

科技评价中专家权重赋值优化研究¹

俞立平 潘云涛 武夷山

摘要：为了对科技评价中专家权重赋值进行优化，本文首先构建指标体系综合评价方法的分类体系，然后利用加法平均和证据理论对不同专家的权重赋值进行合成进而进行评价，并将评价结果按 20: 60: 20 的比例进行分级。接着比较不同专家分级评价结果与加法平均及证据理论分级结果的一致性，计算每个专家的灵敏度，可以找出权重赋值意见分歧最大的专家，从而提供了一种专家修正自己权重的定量比较方法。本文研究发现，证据理论合成能够充分尊重每个专家的意见，但在专家们对某个指标权重打分都比较低的情况下，该指标权重合成的结果可能极小甚至为 0，因此适用于一些不太成熟的科技评价或者用于新生事物的评价。对于一般的科技评价宜用加法平均进行合成，然后再进行灵敏度分析进而进行专家权重调整。将本方法结合传统的权重直接比较和评价结果直接比较可以取得良好的效果。

关键词：科技评价 权重合成 灵敏度 证据理论 权重优化

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。目前国内外综合评价方法有数十种之多，根据权重确定方式结合评价原理，本文将其分为三大类（如图1所示）：第一类是主观评价法，其基本原理是进行指标主观赋权，然后将数据标准化后加权汇总，如专家会议法、德尔菲法、层次分析法等等。第二类是客观评价法，包括两种，一种是采用客观赋权法确定指标权重，然后进行加权汇总，如熵权法、变异系数法、复相关系数法等；另一种是不需要赋权的系统方法，如主成分分析法、因子分析法、TOPSIS等。第三种是主客观相结合的赋权法，首先采用主观赋权方法确定权重，然后采用系统方法进行综合处理，如ELECTRE法、模糊综合评价法、PROMETHEE等等。这里的系统方法主要指综合运用运筹学、模糊数学、系统工程等领域的方法进行评价。指标体系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，不同评价方法得出的评价结果不一致，结果难以得到公认。

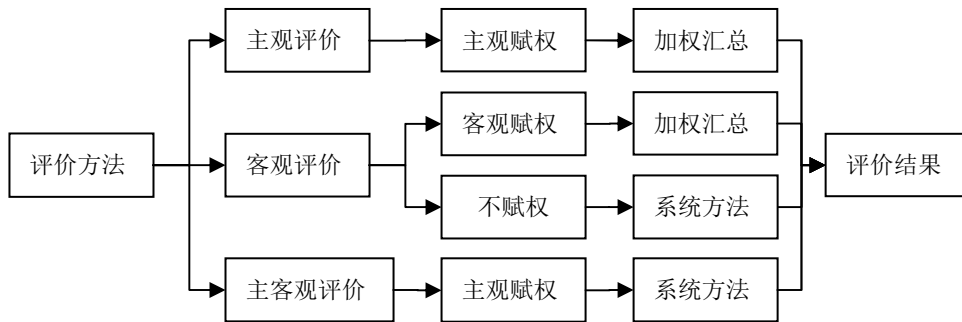


图1 评价方法体系结构

单纯的客观评价方法由于不考虑不同指标的重要性，完全凭数据说话，因此存在一些争议。主观赋权法有德尔菲法、专家会议法等。Helmer,O和 Gordon,T.J. (1946)^[1]创立了德尔菲法，后来日趋成熟。当时美国兰德公司为避免集体讨论存在的屈从于权威或盲目服从多数的缺陷，首次将德尔菲法用于技术预测中，之后便迅速地应用到了世界各国，在技术预测

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，扬州职业大学副教授，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

和新品市场需求预测等方面得到了较为普遍的应用。专家会议法是指根据规定的原则选定一定数量的专家,按照一定的方式组织专家会议,发挥专家集体的智能结构效应,从而做出判断的方法,与德尔菲法相比,专家是见面的。层次分析法(Analytical Hierarchy Process, AHP)是美国匹兹堡大学 A. L. Saaty (1974)^[2]应用网络系统理论和多目标综合评价方法,提出的一种层次权重决策分析方法。这种方法的特点是在对复杂决策问题的本质、影响因素及其内在关系等进行深入分析的基础上,利用较少的定量信息使决策的思维过程数学化,从而为多目标、多准则或无结构特性的复杂决策问题提供简便的决策方法。在多属性评价领域,我国学者田凤调(1987)^[3]发明了秩和比法(Rank Sum Ratio),邓聚龙(1984)^[4]创立了灰色系统理论。

主观赋权法离不开专家的权重,关于不同专家权重的合成方式,人们主要采用加法平均法对不同专家的权重进行合成。自Dempster(1967)^[5]创立证据理论以来,一些学者开始将其用于不同专家权重的合成,但这方面的应用迄今较少,就不同专家权重采取加法平均合成与证据理论合成所带来的差异开展实证研究的也不多。

无论哪种主观赋权方法,专家都必需以某种方式参考其他专家的赋权结果,进而修正自己的判断,但是迄今为止,除了简单进行专家间权重比较或评价结果的比较外,尚没有定量方法得知某个专家设定权重的评价结果与所有专家权重合成后评价结果的差异。若设法找到这种差异,会有利于专家们调整自己的权重赋值,因为主观赋权的最终结果总是专家们经过若干轮循环后达到的一种状态。本文利用中国科学技术信息研究所的医学科技期刊原始数据,首先分析加法平均合成与证据理论合成的特点,然后对每个专家权重的灵敏度进行分析,找出意见分歧较大的专家,从而为专家们进一步调整自己的权重提供数据分析支持。

2 方法

2.1 证据理论

证据理论是由 Dempster 首先提出,后来由他的学生 Shafer (1976)^[6]发展起来的,它是经典概率论的一种推广。证据理论提出,把对“假设”能构成影响的所有可能的证据收集起来,分解成一些相互独立的“元证据”(具单一因素的证据成份),组成一个证据空间,然后对这些元证据所有可能的组合赋以一个满足一定约束条件(比概率约束要弱)的值,从而得到一个定义在证据空间幂集上的一个函数,称为基本概率分配函数。

由于证据空间的子集不再互相独立,而且约束条件比概率弱,为了获得类似于概率的可信度意义,在基本概率函数基础上再设计一个函数,称为类概率函数,它满足类似于概率的约束条件,用这个类概率函数来表示证据的可信度。同时利用基本概率分配函数可以将“不知道”和“不确定”两种成份分开,这是一种对“不精确性”更深入细致的描述。证据理论的上述特点,使之非常适合于专家权重的合成。

2.2 灵敏度分析与专家赋权修正

灵敏度分析有多种,比如可以从指标权重的角度入手分析指标权重的变化对评价结果的影响,也可以从评价对象自身数据变化的角度进行分析。本文立足专家们权重加法合成与证据理论合成的评价结果,分别计算每个专家评价结果的灵敏度,在此基础上专家可以对自己的权重赋值进行优化,其步骤如下:

1、进行专家赋权

搜集第一轮专家评议结果,供后续分析。

2、合成评价结果

分别采用加法平均和证据理论合成,得到专家们权重的合成结果,共有两套赋权方案,分别为方案 A 和方案 B。

3、利用两套赋权方案进行评价

分别利用两套赋权方案进行评价,得到两套评价结果,分别为评价结果 A 和评价结果 B。

4、分级打分评价结果

根据正态分布的原理,科技期刊中较好的期刊和较差的期刊总是少数,参考日常生活中的二八定律,在最好的科技期刊差不多占 20%, 最差科技期刊也约占 20%的,可以据此进行分级。首先将评价结果 A 和评价结果 B 进行降序排列,然后按 20: 60: 20 的比例进行分级,最后将结果分别用 A、B、C 三个等级打分,当然共有两套,分别为打分 A 和打分 B。

5、根据单个专家权重进行评价和分级

利用每个专家的权重进行评价,然后根据上述原理对评价结果进行分级,并且根据等级打分为 A、B、C。

6、计算每个专家权重的灵敏度

分别筛选出每个专家打分结果与打分 A、打分 B 完全一致的个数,计算完全一致结果的百分比,得到每个专家加法合成的灵敏度和证据理论合成的灵敏度,灵敏度本质上就是每个专家的评价结果与所有专家权重合成评价结果的一致性程度。

7、向专家们提供灵敏度分析结果

向专家们提供加法平均合成与证据理论合成灵敏度分析的结果,供专家们进一步修正自己的判断。

以上步骤可以进行数轮,直到专家们不再更改自己的判断为止。

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,本文以基础医学与综合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 年的基础医学与综合医学类数据,共 136 个基础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。

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。此外,被引半衰期是反向指标,也进行了必要的处理。

4 实证结果

4.1 专家权重赋值及其合成

为了对专家主观权重赋值方法进行进一步分析,假定有 6 位专家,分别给 5 个指标赋予权重,然后用加法平均合成与证据理论合成进行专家权重的合成,其结果如表 1 所示。

表 1 专家权重及其合成

专家权重	总被引频次	被引半衰期	影响因子	即年指标	基金论文比
专家 1	0.5	0.05	0.3	0.05	0.1
专家 2	0.4	0.1	0.3	0.1	0.1
专家 3	0.2	0.1	0.2	0.3	0.2
专家 4	0.3	0.1	0.4	0.1	0.1
专家 5	0.1	0.2	0.3	0.2	0.2
专家 6	0.2	0.15	0.3	0.15	0.2
加法平均合成	0.283	0.117	0.3	0.15	0.15
证据理论合成	0.266	0.002	0.718	0.005	0.009

加法平均合成的权重与证据理论合成的权重相差较大,其中影响因子的权重在证据理论合成时被大大增加,达到 0.718,而均值只有 0.3;一些权重均值较小的指标在采用证据理

论合成时被大大削弱，如被引半衰期、即年指标、基金论文比。

根据证据理论的原理，专家权重的合成方法本质上是乘法合成，如果将证据理论合成的方法进行进一步化简可以分为三步，首先进行每个指标每个专家权重连乘，然后将连乘结果汇总，最后将每个指标连乘的结果除以连乘结果汇总，从而得到证据理论每个指标的权重。

根据证据理论，如果专家们对某个指标意见比较一致，虽然权重打分不尽相同，但总体权重打分偏低，那么连乘的结果也会非常小，从而导致该指标权重极低，如被引半衰期、即年指标、基金论文比就是这种情况。乘法合成还有另外一个特点，就是充分尊重每个专家权重的打分，这和加法平均是不同的，如果某个专家对某个指标打分偏低，即使其他专家打分都较高，那么连乘的结果受该专家打分的影响仍然较大。

4.2 评价分级及专家权重灵敏度分析

根据加法平均合成与证据理论合成得到的指标权重，分别进行评价，然后将评价结果按 20: 60: 20 的比例进行分级，由于共有 136 家期刊，因此 A 类期刊 27 种，B 类期刊 82 种、C 类期刊 27 种。由于期刊较多，表 2 给出了部分加法平均合成、证据理论合成以及专家 5 评价的分级结果，按加法平均得分降序排列，取前 30 种期刊。

表 2 部分评价及分级结果

刊名	加法平均得分	加法平均分级	证据理论得分	证据理论分级	专家 5 得分	专家 5 分级
中华医院管理杂志	66.46	A	85.55	A	66.98	A
中华医学杂志	62.30	A	73.36	A	50.23	A
中国危重病急救医学	59.56	A	76.30	A	58.91	A
细胞与分子免疫学杂志	42.22	A	39.85	A	49.48	A
CHINESE MEDICAL JOURNAL	40.58	A	43.10	A	40.64	A
中国病理生理杂志	39.39	A	37.64	A	39.33	A
中国疫苗和免疫	38.97	A	46.35	A	43.60	A
中国现代医学杂志	38.35	A	42.83	A	32.56	A
医学研究生学报	38.01	A	36.03	A	44.27	A
浙江大学学报医学版	37.96	A	30.03	B	46.66	A
中华医学遗传学杂志	36.87	A	37.92	A	41.09	A
生理学报	35.23	A	31.96	A	40.10	A
中华男科学杂志	34.69	A	37.76	A	38.85	A
中华麻醉学杂志	34.64	A	49.95	A	30.94	B
航天医学与医学工程	33.78	A	25.52	B	40.37	A
中华血液学杂志	33.71	A	43.47	A	31.31	B
ASIAN JOURNAL OF ANDROLOGY	33.56	A	41.56	A	39.71	A
吉林大学学报医学版	33.35	A	24.03	B	38.39	A
免疫学杂志	33.20	A	30.74	A	38.46	A
第四军医大学学报	33.17	A	34.60	A	29.44	B
中山大学学报医学科学版	32.70	A	32.06	A	37.55	A
中国寄生虫学与寄生虫病杂志	32.53	A	30.20	A	35.68	A
第三军医大学学报	32.40	A	31.46	A	31.79	B
中华病理学杂志	31.32	A	42.32	A	30.89	B
解剖学报	31.01	A	30.46	A	34.65	A
医疗设备信息	30.87	A	35.44	A	33.96	A
中国实验血液学杂志	30.43	B	28.32	B	36.63	A

南方医科大学学报	30.36	B	27.53	B	32.74	A
北京大学学报医学版	30.32	B	37.42	A	30.82	B
中国医院管理	30.30	B	30.40	A	32.84	A

很明显，权重赋值不同，评价结果不同，当然分级也不一样。如果根据证据理论合成后的评价结果进行降序排列，选前 30 种期刊，则结果必然与表 2 中的不相同。

灵敏度分析是用来比较不同专家评价分级与加法平均及证据理论评价分级的一致性，如表 3 所示。

表 3 各专家评价结果的灵敏度分析

	专家 1	专家 2	专家 3	专家 4	专家 5	专家 6	平均值
加法平均一致期刊 (X1)	116	123	121	128	120	128	122.67
加法平均灵敏度 (X1/136)	0.85	0.90	0.89	0.94	0.88	0.94	0.90
证据理论一致期刊 (X2)	113	120	102	122	104	108	111.50
证据理论灵敏度 (X2/136)	0.83	0.88	0.75	0.90	0.76	0.79	0.82

在 136 种期刊中，6 位专家赋权评价分级与加法平均评价分级几乎是不可能完全一致的，专家 4 和专家 6 赋权评价结果与所有专家评价赋权加法平均后的结果完全一致的期刊都是 128 种，灵敏度最高，达到 94%，专家 1 的灵敏度最低，仅 85%，总体平均灵敏度为 90%。

同样，在 136 种期刊中，6 位专家赋权评价分级与证据理论评价分级也不可能完全一致的，专家 4 的灵敏度最高，达到 90%，专家 3 的灵敏度最低，仅 75%，总体平均灵敏度为 82%。

证据理论的灵敏度小于加法平均的灵敏度，这是由证据理论的特点所决定的。

知道了每个专家赋权的灵敏度，就可以找出权重赋值分歧最大的专家，专家就可以在此基础上进行进一步的分析和思考，再次确认自己的权重，然后重新计算灵敏度，直到专家们不再更改自己的权重为止。

5 结论与讨论

5.1 证据理论合成与加法平均的适用范围不同

在专家权重的合成过程中，证据理论实际上是充分尊重每位专家意见的一种方法，但也正因为如此，在专家们对某个指标权重打分都比较低的情况下，该指标权重合成的结果可能极小甚至为 0，因此，证据理论合成比较适合于科技评价中对一些新生事物的评价，或者说适用一些不成熟的评价，在这种情况下，专家们对指标权重的分歧往往较大，但是每个专家的意见都得到了充分的尊重。

加法平均合成比较中规中矩，适用于一些成熟的评价，因此应用范围也比较广泛，是目前较常见的合成方法。

由于二者的原理不同，加法平均合成的灵敏度总体上要大于证据理论合成的灵敏度。究竟采取什么方法进行合成，要看具体的评价对象特点，一般只要选取一种权重合成方法，不需要两种都用。

5.2 灵敏度分析是专家修正自己权重的一种较好的定量方法

以往专家们在进行权重赋值时，可以参考其他专家的权重，但由于权重之间的互相依赖关系（权重之和一般为 1）以及数据众多，因此专家往往难以进行调整，而通过灵敏度分析，专家可以对比自己权重与所有专家权重合成的差距，同样也可以对比自己的评价结果与所有专家权重合成后评价结果的一致性，从而使自己的权重调整有章可循。

5.3 几种权重优化方法的比较

专家权重的优化方法主要有 3 种，第一种是专家比较自己的权重与其他专家的权重及合成权重，具体操作较复杂。第二种是专家比较自己赋权后的评价结果与其他所有专家赋权的

评价结果及合成权重的评价结果,往往难以比较,因为几乎每种结果的排序都不一致。第三种就是本文提出的方法,可操作性强。将以上三种方法结合在一起进行权重优化是一种较好的处理方式。

5.4 合成后的权重可用于进一步评价

合成后的权重是进一步评价的基础,至于用何种方法进行评价,还要综合评价目的、评价对象的特点等因素进行选取。由于客观评价方法只认数据说话,因此只能选取主观评价方法或主客观相结合的评价方法。

参考文献

- [1] Gordon, T. J. Helmer, O. (1964):Report on a Long-Range Forecasting Study, Rand Corporation, Santa Monica/ California.
- [2]A. L. Saaty. Measuring the Fuzziness of Sets[J]. Journal of Cybernetics,1974(4): 53-61.
- [3]Dempster, A. P. Upper and lower probabilities induced by a multivalued mapping. Annals of Mathematical Statistics, 1967, 38(2): 325-339.
- [4]田凤调. 秩和比法及其应用[M]. 北京:中国统计出版社, 1993, 1-93
- [5]邓聚龙. 灰色系统理论[M]. 武汉:华中工学院出版社, 1984,1-30
- [6] Shafer, G.. A Mathematical Theory of Evidence[J]. Princeton University Press, 1976.