成相关系数为 0.784,也不算高,这和理论分析是一致的。平方平均与调和平均虽然同属加法合成,但相关系数也不高,仅为 0.709。

表 1 不同合成方法结果的相关系数

	调和	加法	平方	乘法	对数	高端代换	低端代换
调和	1.000	0.862	0.709	0.972	0.918	0.679	0.966
加法	0.862	1.000	0.953	0.949	0.931	0.903	0.947
平方	0.709	0.953	1.000	0.823	0.811	0.945	0.836

乘法	0.972	0.949	0.823	1.000	0.964	0.784	0.992
对数	0.918	0.931	0.811	0.964	1.000	0.759	0.961
高端代换	0.679	0.903	0.945	0.784	0.759	1.000	0.791
低端代换	0.966	0.947	0.836	0.992	0.961	0.791	1.000

除了理论上乘法合成与对数合成排序一致外，不同合成方法结果的排序是不尽一致的，由于数据较多，本文仅公布加法合成前 20 个期刊结果，如表 2 所示。需要注意的是，高端代换由于其自身的特点，只要一个指标达到 100，总评价价值就是满分，所以表 2 中有 4 个期刊排名第一。

表 2 不同合成方法排序结果的比较（部分）

刊名	加法	调和	平方	乘法/对数	高端代换	低端代换
中国危重病急救医学	1	1	2	1	5	1
中华医院管理杂志	2	4	1	3	1	3
中华医学杂志	3	2	3	2	1	2
生理学报	4	10	5	6	8	6
中国寄生虫学与寄生虫病杂志	5	6	8	5	6	4
航天医学与医学工程	6	23	6	11	9	10
生理科学进展	7	39	4	26	1	22
中华血液学杂志	8	7	10	8	7	7
细胞与分子免疫学杂志	9	9	14	10	19	11
浙江大学学报医学版	10	28	13	13	21	16
中国病理生理杂志	11	8	19	9	23	9
中华医学遗传学杂志	12	5	22	7	25	8
解剖学报	13	33	9	22	11	19
吉林大学学报医学版	14	26	7	23	1	23
CHINESE MEDICAL JOURNAL	15	3	32	4	35	5
免疫学杂志	16	18	20	14	24	14
北京大学学报医学版	17	21	15	16	14	12
中山大学学报医学科学版	18	27	12	21	18	21
中国疫苗和免疫	19	16	24	12	28	13
现代免疫学	20	37	17	29	20	28

4.3 区分度与打分强化程度的比较

1. 区分度的分析

各种合成方法的基本统计指标、区分度类指标、打分强化指标的计算结果如表 3 所示。调和平均、加法合成、平方平均本质上都是加法合成，只不过幂次从-1、1、2 不断升高而已，随着幂次升高，离散系数和极差比率逐渐降低，评价对象之间的距离越来越近，区分度逐渐降低。乘法合成、高端代换与低端代换本质上都是乘法合成，无论是离散系数还是极差比率，低端代换较高，其次是乘法合成，高端代换最低，但超过加法合成与平方合成，离散系数与极差比率最低的是对数合成。它们的离散系数和极差比率关系大致如下：

调和平均 > 低端代换 > 乘法合成 > 高端代换 > 加法合成 > 平方平均 > 对数合成 公式 (9)

注意公式 (9) 中低端代换、乘法合成和高端代换排列在一起，因为本质上他们是乘法类合成。

加法类合成方法调和平均、加法合成、平方平均的高端区分度在逐渐降低，但低端区分度在逐渐提高。乘法类合成方法乘法合成、高端代换、低端代换的高端区分度比较接近，都比较高，仅次于调和平均；低端区分度也比较接近，总体较低，但高于调和平均，低于加法

合成与平方合成。对数合成的高端区分度最低，低端区分度最高。高端区分度的排序结果如下：

调和平均>高端代换>乘法合成>低端代换>加法合成>平方平均>对数合成 公式(10)
 低端区分度的排序结果如下：

调和平均<乘法合成<低端代换<高端代换<加法合成<平方平均<对数合成 公式(11)

公式(10)和公式(11)中乘法合成、低端代换、高端代换3种乘法类合成方法也排在一起。总体区分度、高端区分度和低端区分度的排序结果非常相似，可以从公式(9)、公式(10)、公式(11)非常明显地看出，基本顺序是调和平均、乘法类合成、加法合成、平方平均、对数合成。

乘法合成的总体区分度和高端区分度都大于加法合成，但是乘法合成的低端区分度小于加法合成，因此，文献[1]是不全面的。

表3 不同合成方法几个相对指标的比较

指标	调和	加法	平方	乘法	对数	高端代换	低端代换
基 平均值	17.03	30.87	38.87	22.65	1.30	38.29	17.66
本 极大值	62.52	67.12	71.62	64.85	1.81	100.00	57.71
统 极小值	1.51	9.40	13.97	5.56	0.74	10.06	0.00
计 标准差	10.84	11.13	10.98	11.28	0.22	16.59	10.76
区 离散系数	0.64	0.36	0.28	0.50	0.17	0.43	0.61
分 极差比率	0.98	0.86	0.80	0.91	0.59	0.90	1.00
度 高端区分度	0.54	0.42	0.35	0.50	0.25	0.51	0.48
低端区分度	0.09	0.17	0.23	0.10	0.30	0.14	0.12
打 汇总比率前 20%÷总分	0.31	0.24	0.22	0.27	0.18	0.23	0.30
分 汇总比率后 20%÷总分	0.05	0.08	0.09	0.06	0.11	0.09	0.04
强 高低端汇总比率	6.20	3.00	2.44	4.50	1.64	2.56	7.50
化							

2. 对较好期刊与较差期刊打分强化分析

从汇总比率看，低端代换对前 20%的科技期刊打分（0.30，列第二）较高，对后 20%的科技期刊打分最低（0.04），从而低端代换的汇总比率最高，达到 7.5，说明低端代换是一种强化较优期刊，弱化较低期刊的合成方法。

调和平均对前 20%的科技期刊打分（0.31）最高，对后 20%科技期刊打分较低（0.05，列倒数第二），导致汇总比率为 6.20，所以调和平均也是一种强化较优期刊，弱化较低期刊的一种合成方法。

乘法合成总体上仍然属于强化较优期刊、弱化较低期刊的合成方法，对前 20%科技期刊打分（0.27）列第三，对后 20%科技期刊打分（0.06）列倒数第三。

对数合成对前 20%科技期刊打分（0.18）最低，对后 20%科技期刊打分（0.11）最高，汇总比率为 1.64，最低。说明对数合成是一种弱化优秀期刊，强化较低期刊的一种合成方法。与以上三种合成方法相反。

调和平均、加法平均、平方平均对前 20%的科技期刊打分越来越低，对后 20%科技期刊打分越来越高，说明三种合成方法总体趋势是弱化较优期刊，强化较低期刊，相应汇总比率也越来越低，分别为 6.20、3.00、2.44。因此，文献[2]的观点是有问题的，幂次越低越是惩罚落后指标（这是正确的），幂次越高，越是奖励先进指标（错误，而是惩罚先进）。可以预计，随着公式(1)中 K 值取值越来越大（比如 K=3 或 K=4），这种趋势会继续下去，达到甚至低于对数合成的汇总比率。

高端代换、加法合成、平方合成的汇总比率总体相近。

4.4 科技期刊指标协调性与乘法类合成方法的关系

首先计算出每种期刊各指标的平均值，然后计算其标准差，接着用标准差除以平均值得到每种科技期刊指标的离散系数，最后分别用离散系数和乘法合成、对数合成、高端代换合成、低端代换合成进行回归，结果如表 4 所示。

表 4 指标离散系数与乘法类合成的关系

	乘法合成	对数合成	高端代换	低端代换
系数值	-0.019*** (-16.296)	-1.037*** (-17.908)	-0.007*** (-5.317)	-0.019*** (-14.267)
R ²	0.665	0.705	0.174	0.603
n	136	136	136	136

所有回归系数的统计检验都是显著的，系数值均为负。离散系数越低，说明科技期刊越是均衡发展，指标评价价值越高，又以乘法合成最为显著，最不显著的高端代换合成，回归系数较低，R²值仅为 0.174，相关程度较低。

4.5 各种合成方法特点的比较

将以上各指标结果加以归纳，结果如表 5 所示。高端代换特点并不明显，因此在极端情况下采用该合成方法外，一般不宜采用。调和平均、低端代换和乘法合成特点相似，区分度都比较高，都强化高端评价结果，弱化低端评价结果，但低端代换和乘法合成的区分度、强化程度、汇总比率等指标都相差不大，因此是鼓励各评价指标协调发展的合成方法，应根据需要选用。加法合成与平方合成强调指标间可以互相替代，在同类指标的综合评价中具有意义，但平方合成强化低端评价结果，可以在需要鼓励“后进”的评价中选用。对数合成弱化高端，强化低端，鼓励指标协调发展。

表 5 各种合成方法特点的比较

合成方法	总体区分度	高端区分度	低端区分度	高端强化程度	低端强化程度	高低端汇总比率	指标协调性强化程度	应用价值
调和	高	高	低	高	低	高	--	高
加法	一般	一般	一般	一般	一般	一般	--	高
平方	一般	一般	一般	一般	较强	一般	--	高
乘法	高	较高	低	较高	一般	较高	高	高
对数	低	低	高	低	强	低	高	高
高端代换	较高	较高	一般	一般	较强	一般	低	低
低端代换	高	较高	一般	高	低	高	高	高

5 结论与讨论

5.1 不同合成方法评价结果排序不同

不同合成方法评价结果的顺序不尽相同，因此一定要慎重选择合成方法。只有对数合成和乘法合成完全一致，当然，这和数据量的多少有关，在数据较少，数据间距较大的情况下，不同合成方法评价结果有可能一致，但这种情况毕竟是有限的。

5.2 应根据需要选择不同的合成方法

如果评价对象较多，比如医学学科科技期刊较多，应该考虑采用区分度较高的合成方法，进一步地，如果较差期刊较多，应考虑低端区分度高的合成方法。在评优时，要考虑采用高端区分度较高的合成方法，如调和平均。在某些二级指标内容相近时，可以考虑采用指标间能互相替代的加法合成，比如总被引频次、影响因子、即年指标三者都是反映被引情况的，可以采用加法合成得到被引指标的总分。在需要考虑评价对象全面发展时，可以考虑采用乘法合成或低端代换合成。在评价一些发展水平相对较低的领域，对数合成由于低端区分

度高，容易发现问题，鼓励落后和协调发展，排除了随机波动造成的干扰，因此具有重要意义。在某些二级指标内容相近，数据量较多，需要鼓励先进，强化高端评价对象，弱化低端评价对象，调和平均是比较合适的方法。

5.3 应注重各种合成方法的综合使用

由于各种合成方法特点迥异，而科技评价本身也比较复杂，因此应根据实际情况选取不同的合成方法。比如一级指标一般是不同属性，最好选取乘法类的合成方法，二级指标一般是同一属性，可以采用加法类合成方法。

5.4 注意评价结果之间的差距

对于普通加法合成而言，评价结果之间的相对差距很好计算，但是对于采用乘法合成、调和平均、对数合成等方法得出的评价结果，由于不是简单的线性函数关系，其相对差距计算比较复杂，不能简单化处理。

总之，由于不同合成方法的特点难以从数学上进行严格证明，因此在实际应用中必须结合评价目的、评价对象、指标选择、数据特点等属性进行综合分析，确定最合适的合成方法。

参考文献

- [1] 邱东. 多指标综合评价方法的系统分析[M]. 北京: 中国统计出版社, 1992
- [2] 苏为华. 多指标综合评价理论与方法研究[M]. 北京: 中国物价出版社, 2002
- [3] 张田玉. 学校教育评价[M]. 北京: 中央民族学院出版社, 1987
- [4] 田赤勇, 陈进殿等. 基于灰色模糊理论的埋地输油管道腐蚀与防护势态综合评判[J]. 全面腐蚀控制, 2006 (12): 20-24
- [5] 商红岩, 宁宣熙. 基于乘法合成的层次分析模型的第三方物流企业绩效评价[J]. 科技进步与对策, 2005 (11): 94-96