

相似权法及其在属性综合评价中的应用

李建明 曲成毅 刘庆欧

【摘要】 目的 探讨综合评价中指标权重的确定。方法 以某 11 个省会城市的大气监测数据为实例,将相似权法应用于属性综合评价系统对该 11 个城市的大气质量进行聚类 and 排序。结果 用相似权作为权重得到的评价结果与按 Topsis 法得到的评价结果基本一致。结论 应用相似权法的属性综合评价既充分利用样本信息又合理体现指标权重,因此评价结果客观合理。

【关键词】 相似权;属性综合评价;大气质量

Similarity weight method and its application based on the attribute synthetic evaluation system LI Jian-ming*, QU Cheng-yi, LIU Qing-ou. *The Department of Mathematics, Shanxi Medical University, Taiyuan 030001, China

Corresponding author: Qu Cheng-yi, Email: quc-y@public.ty.sx.cn

【Abstract】 Objective To explore a way of determining the weights of variables in synthetic evaluation. **Methods** Based on the attribute synthetic evaluation system, similarity weight was used to cluster and rank the atmospheric quality on 11 large cities. **Results** Realized the classification and ranking for the atmospheric quality of the cities given, the results were almost comparable to that achieved by Topsis evaluation. **Conclusion** The method of comprehensive assessment using similarity weight in the attribute synthetic evaluation system seemed to be objective and rational. Not only it embodied the weights of variables involved, but also exploiting the information presented by the sample.

【Key words】 Similarity weight; Attribute synthetic evaluation; Atmospheric quality

综合评价是对一个复杂系统的多个指标进行总评价。一种好的评价方法,不但要能利用原始数据充分挖掘信息,还应有较好地确定指标权重的方法。基于属性测度的属性综合评价从原理上解决了模糊综合评价应用中出现的“分类不清,分类结果不合理”问题^[1,2],但在有关属性综合评价的文献资料中,指标权重的确定多用属于主观赋权法的属性 AHM 赋权法,鉴于相似权法可较好地利用原始数据的特点,也为了丰富属性综合评价内容,将相似权法引入属性综合评价,并以流行病学现场调查大气质量评价为例阐述其应用。

原理与方法

1. 属性综合评价系统:设评价对象为 $X = (x_1, x_2, \dots, x_n)$, 评价等级为 $C = (c_1, c_2, \dots, c_K)$, 评价指标为 $I = (I_1, I_2, \dots, I_m)$, 指标实测值为矩阵 $(x_{ij})_{n \times m}$ 。评价目标是依据 $(x_{ij})_{n \times m}$ 对 X 中的元素

聚类或排序。其步骤如下:

(1) 求单指标属性测度 μ_{ijk}, μ_{ik} 表示第 i 个对象按第 j 个指标属于第 k 个评价类的属性测度。

(2) 求综合属性测度 μ_{ik}, μ_{ik} 表示第 i 个对象属于第 k 个评价类的综合测度, μ_{ik} 按 m 个指标的权重对 μ_{ijk} 加权求和得到。设第 j 个指标的权重为 W_j , 则

$$\mu_{ik} = \sum_{j=1}^m w_j \mu_{ijk}, (1 \leq k \leq K) \tag{1}$$

由式(1)可得到综合属性测度矩阵 $(\mu_{ik})_{n \times K}$ 。其第 i 个行向量 $(\mu_{i1}, \mu_{i2}, \dots, \mu_{iK})$ 为 x_i 的综合测度评价向量。

(3) 按置信度准则聚类。令 λ 为置信度 $(0.5 < \lambda \leq 1)$, 若

$$k_0 = \min \left\{ k : \sum_{i=1}^n \mu_{ik} \geq \lambda \right\} \tag{2}$$

则称 $x_i \in c_{k_0}$ 类。按评分准则排序,对象 x_i 的得分为

$$q_i = \sum_{k=1}^K n_k \mu_{ik} \tag{3}$$

当 $c_1 > c_2 > \dots > c_K$ 时, $n_k = K + 1 - k$, 对 $c_1 < c_2 < \dots$

作者单位:030001 太原,山西医科大学数学教研室(李建明、刘庆欧),流行病学教研室(曲成毅)

通讯作者:曲成毅,Email: quc-y@public.ty.sx.cn

$< c_K$, 取 $n_k = k$ 。

2. 用相似权重法确定指标权重: 当评价指标是客观型指标, 其权重无法由专家凭经验给出时, 可初步假定各指标有相同的权重, 即 $w_j = \frac{1}{m}$ ($j = 1, 2, \dots, m$)。再用公式 $\mu_{ik} = \frac{1}{m} \sum_{j=1}^m \mu_{ijk}$ 可求出 x_i 的综合属性测度向量 $(\mu_{i1}, \mu_{i2}, \dots, \mu_{iK})$, 该向量在平均的意义下反映了 x_i 的总体评价情况。这样, 单指标属性测度评价向量 $(\mu_{ij1}, \mu_{ij2}, \dots, \mu_{ijK})$ 与综合属性测度评价向量 $(\mu_{i1}, \mu_{i2}, \dots, \mu_{iK})$ 的“相近”程度体现了指标 I_j 反映总体情况的能力^[3], 两者越相近, 说明 I_j 越能体现总体情况, 权重应越大。令 r_j 为相似系数, w_j 为相似权。

$$r_j = \frac{1}{n} \sum_{i=1}^n (\mu_{ij1}, \mu_{ij2}, \dots, \mu_{ijK}) \cdot (\mu_{i1}, \mu_{i2}, \dots, \mu_{iK})^T \quad (4)$$

$$= \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K \mu_{ijk} \cdot \mu_{ik}$$

$$w_j = r_j / \sum_{j=1}^m r_j \quad (5)$$

则将相似权代入式(1)重新求综合属性测度。

实例分析

监测我国 11 个省会城市的大气质量指标^[4], 按照国家大气环境质量标准(GB 83095-82)的规定, 选取 SO_2 、 NO_x 、TSP、降尘四个指标, 把大气环境质量分为 5 类: $c_1 = \{\text{清洁}\}$ 、 $c_2 = \{\text{轻污染}\}$ 、 $c_3 = \{\text{中污染}\}$ 、 $c_4 = \{\text{重污染}\}$ 、 $c_5 = \{\text{严重污染}\}$ 。对该 11 个城市的大气质量进行评价(表 1)。依实际情况, 确定研究对象空间为: $X = \{x_1, x_2, \dots, x_{11}\}$, 评价指标集为 $I = \{I_1, I_2, I_3, I_4\}$ 。即 $I = (SO_2, NO_x, TSP, \text{降尘})$ 。

表1 我国 11 个省会城市大气质量评价指标分级标准

指标	级 别				
	清洁	轻污染	中污染	重污染	严重污染
$SO_2(x_1)$	0.05	0.15	0.25	0.50	0.85
$NO_x(x_2)$	0.05	0.10	0.15	0.30	0.50
TSP(x_3)	0.15	0.30	0.50	1.00	0.70
降尘(x_4)	3.00	13.00	22.00	44.00	75.00

1. 建立单指标属性测度函数: 将指标实测值, 对照图 1 并按属性测度函数的分段线性函数^[2]求单指标属性测度矩阵 $(\mu_{ijk})_{4 \times 5}$, ($i = 1, 2, \dots, 11$)。

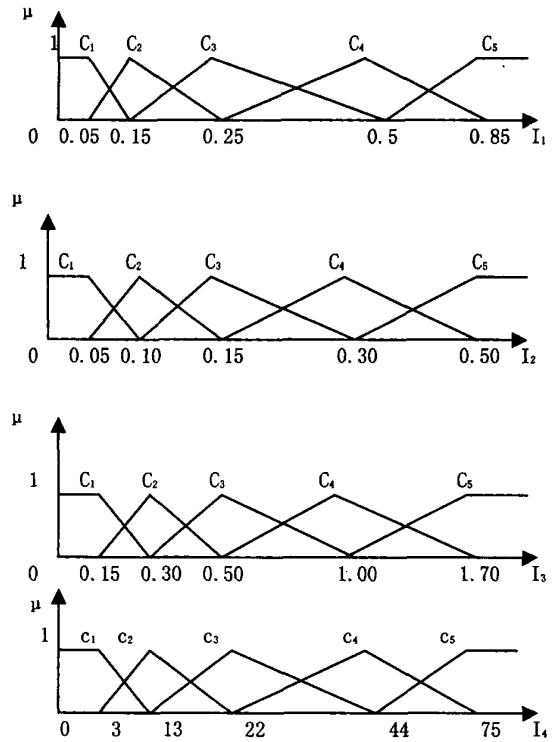


图1 单指标属性测度函数

单指标属性测度矩阵为:

$$\begin{aligned}
 x_1: \mu_{1jk} &= \begin{bmatrix} 0.500 & 0.500 & 0.000 & 0.000 & 0.000 \\ 1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.067 & 0.933 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.430 & 0.570 & 0.000 & 0.000 \end{bmatrix} \\
 x_2: \mu_{2jk} &= \begin{bmatrix} 0.330 & 0.670 & 0.000 & 0.000 & 0.000 \\ 0.380 & 0.620 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.225 & 0.775 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.447 & 0.523 & 0.000 \end{bmatrix} \\
 x_3: \mu_{3jk} &= \begin{bmatrix} 0.560 & 0.440 & 0.000 & 0.000 & 0.000 \\ 0.220 & 0.780 & 0.000 & 0.000 & 0.000 \\ 0.940 & 0.060 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.936 & 0.064 & 0.000 \end{bmatrix} \\
 x_4: \mu_{4jk} &= \begin{bmatrix} 1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.880 & 0.120 & 0.000 & 0.000 & 0.000 \\ 0.133 & 0.867 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.640 & 0.360 & 0.000 & 0.000 \end{bmatrix} \\
 x_5: \mu_{5jk} &= \begin{bmatrix} 0.270 & 0.730 & 0.000 & 0.000 & 0.000 \\ 0.760 & 0.240 & 0.000 & 0.000 & 0.000 \\ 0.873 & 0.127 & 0.000 & 0.000 & 0.000 \\ 0.068 & 0.932 & 0.000 & 0.000 & 0.000 \end{bmatrix} \\
 x_6: \mu_{6jk} &= \begin{bmatrix} 0.000 & 0.000 & 0.976 & 0.024 & 0.000 \\ 0.620 & 0.380 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.852 & 0.148 & 0.000 \\ 0.000 & 0.000 & 0.482 & 0.518 & 0.000 \end{bmatrix} \\
 x_7: \mu_{7jk} &= \begin{bmatrix} 0.730 & 0.270 & 0.000 & 0.000 & 0.000 \\ 0.200 & 0.800 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.970 & 0.030 & 0.000 \\ 0.000 & 0.000 & 0.911 & 0.089 & 0.000 \end{bmatrix} \\
 x_8: \mu_{8jk} &= \begin{bmatrix} 0.600 & 0.400 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.360 & 0.640 & 0.000 & 0.000 \\ 1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.344 & 0.656 & 0.000 & 0.000 & 0.000 \end{bmatrix}
 \end{aligned}$$

$$x_9: \mu_{9jk} = \begin{pmatrix} 0.000 & 0.280 & 0.720 & 0.000 & 0.000 \\ 0.000 & 0.220 & 0.780 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.586 & 0.414 & 0.000 \\ 0.000 & 0.000 & 0.828 & 0.172 & 0.000 \end{pmatrix}$$

$$x_{10}: \mu_{10jk} = \begin{pmatrix} 1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.490 & 0.510 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.741 & 0.259 & 0.000 \end{pmatrix}$$

$$x_{11}: \mu_{11jk} = \begin{pmatrix} 0.600 & 0.400 & 0.000 & 0.000 & 0.000 \\ 0.960 & 0.040 & 0.000 & 0.000 & 0.000 \\ 0.893 & 0.107 & 0.000 & 0.000 & 0.000 \\ 0.378 & 0.622 & 0.000 & 0.000 & 0.000 \end{pmatrix}$$

2. 求相似权与综合属性测度评价矩阵:先假定大气质量四个评价指标 SO_2 、 NO_x 、TSP、降尘具有相

同权重即: $(w_1 = w_2 = w_3 = w_4) = \frac{1}{m} = 0.25$ 。将单

指标属性测度矩阵 $(\mu_{ijk})_{4 \times 5}$, $(i = 1, 2, \dots, 11)$ 代入

公式 $\mu_{ik} = \frac{1}{m_j} \sum_{j=1}^m \mu_{ijk}$ 求得初始综合测度评价矩阵为:

$$(\mu_{ik})_{11 \times 5} = \begin{pmatrix} 0.392 & 0.466 & 0.142 & 0.000 & 0.000 \\ 0.178 & 0.379 & 0.313 & 0.131 & 0.000 \\ 0.430 & 0.320 & 0.236 & 0.016 & 0.000 \\ 0.503 & 0.407 & 0.090 & 0.000 & 0.000 \\ 0.493 & 0.507 & 0.000 & 0.000 & 0.000 \\ 0.155 & 0.095 & 0.577 & 0.173 & 0.000 \\ 0.233 & 0.268 & 0.470 & 0.030 & 0.000 \\ 0.486 & 0.354 & 0.160 & 0.000 & 0.000 \\ 0.000 & 0.125 & 0.729 & 0.146 & 0.000 \\ 0.500 & 0.122 & 0.313 & 0.065 & 0.000 \\ 0.708 & 0.292 & 0.000 & 0.000 & 0.000 \end{pmatrix}$$

由式(4)得相似数向量: $(r_1, r_2, r_3, r_4) = (0.452, 0.403, 0.451, 0.368)$ 。由式(5)得相似权: $(w_1,$

$w_2, w_3, w_4) = (0.270, 0.241, 0.269, 0.220)$ 。以

求得的相似权作为权重,再将单指标属性测度矩阵

$(\mu_{ijk})_{4 \times 5}$, $(i = 1, 2, \dots, 11)$ 代入式(1)重新求得综合

属性测度评价矩阵如下:

$$(\mu_{ik})_{11 \times 5} = \begin{pmatrix} 0.394 & 0.481 & 0.125 & 0.000 & 0.000 \\ 0.181 & 0.391 & 0.314 & 0.115 & 0.000 \\ 0.457 & 0.323 & 0.206 & 0.014 & 0.000 \\ 0.518 & 0.403 & 0.079 & 0.000 & 0.000 \\ 0.506 & 0.494 & 0.000 & 0.000 & 0.000 \\ 0.149 & 0.092 & 0.599 & 0.160 & 0.000 \\ 0.245 & 0.266 & 0.461 & 0.028 & 0.000 \\ 0.507 & 0.339 & 0.154 & 0.000 & 0.000 \\ 0.000 & 0.129 & 0.722 & 0.149 & 0.000 \\ 0.511 & 0.132 & 0.300 & 0.057 & 0.000 \\ 0.717 & 0.283 & 0.000 & 0.000 & 0.000 \end{pmatrix} \quad (6)$$

3. 聚类:取置信度 $\lambda = 0.6$, 由综合属性测度评价

矩阵(6)及置信度准则式(2)识别得:天津、大连、武汉、杭州、广州、呼和浩特、福州属于 c_2 类即“轻污染”;沈阳、太原、郑州、乌鲁木齐属于 c_3 类即“中污染”。

4. 排序:因 (c_1, c_2, \dots, c_5) 是有序分割类,且 $c_1 > c_2 > c_3 > c_4 > c_5$, 由评分准则式(3), 并取 $n_i = 6 - i$, 求得 11 个城市大气质量得分并优劣排序如表 2。

表2 我国 11 个省会城市大气质量排序

城市	得分	排序	Topsis 法
福州	4.708	1	4
杭州	4.493	2	6
武汉	4.413	3	3
广州	4.326	4	1
天津	4.249	5	5
大连	4.164	6	8
呼和浩特	4.058	7	2
郑州	3.703	8	7
沈阳	3.603	9	9
太原	3.232	10	10
乌鲁木齐	2.979	11	11

讨 论

由表 2 应用配对设计差值的符号秩和检验, $T_- = 10.5, \alpha > 0.05$, 可认为将相似权法应用于属性综合评价对 11 个城市的大气质量的评价结果与 Topsis 法的评价结果差异不显著。但排序上尚有一定的不同,这主要是由于一方面应用 Topsis 法不考虑指标权重所致;另一方面 Topsis 法是按样本点距离最优样本点的相对距离来排序。事实上,从有限样本中确定的最优点与最劣点不一定是客观的最优与最劣点。

关于指标无量纲化:单指标属性测度 μ_{ijk} 与综合属性测度 μ_{ik} 的值都介于 $[0, 1]$ 之间,在计算 μ_{ijk} 时自然而然地对指标进行了无量纲化。因此,应用属性综合评价不用专门对指标无量纲化。

关于参评指标的选取^[5]:本文中实例应用是对大气质量的综合评价,按 GB 83095-82 的规定,应选取 SO_2 、 NO_x 、TSP、降尘等四项指标。若实际问题无标准可依据且指标较多时,可先对指标聚类,再用条件广义方差极小法、极大不相关法等选取典型指标。

关于指标权重:作为客观赋权法的相似权法,可以充分利用样本资料信息,避免了确定指标权重时的主观随意性。缺点是手工操作时工作量较大,但由于其原理简明,易于编程运算。

关于属性综合评价方法:属性综合评价建立于

属性数学基础上,其立足点是属性集与属性测度。属性测度满足测度论关于测量的基本准则:整体测度等于部分测度之和。属性综合评价置信度准则,解决了模糊综合评价中最大隶属度准则失效的问题^[1,2]。另外,该方法优点还在于不必考虑指标的分布型,也无需对变量标准化。

参 考 文 献

1 程乾生. 属性识别理论模型及其应用. 北京大学学报(自然科学

版),1997,33:12-20.

2 程乾生. 属性集和属性综合评价系统. 系统工程理论与实践, 1997,17:1-8,42.

3 庞彦军,刘开第,张博文. 综合评价系统客观性指标权重的确定方法. 系统工程理论与实践,2001,21:37-42.

4 姜成英,胡珍姣. 大气污染程度的 Topsis 分析方法. 数理医药学杂志,1998,11:299-300.

5 胡永宏,贺思辉. 综合评价方法. 北京:科学出版社,2000. 15-22.

(收稿日期:2004-01-30)

(本文编辑:张林东)

• 疾病控制 •

膳食纤维干预改善便秘作用的研究

江国虹 杨溢 李静 常改

为探索社区开展膳食纤维干预改善便秘的方法及在人群中开展相关干预奠定基础,通过社区医院收集便秘患者,在知情同意、问卷调查、膳食调查和健康体检的基础上采用双盲的随机对照试验,将研究对象随机编号(从 1 编到 120 号),预先规定偶数为试验组(服用膳食纤维素片),奇数为对照组(服用安慰剂),研究课题由天津市卫生防病中心营养与食品卫生学部负责质量控制。应用 SPSS 软件分析自身前后及组间便秘程度的改善。每个试验对象在试食前后均按时分别完成连续 3 天膳食记录表及连续 7 天的排便记录表,排便记录包括(排便次数、排便状况、粪便性状);每日服用摄入膳食纤维或安慰剂的情况(包括不良反应)的记录;并采用电话提示和面对面咨询进行随访,人均 1~2 次/周,调查表复合率为 70%,说明研究对象的依从性较好。

试食前后两组便秘程度的组间比较见表 1。试食前后试验组便秘程度的自身比较见表 2。试食前后两组纤维素类食物摄入量的自身比较见表 3。

本研究结果显示试验组与对照组三项纳入指标差异无统计学意义, P 值均 > 0.05,说明该研究贯穿了随机的原则,保证了两组的均衡;试食前后两组纤维素类食物摄入量的自身比较无差异(P > 0.05),说明本研究是在维持常规饮食的前提下,观察适量增加膳食纤维对便秘的改善作用,体现本研究排除了相关干扰因素;试验组自身前后的排便次数、排便状况明显改变(P < 0.01, P < 0.05);试食后组间的排便次数、排便状况、粪便性状也均有差异,分别是 P < 0.01、P < 0.05,说明膳食纤维素具有较好的通便功能,不仅有效的改善了人体的便秘状态,而且达到了社区健康促进的目的。

表1 试食前后两组便秘程度的组间比较($\bar{x} \pm s$)

组 别	例数	试 食 前			试 食 后		
		排便次数	排便状况	粪便性状	排便次数	排便状况	粪便性状
试验组	60	3.42 ± 1.34	5.36 ± 2.17	4.21 ± 1.95	4.50 ± 2.21	4.61 ± 2.01	3.96 ± 1.95
对照组	58	3.58 ± 2.12	5.53 ± 2.46	4.52 ± 2.68	3.64 ± 1.72	5.25 ± 2.11	4.58 ± 2.74
P 值		0.5093	0.5449	0.2701	<0.01	<0.05	<0.05

表2 试食前后试验组便秘程度的自身比较($\bar{x} \pm s$)

试食	例数	排便次数	排便状况	粪便性状
前	60	3.42 ± 1.34	5.36 ± 2.17	4.21 ± 1.95
后	60	4.50 ± 2.21	4.61 ± 2.01	3.96 ± 1.95
P 值		<0.01	<0.01	0.3278

表3 试食前后两组纤维素食物摄入情况的自身比较($\bar{x} \pm s$)

组别	纤维素类食物摄入量* (g/d)	P 值
试验组	95.625 ± 61.336	0.0804
对照组	84.386 ± 58.151	0.1523

* 纤维素食物摄入量为试食前后差值的均数

(收稿日期:2004-07-01)

(本文编辑:张林东)

作者单位:300011 天津市卫生防病中心