【基础理论与应用研究】

基于主成分分析的指标权重确定方法

韩小孩[1]，张耀辉[1]，孙福军[2]，王少华[1]

(1 装甲兵工程学院 维修工程教研室，北京　100072; 2 71602 部队，山东 潍坊　261055)

摘要: 介绍一种基于主成分分析的权重确定方法。通过分析主成分分析方法的应用现状，提出了基于主成分分析的权重确定思路。给出了主成分分析方法的基本原理及其计算过程，并在此基础上提出了权重确定的假设，建立了基于主成分分析的权重确定模型。最后结合例子实现了指标权重的确定。

关键词: 主成分分析; 综合评价函数; 综合得分值

中图分类号: E919　　文献标识码: A　　文章编号: 1006 - 0707(2012)10 - 0124 - 03

主成分分析(principal component analysis，PCA)，也称主分量分析或矩阵数据分析。它通过变量变换的方法把相关的变量变为若干不相关的综合指标变量[1]，从而实现对数据集的降维，使得问题得以简化。现行的关于主成分分析的应用研究中大多集中于数据的简化处理或综合评价上[2-3]。文献[4]中介绍主成分的在权重确定方面的研究，虽提出了权重确定的一般方法，但由于所需样本数据较多，在实际应用时通用性不强。本文旨在研究一种权重确定方法，在无需指标样本数据的情况下利用主成分分析方法基本原理，解决权重确定问题。

1　主成分方法

1.1　基本原理

主成分分析的原理可以简单的陈述如下: 借助一个正交变换，将其分量相关的原随机向量

$$X = (x_1, x_2, \cdots, x_p)^{\mathrm{T}}$$

转化成其分量不相关的新随机变量

$$U = (u_1, u_2, \cdots, u_p)^{\mathrm{T}}$$

使之指向样本点散布最开的 p 个正交方向，然后对多维变量系统进行降维处理，使之能以一个较高的精度转换成低维变量系统[5-7]。

1.2　计算步骤

1) 构造样本阵

$$X = \begin{bmatrix} x_1^{\mathrm{T}} \\ x_2^{\mathrm{T}} \\ \vdots \\ x_n^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

其中，x_{ij} 表示第 i 组样本数据中的第 j 个变量的值。

2) 对样本阵 X 进行变换得 $Y = [y_{ij}]_{n\times p}$，其中

$$y_{ij} = \begin{cases} x_{ij}, & \text{对正指标} \\ -x_{ij}, & \text{对逆指标} \end{cases}$$

3) 对 Y 做标准化变换得标准化阵

$$Z = \begin{bmatrix} z_1^{\mathrm{T}} \\ z_2^{\mathrm{T}} \\ \vdots \\ z_n^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1p} \\ z_{21} & z_{22} & \cdots & z_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{np} \end{bmatrix}$$

其中，$z_{ij} = \dfrac{y_{ij} - \bar{y}_j}{s_j}$，$\bar{y}_j$，$s_j$ 分别为 Y 阵中第 j 列的均值和标准差。

4) 计算标准化阵 Z 的样本相关系数阵

$$R = [r_{ij}]_{p\times p} = \frac{Z^{\mathrm{T}} Z}{n-1}$$

5) 求特征值

$$|R - \lambda I_p| = 0$$

解得 p 个特征值 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p \geqslant 0$。

6) 确定 m 值，使信息的利用率达到 80% 以上。确定方法为

$$\frac{\sum_{j=1}^{m} \lambda_j}{\sum_{j=1}^{p} \lambda_j} \geqslant 0.8$$

对每个 λ_j，$j = 1, 2, \cdots, m$。解方程组 $Rb = \lambda_j b$，得单位向量 $b_j^0 = \dfrac{b_j}{\| b_j \|}$。

7) 求出 $z_i = (z_{i1}, z_{i2}, \cdots, z_{ip})^{\mathrm{T}}$ 的 m 个主成分分量

$$u_{ij} = z_i^{\mathrm{T}} b_j^0, j = 1, 2, \cdots, m$$

得决策矩阵

收稿日期: 2012 - 06 - 11

作者简介: 韩小孩(1987—)，男，硕士研究生，主要从事维修理论与技术研究。

$$U=\begin{bmatrix}u_1^{\mathrm{T}}\\u_2^{\mathrm{T}}\\\vdots\\u_p^{\mathrm{T}}\end{bmatrix}=\begin{bmatrix}u_{11}&u_{12}&\cdots&u_{1m}\\u_{21}&u_{22}&\cdots&u_{2m}\\\vdots&\vdots&\cdots&\vdots\\u_{p1}&u_{p2}&\cdots&u_{pm}\end{bmatrix}$$

其中,u_i 为第 i 个变量的主成分向量。

1.3　建立权重模型

1)提出假设

假设需确定权重的指标个数为 h 个。现分别咨询 L 位专家得出 h 组权重评分值,其中每组评分值中均有 L 个元素。具体形式可由表 1 表示。

表 1　专家打分

指标＼专家	w_1	w_2	$\cdots$	w_L
v_1	p_{11}	p_{12}	$\cdots$	p_{1L}
v_2	p_{21}	p_{22}	$\cdots$	p_{2L}
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
v_h	p_{h1}	p_{h2}	$\cdots$	p_{hL}

由于各位专家所研究方向不同,其打分也存在一定的偏向,从而给权重的确定带来一定的模糊性。研究发现,专家人数越多,得到的权重越科学,与此同时权重的确定也就越模糊。在此基础上提出以下假设,即在专家人数不变的情况下,利用各位专家评分间的线性关系对实际评分专家数进行类似的简化,从而实现权重评判的精确性。经分析得,思路符合主成分分析的基本原理,故可尝试用主成分分析方法来确定权重。

2)权重确定过程

根据上述条件可知,权重的确定过程其实就是主成分分析求综合评价函数的过程。在此过程中,原评价系统中的指标变为样本;现有指标为各位专家。具体的权重确定流程可用图 1 表示。

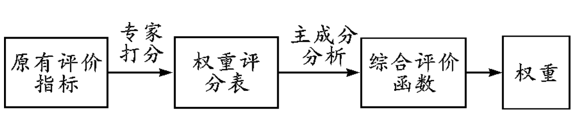

图 1　权重确定流程

3)权重模型

首先确定的初级权重模型即是主成分模型

$$\begin{cases}F_1=u_{11}w_1+u_{21}w_2+\cdots+u_{L1}w_L\\F_2=u_{12}w_1+u_{22}w_2+\cdots+w_{L2}w_L\\\vdots\\F_m=u_{1m}w_1+u_{2m}w_2+\cdots+u_{Lm}w_L\end{cases}\tag{1}$$

式中,$F_1,F_2,\cdots,F_m$ 为分析后得到的 m 个主成分;u_{ij}为决策矩阵中系数。需要指出的是,在用 SPSS 软件进行主成分分析时,得到不是决策矩阵系数 u_{ij}而是初始因子载荷 f_{ij}。二者满足如下关系

$$u_{ij}=\frac{f_{ij}}{\sqrt{\lambda_j}},j=1,2,\cdots,m\tag{2}$$

在此基础上构建综合评价函数:

$$F_Z=\sum_{j=1}^{m}(\lambda j/\kappa)Fj=a_1w_1+a_2w_2+\cdots+a_Lw_L,$$
$$\kappa=\lambda_1+\lambda_2+\cdots+\lambda_m\tag{3}$$

式中,$a_1,a_2,\cdots a_L$ 即指标 $w_1,w_2,\cdots,w_L$ 在主成分中的综合重要度。在此基础上结合专家实际打分,可算出原有指标得分综合值。

$$V_{Zi}=\sum_{j=1}^{L}a_jp_{ij},i=1,2,\cdots,h\tag{4}$$

可得各指标权重为

$$\omega_i=V_{Zi}/\sum_{i=1}^{h}V_{Zi}\tag{5}$$

由式(3)、式(4)、式(5)可得二级权重模型

$$\begin{cases}F_Z=\sum_{j=1}^{m}(\lambda j/\kappa)Fj=a_1w_1+a_2w_2+\cdots+a_Lw_L\\V_{Zi}=\sum_{j=1}^{L}a_jp_{ij}\\\omega_i=V_{Zi}/\sum_{i=1}^{h}V_{Zi}\end{cases}\tag{6}$$

因此可确定总的权重模型如图 2 所示。

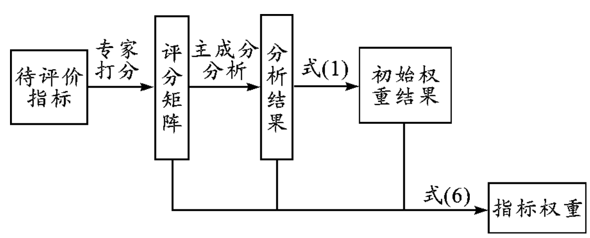

图 2　指标权重确定模型

2　示例分析

现有任务成功性评定指标集:(任务强度,环境等级,寿命等级,人员素质,技术状态)。

假定变量 $V=(v_1,v_2,\cdots,v_5)$ 与指标集中元素满足如下对应关系:v_1 – 任务强度;v_2 – 环境等级;v_3 – 寿命等级;v_4 – 人员素质;v_5 – 技术状态。另有,w_1 – 专家 1;w_2 – 专家 2;w_3 – 专家 3;w_4 – 专家 4;w_5 – 专家 5;w_6 – 专家 6。

由 6 位专家采取 5 分制原则对 5 个指标进行评分得表 2。

以原指标项为样本,专家项为指标对评分表进行主成分分析。经 SPSS 软件分析得表 3、表 4。

表2 评价指标得分表

指标＼专家	专家1	专家2	专家3	专家4	专家5	专家6
任务强度	3	4	3	3	3	4
环境等级	2	3	2	3	2	2
寿命等级	5	4	4	3	3	4
人员素质	3	2	3	2	3	3
技术状态	5	3	4	4	3	4

表中:5－非常重要;4－比较重要;3－一般重要;2－不太重要;1－不重要

表3 成分矩阵

	成分	
	1	2
专家1	0.930	－0.021
专家2	0.491	0.625
专家3	0.960	－0.161
专家4	0.518	0.711
专家5	0.802	－0.542
专家6	0.958	－0.069

表4 方差解释表

成分	初始特征值		
	合计	方差的%	累积%
1	3.860	64.328	64.328
2	1.222	20.362	84.690
3	0.709	11.814	96.504
4	0.210	3.496	100.000
5	－6.417E－17	－1.070E－15	100.000
6	－2.059E－16	－3.432E－15	100.000

将表(3)结果经式(2)转换后代入初始权重模型即式(1)可得

$$\begin{cases} F_1 = 0.473\,4w_1 + 0.249\,9w_2 + 0.488\,6w_3 + \\ \quad 0.263\,7w_4 + 0.408\,2w_5 + 0.487\,6w_6 \\ F_2 = -0.019w_1 + 0.565\,4w_2 - 0.145\,6w_3 + \\ \quad 0.643\,2w_4 - 0.490\,3w_5 - 0.062\,4w_6 \end{cases}$$

再结合上述结果以及评分表和表(4)结果共同代入二级权重模型即式(6)可得

$$F_Z = 0.355w_1 + 0.325\,7w_2 + 0.336\,1w_3 + 0.354\,9w_4 + 0.192\,2w_5 + 0.355\,4w_6$$

$$\omega = (\omega_1, \omega_2, \omega_3, \omega_4, \omega_5) = (0.207\,5, 0.145\,6, 0.241\,2, 0.163\,6, 0.242\,1)$$

即指标集(任务强度,环境等级,装备寿命状态等级,人员素质等级,装备技术状态等级)对应的权重集为(0.207 5,0.145 6,0.241 2,0.163 6,0.242 1)。

3 结束语

本文针对主成分分析方法缺少在指标权重确定方面的应用这一问题,提出了一种无需多组样本数据的基于主成分分析的权重确定方法。建立了基于主成分分析的权重确定模型,并结合例子进行了分析。该模型通用性较强,可有效应用于各种指标权重确定问题中,为各类评估问题的进行奠定了基础。

参考文献:

[1] 汪应洛.系统工程[M].4版.北京:机械工业出版社,2008.

[2] 孙晓东,田澎.类加权主成分分析在企业物流绩效评价中的应用[J].工业工程与管理,2007(1):57－63.

[3] 侯文.对应用主成分法进行综合评价的探讨[J].数理统计与管理,2006,25(2):211－214.

[4] 杨春周,滕克难,程月波.作战效能评估指标权重的确定[J].计算机仿真,2008,25(10):5－7.

[5] 王学民.应用多元分析[M].上海:财经大学出版社,2004.

[6] 高惠璇.应用多元统计分析[M].北京:北京大学出版社,2005.

[7] 秦寿康.综合评价原理与应用[M].北京:电子工业出版社,2003.

(**责任编辑** 杨继森)