

文章编号:1001-9081(2007)09-2346-03

PCA 在克服变量多重相关性中的局限作用

郭凯红

(辽宁大学 信息科学与技术学院, 沈阳 110036)

(guokh@126.com)

摘要:针对主成分分析(PCA)在非线性的观测变量中应用的局限作用,对 PCA 进行了理论研究。基于欧氏空间和统计方法,讨论了 PCA 的数学本质,以变量高度多重相关为例,分析了非线性系统结构,提出并证明了 PCA 在克服变量多重相关性和多指标系统评估中存在局限性的必然原因。针对一些具体的非线性问题,提出了若干改进的 PCA 方法,以及消除其局限性的方法和建议。

关键词:主成分分析;多指标;多重相关;数据变异与相似

中图分类号: TP301.6 **文献标志码:** A

Limitations of PCA in solving the multiple correlations between variables

GUO Kai-hong

(College of Information Science and Technology, Liaoning University, Shenyang Liaoning 110036, China)

Abstract: In view of the limitations of PCA's application to variables observed with nonlinear features, Principal Component Analysis (PCA) was studied theoretically. Based on the coordinates of Euclidean-space and statistical method, the study discussed the mathematical essence of PCA, and analyzed the structure of nonlinear system, and then proposed and proved the causes of the limitations of PCA in (a) solving the multiple correlations between variables efficiently and (b) using in multiple indexes system evaluation under the condition of the multiple correlations by an example of multiple correlations between variables in high degree. For some nonlinear problems, the improvement on PCA and solutions to the limitations were presented accordingly.

Key words: Principal Component Analysis(PCA); multiple indexes; multiple correlations; data variety and similarity

0 引言

主成分分析(PCA)是一种实用的多元统计决策方法,在经济管理与计算机决策领域有着广泛的应用。它的工作目标,是要对高维多变量数据系统进行最佳综合简化,在力保数据信息损失最少的原则下,对高维变量空间进行降维处理^[1]。然而,通过对主成分分析深入研究后发现,当原指标变量多重相关^[2,3],特别是高度多重相关时,使用主成分分析比较系统中各个样本点间的相似性和差异性^[4],依然不能完全有效地克服这种相关性;并且在这种情况下,使用主成分分析进行多指标系统评估,在某种程度上将具有双重夸大重叠信息、扭曲真实数据的不良作用。本文基于欧氏空间讨论了主成分分析的数学本质,分析并证明了上述问题存在的必然性,以及主成分分析在多指标系统评估中的局限作用,最后提出了改进的主成分分析及消除其局限性的有关方法。

1 PCA 本质概述

对于一个指标数量为 p 的多指标决策,可以用一个 p 维随机向量 $X = (X_1, X_2, \dots, X_p)^T$ 来对特性指标加以描述,其中 X_i 是表示第 i 个特性指标的随机变量 ($i = 1, 2, \dots, p$)。设有 n 个样本数据,则构成一个 $n \times p$ 维的样本数据矩阵:

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}_{n \times p} = \begin{bmatrix} e_1^T \\ e_2^T \\ \vdots \\ e_n^T \end{bmatrix}_{n \times 1} =$$

$$[X_1 \ X_2 \ \dots \ X_p]_{1 \times p}$$

其中:数据矩阵 X 的每一行称为一个样本点,用 $e_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T \in R^p$ 表示 ($i = 1, 2, \dots, n$);每个样本点均由 p 个特性指标来描述,这些特征指标 $X_j \in R^n$ 称为数据变量 ($j = 1, 2, \dots, p$)。所有这样的样本点 e_i 构成的空间称为样本点空间 F ,这 n 个样本点组成了一个群点集合 N_i ^[5,6];所有变量 X_j 构成的空间称为变量空间 E ,变量的群点记为 N_j ^[5,6]。

对于一个多变量高维数据系统,某一方向上数据的方差越大,所反映的数据信息就越多。数据方差最大的方向被称为数据变异的极大方向。对高维系统降维处理的核心思想,就是省却数据变异不大的变量方向,文献[1,4]对此进行了比较深入的探讨,这里做简要介绍。

设原始数据矩阵 X 的变量为 X_1, X_2, \dots, X_p 。主成分分析的过程本质上就是对原坐标系进行平移和正交旋转变换,使得新坐标的原点与数据群点的重心重合,新坐标系的第一轴与数据变异的极大方向对应,新坐标系的第二轴与第一轴标准正交,并且对应于数据变异的第二大方向,以此类推。这些新轴分别被称为第一主轴 u_1 ,第二主轴 $u_2 \dots$ 。若舍弃少量信息后,主轴 ($m < p$) 能够十分有效地表示原数据的变异情况,则原来的 p 维空间就被降至 m 维。生成的空间 $U(u_1, u_2, \dots, u_m)$ 称为 m 维主超平面。当 $m = 2$ 时,就称其为主平面。可以用原样本群点 N_i 在主超平面上的投影来近似地表达原群点。

主成分分析能在数据信息损失最小的前提下,对高维变量空间进行降维处理,通过数据变异和相似性等信息实现对

收稿日期:2007-03-26;修回日期:2007-06-15。

作者简介:郭凯红(1973-),男,辽宁沈阳人,讲师,硕士,主要研究方向:管理决策支持、智能知识处理、人工智能。

数据的描述。数据变异信息是指原数据的全部变异信息,可以用点与点间的相对距离总和的平均水平来表示:

$$\sum_{i=1}^n \sum_{k=1}^n p_i p_k d^2(e_i, e_k) = \sum_{i=1}^n \sum_{k=1}^n p_i p_k \|e_i - e_k\|_M^2$$

其中: $e_j, e_k \in F, p_i, p_k$ 分别为 e_i, e_k 上的权重。

数据相似性同样是以样本群点间的相对距离来加以描述的。在主成分分析理论中可以证明,若试图在一个 m 维超平面上近似地表达原样本空间中的群点,则在所有的 m 维超平面中,只有主超平面可以使群点间的相似性改变最小,这样就能够主超平面上,通过观察样本点的相对位置判断高维空间中样本点间的相互关系。

2 PCA 对变量多重相关性的局限作用

假设在主成分分析之前,原始的 p 个变量 X_1, X_2, \dots, X_p 是存在多重相关性的,即 $Cov(X_j, X_k) \neq 0, j \neq k$ 。而主成分分析可以将原 p 个变量综合简化成 m 个新变量 Y_1, Y_2, \dots, Y_m ($m < p$)。已经证明,主成分存在以下性质^[1,4]:

- 1) 期望为零,即 $E(Y_i) = 0; i = 1, 2, \dots, m$ 。
- 2) 方差递减,即 $V(Y_i) = \lambda_i; i = 1, 2, \dots, m, V(Y_1) \geq V(Y_2) \geq \dots \geq V(Y_m)$ 。
- 3) 主成分 Y_1, Y_2, \dots, Y_m 是彼此无关的变量,即 $Cov(Y_j, Y_k) = 0; j \neq k$ 。

由此,一些文献指出,在变换后的新变量系统中,已消除了原变量系统中的多重相关性。然而,这种看法并没有把握主成分分析计算结果的本质。事实上,变量多重相关性必然从方向和数量两个方面扭曲真实的数据信息。变量多重相关度越高,经主成分分析处理后的扭曲程度就越大。为了证明的目的,假定变量处于高度多重相关状态。下面给出证明过程。

设某系统有重要性等同的两个独立特性指标 X_I 和 X_{II} 。对 X_I ,不妨用 $p-1$ 个完全相同的变量 $X_1 = X_2 = \dots = X_{p-1}$ 来描述,而对 X_{II} ,则用一个变量 X_p 来描述。记 $X = (X_1, X_2, \dots, X_p)^T$ 。设 X_i 均为标准化变量($i = 1, 2, \dots, p$),可得相关系数矩阵:

$$R = \begin{bmatrix} 1 & 1 & \dots & 1 & 0 \\ 1 & 1 & \dots & 1 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & 1 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}_{p \times p}$$

矩阵 R 的特征值为:

$$\lambda_1 = n - 1, \lambda_2 = 1, \lambda_3 = \lambda_4 = \dots = \lambda_p = 0$$

相应的特征向量为:

$$l_1 = \frac{1}{\sqrt{n-1}} \cdot \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \end{bmatrix} \quad l_2 = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

$$l_3 = \sqrt{\frac{1}{2}} \cdot \begin{bmatrix} -1 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \quad l_4 = \sqrt{\frac{2}{3}} \cdot \begin{bmatrix} -1 \\ 2 \\ -1 \\ 1 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

$$\dots \quad l_p = \sqrt{\frac{p-2}{p-1}} \cdot \begin{bmatrix} -1 \\ p-2 \\ -1 \\ p-2 \\ -1 \\ p-2 \\ \vdots \\ 1 \\ 0 \end{bmatrix}$$

其中: l_1, l_2 已经单位化, l_3, l_4, \dots, l_p 已经施密特正交化并单位化,由此构成特征矩阵:

$$L = [l_1 \quad l_2 \quad l_3 \quad \dots \quad l_p]$$

显然 L 为正交矩阵。记 $Y = (Y_1, Y_2, \dots, Y_p)^T$, 则主成分 $Y = L^T X$ 。

1) 对主变异方向的影响

对于主成分 Y ,它的前两个分量分别为:

$$Y_1 = l_1^T X = \frac{1}{\sqrt{n-1}}(X_1 + X_2 + \dots + X_{p-1})$$

$$Y_2 = l_2^T X = X_p$$

以累计贡献率为测度可知,舍弃 Y_3, Y_4, \dots, Y_n ,以二维变量 Y_1, Y_2 来表示原 p 维变量系统,其精度为 100%。

很显然, Y_1 为第一主成分,它对应了数据变异的最大方向。在此方向上,可反映 $(p-1)/p$ 的原数据信息,而 Y_2 则反映了 $1/p$ 的原数据信息。如果要以 $(p-1)/p$ 的精度选取新变量系统,唯有 Y_1 当选,而 Y_2 则被完全忽略掉。然而 $Y_2 = X_p$,恰好表示了第二个重要特征 X_{II} 。在分析前,假设 X_{II} 与 X_I 是相互独立且同等重要的。由此可见,由于对特性指标 X_I 采用重复变量 X_1, X_2, \dots, X_{p-1} 来描述,这一高度多重相关性造成数据系统的主要变异方向指向 X_I ,这与客观事实不符。

2) 对数量的影响

在主成分分析之前,设有任意两个样本点 $e_i, e_k \in R^p$, $e_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T, e_k = (x_{k1}, x_{k2}, \dots, x_{kp})^T$,它们间的欧氏距离为:

$$d^2(e_i, e_k) = (e_i - e_k)^T(e_i - e_k) = \sum_{j=1}^p (x_{ij} - x_{kj})^2 = \sum_{j=1}^{p-1} (x_{ij} - x_{kj})^2 + (x_{ip} - x_{kp})^2 = (p-1)(x_{iI} - x_{kI})^2 + (x_{iII} - x_{kII})^2$$

由于 X_I 采用重复信息,该特征指标在计算中被重复了 $p-1$ 次,其重要性被人为地加强。

在主成分分析之后,假设不舍弃任何数据信息,这两样本点 e_i, e_k 在新的变量系统中坐标轴上的投影为 $z_i, z_k \in R^p$,有

$$z_i = (y_{i1}, y_{i2}, \dots, y_{ip})^T = L^T e_i$$

$$z_k = (y_{k1}, y_{k2}, \dots, y_{kp})^T = L^T e_k$$

这时 z_i 与 z_k 的欧氏距离为:

$$d^2(z_i, z_k) = \sum_{j=1}^p (y_{ij} - y_{kj})^2 = (z_i - z_k)^T(z_i - z_k) = (L^T e_i - L^T e_k)^T(L^T e_i - L^T e_k) = (e_i - e_k)^T L L^T (e_i - e_k) = (e_i - e_k)^T(e_i - e_k) = (p-1)(x_{iI} - x_{kI})^2 + (x_{iII} - x_{kII})^2$$

从上式可以看出,经过主成分计算数据系统的重叠信息,从量的方面也未做任何实质性的改善。事实上,正如前所述,

主成分分析过程本质是对原坐标系进行正交旋转变换,而坐标系正交旋转变换最突出的优点是欧氏距离保持不变^[1,4]。这样,主成分分析后的样本点间的变异与相似情况(即它们间的距离)完全等同于变换前的情形。

3 PCA 对多目标系统评估的局限作用

在不进行任何条件限制的情况下,一些文献指出,可以采用多个主成分建立系统评估函数,以期待提高评估函数的精确度,即:

$$Z = \frac{1}{p} \sum_{i=1}^m \lambda_i Y_i$$

其中: Y_i 是主成分 Y 的第 i 个分量, λ_i 是样本协方差矩阵 X 的第 i 个特征值, p 为原数据变量的个数。

当原变量系统出现高度多重相关性时,这样建立的评估函数将具有双重夸大重叠信息的作用。现仍采用上节的例子证明之。

由前述可知,取主成分 Y 的前两个分量 Y_1, Y_2 来表示原 p 维变量系统,其精度可达 100%。按照上述方法建立评估函数,有

$$\begin{aligned} Z &= \frac{\lambda_1}{p} Y_1 + \frac{\lambda_2}{p} Y_2 = \\ &= \frac{n-1}{p} \cdot \frac{1}{\sqrt{n-1}} (X_1 + X_2 + \dots + X_{p-1}) + \frac{1}{p} \cdot X_p = \\ &= \frac{\sqrt{n-1}}{p} \cdot (p-1) \cdot X_I + \frac{1}{p} \cdot X_{II} = \\ &= \frac{p-1}{p} \cdot \sqrt{n-1} \cdot X_I + \frac{1}{p} \cdot X_{II} \end{aligned}$$

显然, X_I 的权重严格大于 X_{II} 的权重,即特性指标 X_I 的重要性被人为地加强了。

事实上,采用 $Y_1 = \frac{p-1}{\sqrt{n-1}} \cdot X_I$ 时,就已经包含了数据

的重叠信息,而由于 $\frac{\lambda_1}{p} = \frac{n-1}{p} > \frac{1}{p} = \frac{\lambda_2}{p}$, 所以,采用 $\frac{\lambda_1}{p}$ 作为 Y_1 的权重时,必然再一次夸大 X_I 在评估中的重要性。此前已经假定 X_I 和 X_{II} 是同等重要的,而这种评估结果与客观实际是相背的。显然,在一般情况下,使用多个主成分建立系统评估函数的方法是不可取的。

由于第一主成分 Y_1 对应于数据变异最大的方向,即 Y_1 是使数据信息损失最小、精度最高的一维综合变量,因此,它就有可能被用于构造系统评估函数^[7]。不仅如此,由于在 Y_1 对应的方向上样本点的散布被拉得最开,即样本点间的方差取到最大,在 Y_1 上样本点的性质差距也最大,因此也最容易判别出样本点的排列顺序和类级^[8],易于对它们进行排序评估。

然而, Y_1 并不是总能被用于评估函数^[9]。只有当所有评估指标变量(即 $X_j, j=1, 2, \dots, p$) 都正相关时,即所有变量均有同增、同减的趋势时, Y_1 才可被用于评估函数。关于这一点,文献[10,11]给出了较详细的论述和相关定义。

4 PCA 方法的改进及相关建议

由于传统的主成分分析是一种线性降维技术,具体表现为其主成分是原始变量的线性组合。因此,当原始数据不具备基本的线性特点时,主成分分析的降维效果便不明显。这时,为了提高主成分分析的降维效果,就要对原始数据作变

换^[12],一般可直接对它们进行函数处理:描绘原始数据列 x_{ij} 的散点图,若散点图呈现出某种曲线特征,如呈现出对数曲线特征时,则可令 $y_{ij} = \ln x_{ij}$,再对 y_{ij} 利用主成分分析方法,便可提高降维效果。

特别地,对于百分比数据 x ,因为 x 的值都在 0 与 1 之间,变化值不大,这时可令 $y = \ln \frac{x}{1-x}$,则 y 的变化范围就处在 $(-\infty, +\infty)$ 之间,利用主成分分析法就更容易处理;若原始数据列中,某些数据是愈大愈好,而另一些数据是愈小愈好,这时则应将它们转化一致。例如,可将愈大愈好的数据保留作为新数据,愈小愈好的数据则取倒数(事实上用一个递减函数作用它们即可)后作为新数据^[13],再利用主成分分析法处理新数据较为合理。

更进一步,如果观测变量不具备线性独立性这一假设,那么,可以使用诸如非线性主成分分析^[14]、分段主成分分析^[15,16]、独立成分分析^[17,18]、流形学习等方法^[19],以不同程度地解决主成分分析对变量多重相关性及多目标系统评估的局限作用。

综上所述,对具有非线性特征的原始数据列进行函数处理后,作为主成分分析的指标就是原始指标的函数,这样求出的主成分就出现非线性的形式。对原始数据列进行变换不仅会明显提高降维效果,用更少的主成分更多反映原始指标的信息,而且会使评价模型更具科学性。文献[20-24]分别介绍了改进的主成分分析方法在实际中的应用。

5 结语

显而易见,变量多重相关性造成的信息重叠,会片面强调某些变量的重要性,而使用主成分分析的计算结果不能完全消除重叠信息的作用。进而,当人为地采用一些无益的相关变量时,无论从方向上还是从数量上都会扭曲客观结论。因此,在主成分分析之前,对变量系统的确定必须采取谨慎的态度,尽可能地避免变量间出现多重相关性。应当注意,主成分分析的前提假定是观测变量具有线性独立性,在此前提下可以通过谱方法来发现内在的能够保持数据方差的最大结构部分,即基于主成分的约简。对于观测变量不具备线性独立性这一假设,也提出了主成分分析的改进方法和相关建议。同时,在使用主成分分析进行系统评估时,应使用第一主成分建立评估函数,从而保证系统评估的结果真实有效。

参考文献:

- [1] RENCHER AC. Methods of multivariate analysis [M]. New York: Wiley & Sons, 1995: 439-498.
- [2] 高惠璇. 两个多重相关变量组的统计分析: 2 [J]. 数理统计与管理, 2002, 21(2): 60-64.
- [3] 高惠璇. 两个多重相关变量组的统计分析: 1 [J]. 数理统计与管理, 2002, 21(1): 57-64.
- [4] 高惠璇. 应用多元统计分析[M]. 北京: 北京大学出版社, 2005: 187-199.
- [5] LIAO H C. Multi-response optimization using weighted principal component [J]. The International Journal of Advanced Manufacturing Technology, 2006, 27(1): 7-8.
- [6] MELLOUL A J. Use of principal components analysis for studying deep aquifers with scarce data-application to the Nubian sandstone aquifer, Egypt and Israel [J]. Hydrogeology Journal, 1995, 3(2): 2-3.
- [7] 孟生旺. 用主成分分析法进行多指标综合评价注意的问题[J]. 统计研究, 1992, (4): 66-68. (下转第 2352 页)

key_width) 进行评价。如果剩余面积不满足适应度函数要求, 则放弃对该按键的布局, 即不放置此类按键。根据优化结果显示各个按键, 如图 4 手机组件初步布局。

3) 用户还可以通过交互方式, 在智能布局面板中对待布局手机组件的参数修改: 外壳序号 ($shell_num$), 键盘序号 (key_num), 屏幕面积比例 ($screen_scale$), 布局方式 ($flag$), 如图 5 用户输入参数模块。相应的参数 t 取值发生改变, 如果其取值落在不同区间, 则自动修改式(4)中的参数 α 、 β 、 γ , 实现算法的自适应调整, 如图 6 改变参数后布局结果。

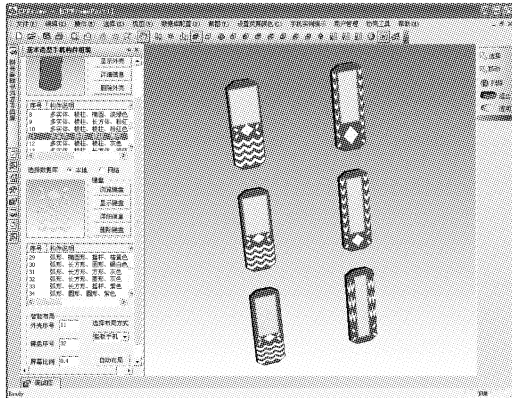


图 7 竖版手机布局方案实例

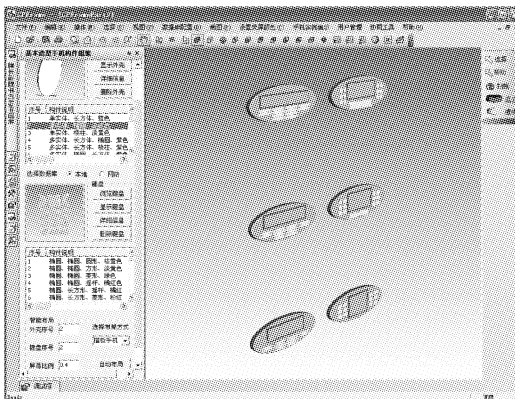


图 8 横版手机布局方案实例

根据记录的优化结果, 按照组件的位置参数 (x , y) 和大小参数 ($length$, $width$), 将满足优化条件的手机造型布局方案进行批量显示。系统设定记录前 6 个最优方案, 生成结果

如图 7、图 8 所示横版、竖版布局方案实例。用户可以根据其位置序号选择其中较满意的设计方案, 通过改变参数对其进行继续优化或修改。

4 结语

在手机组件的布局问题中, 利用微粒群算法求解可以得到较满意的布局方案, 并且得到的手机外观设计方案有一定的创新性, 可见 PSO 算法对于计算机辅助创新设计系统的开发是一个非常有利的工具。

结合 PSO 的具体应用情况来看, 还有不少问题需做进一步研究, 如多微粒之间的协同进化、设计方案的智能评价和优化筛选等问题。这些问题的解决, 将对 PSO 算法性能的提高和新的计算机辅助创新设计系统的开发有重要的意义。

参考文献:

- [1] KENNEDY J, EBERHART R C. Particle swarm optimization[C]// Proceeding IEEE International Conference on Neural Networks. [S. l.]: IEEE Press, 1995.
- [2] ZHANG B, TENG H F. Human-computer cooperative design method with modified particle swarm optimization[C]// Processing of the 6th International Conference on Frontiers of Design and Manufacturing. Beijing, Science Press, 2004: 125 - 127.
- [3] 刘飞, 孙明, 李宁, 等. 粒子群算法及其在布局优化中的应用[J]. 计算机工程与应用, 2004, 40(12): 71 - 73.
- [4] 袁苗龙, 周济, 张新访. 三维几何布局的一类启发式求解算法[J]. 计算机学报, 1999, 22(9): 923 - 930.
- [5] 吴一帆, 李思昆. 基于神经网络的自动布局算法[J]. 计算机工程与科学, 2000, 22(2): 77 - 79.
- [6] CAGAN J. Shape annealing solution to the constrained geometric knapsack problem [J]. Computer Aided Design, 1994, 26(10): 763 - 770.
- [7] 钱志勤, 滕弘飞, 孙志国. 人机交互的遗传算法及其在约束布局优化中的应用[J]. 计算机学报, 2001, 24(5): 553 - 559.
- [8] CLERC M, KENNEDY J. The particle swarm-explosion, stability, and convergence in a multidimensional complex space [J]. IEEE Transaction on Evolutionary Computer, 2002, 6(1): 58 - 73.
- [9] 孙守迁, 包恩伟, 潘云鹤. 面向产品布局设计的组件特征模型[J]. 计算机辅助设计与图形学学报, 1999, 11(1): 21 - 25.
- [10] 熊云飞, 刘刚. 手机键盘布局设计的人机工程实验研究[J]. 包装工程, 2006, 27(2): 171 - 174.
- [11] 阎慈琳. 关于主成分分析做综合评价的若干问题[J]. 数理统计与管理, 1998, 17(2): 22 - 25.
- [12] 阮敏. 主成分方法在经济管理综合评价应用中的误区[J]. 统计与决策, 2005, (4): 23 - 24.
- [13] 任若恩, 王惠文. 多元统计数据分析[M]. 北京: 国防工业出版社, 1997: 125 - 141.
- [14] 彭勇行. 管理决策分析[M]. 北京: 科学出版社, 2000: 222 - 229.
- [15] 黄宁. 关于主成分分析应用的思考[J]. 数理统计与管理, 1999, 18(5): 44 - 52.
- [16] 何斌, 蒙清. 主成分分析法的几个注记[J]. 云南师范大学学报: 自然科学版, 2002, 22(2): 6 - 8.
- [17] 钱道翠. 关于主成分分析的进一步探讨[J]. 嘉兴学院学报, 2002, 14(3): 66 - 68.
- [18] 徐兵, 邓群钊. 分块加权主成分分析法[J]. 南昌大学学报: 理科版, 2003, 27(3): 211 - 215.
- [19] 陈伏兵, 韦相和, 严云洋, 杨静宇. 分块二维主成分分析鉴别特征抽取能力研究[J]. 计算机工程与应用, 2006, (27): 69 - 75.
- [20] 刘洋, 吴新杰. 独立成分分析方法的研究及应用[J]. 沈阳教育学院学报, 2006, 8(1): 125 - 126.
- [21] 张建明, 林亚平. 独立成分分析的研究进展[J]. 系统仿真学报, 2005, 18(6): 992 - 1001.
- [22] 徐蓉, 姜峰, 姚鸿勋. 流形学习概述[J]. 智能系统学报, 2000, 1(1): 44 - 50.
- [23] FIORI S. An experimental comparison of three PCA neural networks [J]. Neural Processing Letters, 2000, 11(6): 13 - 15.
- [24] 姬春煦, 张骏. 基于主成分分析的股票指数预测研究[J]. 计算机工程与科学, 2006, 28(8): 122 - 124.
- [25] 李三平, 岳振军. 基于小波分析和非线性 PCA 的图像压缩算法[J]. 系统工程与电子技术, 2005, 27(9): 1532 - 1535.
- [26] 王文博, 陈秀芝. 多指标综合评价中主成分分析和因子分析方法的比较[J]. 统计与信息论坛, 2006, 21(5): 19 - 22.
- [27] 李成友, 王学周, 王芳, 等. 主成分分析法在网络教学评价系统中的应用[J]. 计算机应用研究, 2003, (12): 128 - 133.

(上接第 2348 页)