学 术 论 文

提升 KPCA 方法特征抽取效率的算法设计

徐 勇，杨静宇，陆建峰
（南京理工大学计算机科学与技术系，南京 210094）

[摘要] 在 PCA 基础上发展出的 KPCA 方法能抽取样本的非线性特征分量。然而，基于 KPCA 的特征抽取需计算所有训练样本与待抽取特征的样本间的核函数，因此，训练集的大小制约着特征抽取的效率。为了提高效率，假设特征空间中变换轴可由一部分训练样本（节点）线性表出，并设计了改进的 KPCA 算法（IKPCA）。该算法抽取某样本特征时，只需计算该样本与节点间的核函数即可。实验结果显示，IKPCA 在对应较好性能的同时，具有明显的效率上的优势。

[关键词] KPCA；IKPCA；特征抽取；特征空间
[中图分类号] TP391 [文献标识码] A [文章编号] 1009－1742（2005）10－0038－05

1 引言

作为线性方法，PCA（主分量分析）方法是最小均方误差意义上的最优维数压缩技术[1]。这种方法基于数据的二阶统计信息（基于相应协方差矩阵）进行分析，抽取不相关的各个特征分量. 应用中，PCA 方法可通过求解特征方程实现，并选择对应较大特征值的特征向量作为变换轴。

如果原始数据的特征存在复杂的非线性关系，相比主分量分析，非线性主分量分析更适合用作特征抽取方法[2]。KPCA（核主分量分析）[3]是成功的一种 NPCA（非线性主分量分析）方法。相比一般的 NPCA 方法，KPCA 并不需要直接对样本数据进行非线性映射，因而其实现是简洁高效的。KPCA 方法广泛地应用于特征抽取，人脸识别，图像处理等问题。如同其他核方法（譬如 KFDA，KMSE 等）一样，基于 KPCA 方法对某样本进行特征抽取时，需计算该样本与所有训练样本间的核函数；训练样本集越大，相应计算量也越大，效率也越低[4~8]，而很多实际的模式分类任务要求系统具有较高的效率。从应用角度分析，有必要对 KPCA 方法进行改进，以提高其效率。笔者假设特征空间中的变换轴可由一部分训练样本线性表出，并据此发展了一种改进的 KPCA（IKPCA）方法。

2 PCA 与 KPCA

2.1 PCA

PCA 方法又称为 K－L 变换，其表述如下[9, 10]。

设 $\boldsymbol{X}$ 是一个 n 维的随机变量，$u_1, u_2, \cdots, u_n$ 是 n 维空间的正交归一化矢量系，即

$$(u_i)^{\mathrm{T}}(u_j) = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases} \tag{1}$$

则可将 $\boldsymbol{X}$ 无误差地表示为

$$X = \sum_{i=1}^{n} y_i u_i \tag{2}$$

其中 $y_i = (u_i)^{\mathrm{T}}\boldsymbol{X}$，$i = 1, 2, \cdots, n$，若用前 r 项估计 $\boldsymbol{X}$，即

$$\hat{X} = \sum_{i=1}^{r} y_i u_i \tag{3}$$

则由此引起的均方误差为

[收稿日期] 2004－09－21；修回日期 2004－11－10
[基金项目] 国家自然科学基金资助项目（60072034）
[作者简介] 徐 勇（1972－），男，四川简阳市人，南京理工大学计算机科学与技术系博士生

$$\varepsilon = \sum_{i=r+1}^{n} E(y_i^2) \tag{4}$$

亦即

$$\varepsilon = \sum_{i=r+1}^{n} (u_i)^{\mathrm{T}} E(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})(u_i) = \sum_{i=r+1}^{n} (u_i)^{\mathrm{T}} \sum (u_i) \tag{5}$$

使用拉格朗日乘子法，可以求出在满足正交归一化条件式（1）下，使得均方误差 ε 取极值的坐标系统。换言之，拉格朗日函数

$$g = \sum_{i=r+1}^{n} (u_i)^{\mathrm{T}} \boldsymbol{\Sigma} (u_i) - \sum_{i=r+1}^{n} \lambda_i ((u_i)^{\mathrm{T}} (u_i) - 1) \tag{6}$$

取得极值时，式（5）也取得极值。而式（6）取得极值应满足条件

$$(\boldsymbol{\Sigma} - \lambda_i \boldsymbol{I})(u_i) = 0 \quad (i = r+1, \cdots, n) \tag{7}$$

令 $r=0$，则可得到如下结论：

以矩阵 $\boldsymbol{\Sigma}$ 的特征向量作为坐标轴来展开 $\boldsymbol{X}$ 时，其截断均方误差具有极值性质，且当取 r 个 u_i（$i=1, \cdots, r$）来表示 $\boldsymbol{X}$ 时，其均方误差为

$$\varepsilon = \sum_{i=r+1}^{n} \lambda_i \tag{8}$$

式中，λ_i 是矩阵 $\boldsymbol{\Sigma}$ 的相应特征值。因此，PCA 方法一般求出协方差矩阵较大特征值对应的特征向量作为变换轴。

2.2　PCA 中核函数的引入

假设 $x_1, x_2, \cdots, x_N$ 为训练样本，用 $\{x_i\}$ 表示输入空间。KPCA 方法的基本思想是通过某种隐式方式将输入空间映射到某个高维空间（常称为特征空间），并在特征空间中实现 PCA。假设相应的映射为 $\boldsymbol{\phi}$，并且由此映射而得的特征空间中数据满足中心化的条件，即

$$\sum_{i=1}^{N} \boldsymbol{\phi}(x_i) = 0 \tag{9}$$

则特征空间中协方差矩阵为

$$\boldsymbol{\Sigma} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\phi}(x_i) \boldsymbol{\phi}(x_i)^{\mathrm{T}} \tag{10}$$

对不满足中心化的情况，可参考文献 [3]。可认为特征空间中最小均方误差意义上的最优变换轴 u_i 必为所有样本的线性组合，也称 u_i 位于 $\phi(x_1)$，$\phi(x_2)$，…，$\phi(x_N)$ 张成的子空间中，并表示为[3]

$$u_i = \sum_{j=1}^{N} \alpha_j^{(i)} \boldsymbol{\phi}(x_j) \tag{11}$$

联合式（6）与式（11），令 $r=0$，得

$$g = \frac{1}{N} \sum_{i=1}^{N} (\alpha^{(i)})^{\mathrm{T}} \boldsymbol{K}\boldsymbol{K}^{\mathrm{T}} \alpha^{(i)} - \sum_{i=1}^{N} \lambda_i (\alpha^{(i)})^{\mathrm{T}} \boldsymbol{K} \alpha^{(i)} + \sum_{i=1}^{N} \lambda_i \tag{12}$$

矩阵 $\boldsymbol{K}$ 中元素为 Mercer 核，即 $(\boldsymbol{K})_{ij} = k(x_i, x_j) = \boldsymbol{\phi}(x_i) \boldsymbol{\phi}(x_j)$。由于此处 $\boldsymbol{K}$ 为对称阵，将式（12）对 $\alpha^{(i)}$ 求导，可得

$$\boldsymbol{K}^2 \alpha^{(i)} / N = \lambda_i \boldsymbol{K} \alpha^{(i)} \tag{13}$$

式（13）可改写为特征方程

$$\boldsymbol{K}\alpha = \lambda' \alpha \tag{14}$$

其中 $\lambda' = N\lambda_i$。

2.3　基于 KPCA 的特征抽取

在特征空间的训练样本集 $\{\boldsymbol{\phi}(x_i)\}$ 上计算特征方程式（14）的非零特征值与相应特征向量。假设按降序排列的若干个较大非零特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_m$（$m \leqslant N$），相应特征向量为 $\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \cdots, \boldsymbol{\alpha}^{(m)}$，并假设特征空间中相应单位特征向量分别为 $\boldsymbol{u}^{(1)}, \boldsymbol{u}^{(2)}, \cdots, \boldsymbol{u}^{(m)}$，则关系式

$$\boldsymbol{u}^{(i)} \triangleq \lambda_i^{-1/2} \sum_{j=1}^{N} \alpha_j^{(i)} \boldsymbol{\phi}(x_j), \quad i = 1, 2, \cdots, m \tag{15}$$

成立。据此容易给出特征空间中样本 $\boldsymbol{\phi}(\boldsymbol{x})$ 在 $\boldsymbol{u}^{(i)}$ 上投影的计算式。若将 $\boldsymbol{\phi}(\boldsymbol{x})$ 在 m 个特征向量上的投影值组成矢量，则可得

$$y = \lambda_1^{-1/2} \Big[\sum_{j=1}^{N} \alpha_j^{(1)} k(x_j, x), \lambda_2^{-1/2} \sum_{j=1}^{N} \alpha_j^{(2)} k(x_j, x), \cdots, \lambda_m^{-1/2} \sum_{k=1}^{N} \alpha_j^{(m)} k(x_j, x) \Big]^{\mathrm{T}} \tag{16}$$

在实际应用中，可根据不同情况选定 m 值，m 也称为 KPCA 方法的主分量数。分类可基于式（16）给出的样本特征进行。

3　提升 KPCA 特征抽取效率的算法设计

3.1　提高 KPCA 方法特征抽取效率的思路

KPCA 算法认为特征空间中的特征向量位于 $\phi(x_1)$，$\phi(x_2)$，…，$\phi(x_N)$ 张成的子空间中，特征抽取由式（16）规定。式（16）表明，为了得出一个样本在特征空间的每一个特征分量，均需计算该样本与所有训练样本间的核函数，并做累加。若训练样本集较大，特征抽取会对应很大的计算量，使得基于 KPCA 的特征抽取效率较低，其他核方法

也面临这一问题。而在一些实时性要求很强的应用中，算法的高效性是必须考虑的重要因素。鉴于此，提升KPCA方法特征抽取效率很有重要意义。

假设，在特征空间中训练样本集中一部分样本的线性组合即可较好地表示主分量（或者称逼近主分量）。换个角度分析，虽然所有样本的线性组合可准确描述式（15）中的特征向量$\boldsymbol{u}^{(i)}$；但可认为特征空间中不同样本在该线性组合中的重要性不完全相同，其中一部分样本的贡献较大，而另一些则相反。假如能从所有样本中找出比较重要的那部分样本，并以其线性组合的形式给出特征空间中主分量的近似表示式，则可减少基于KPCA的特征抽取的计算代价。相似的思路曾在KFD（核Fisher鉴别分析）中得到很好的应用[11, 12]。

根据PCA方法的特点，提出按照相应特征值的大小判断训练样本的重要程度，PCA方法中特征值越大，相应主分量对原数据的描述能力越强，由此抽取出的特征包含原数据的信息越多。

假设

$$u_i = \sum_{j=1}^{s} \alpha_j^{(i)} \phi(x_j^{(0)}), s < N \tag{17}$$

称$\phi(x_j^{(0)})$为特征空间中的节点。令$\boldsymbol{\alpha}^{(i)} \triangleq [\alpha_1^{(i)} \alpha_2^{(i)} \cdots \alpha_N^{(m)}]^{\mathrm{T}}$。此时式（12）相应变形为

$$g = \frac{1}{N}\sum_{i=1}^{N} (\boldsymbol{\alpha}^{(i)})^{\mathrm{T}} \boldsymbol{K}_1 (\boldsymbol{K}_1)^{\mathrm{T}} \boldsymbol{\alpha}^{(i)} - \sum_{i=1}^{N} \boldsymbol{\lambda}_i (\boldsymbol{\alpha}^{(i)})^{\mathrm{T}} \boldsymbol{K}_2 \alpha^{(i)} + \sum_{i=1}^{N} \lambda_i \tag{18}$$

其中

$$\boldsymbol{K}_1 = \begin{bmatrix} k(x_1^{(0)}, x_1) & \cdots & k(x_1^{(0)}, x_N) \\ k(x_2^{(0)}, x_1) & \cdots & k(x_2^{(0)}, x_N) \\ \vdots & & \vdots \\ k(x_s^{(0)}, x_1) & \cdots & k(x_s^{(0)}, x_N) \end{bmatrix}$$

$$\boldsymbol{K}_2 = \begin{bmatrix} k(x_1^{(0)}, x_1^{(0)}) & \cdots & k(x_1^{(0)}, x_s^{(0)}) \\ k(x_2^{(0)}, x_1^{(0)}) & \cdots & k(x_2^{(0)}, x_s^{(0)}) \\ \vdots & & \vdots \\ k(x_s^{(0)}, x_1^{(0)}) & \cdots & k(x_s^{(0)}, x_s^{(0)}) \end{bmatrix}$$

将式（18）对$\boldsymbol{\alpha}^{(i)}$求导，可得广义特征方程

$$\boldsymbol{K}_1 (\boldsymbol{K}_1)^{\mathrm{T}} \boldsymbol{\alpha}^{(i)} / N = \lambda_i \boldsymbol{K}_2 \boldsymbol{\alpha}^{(i)} \tag{19}$$

$\boldsymbol{K}_2$可逆条件下，令$\lambda' = N\lambda_i$，该特征方程可改写为如下等价形式

$$(\boldsymbol{K}_2)^{-1} \boldsymbol{K}_1 (\boldsymbol{K}_1)^{\mathrm{T}} \alpha = \lambda'_i \alpha \tag{20}$$

若$\boldsymbol{K}_2$非可逆，则可采用$(\boldsymbol{K}_2 + \mu \boldsymbol{I})^{-1} \boldsymbol{K}_1 (\boldsymbol{K}_1)^{\mathrm{T}} \boldsymbol{\alpha} = \lambda'_i \boldsymbol{\alpha}$求解$\boldsymbol{\alpha}$，其中$\boldsymbol{I}$为单位矩阵，$\mu$为正常数。

3.2 算法设计

Step 1 选出第一个节点：

对单个训练样本x_i，$i = 1, 2, \cdots, N$，首先计算相应$\boldsymbol{K}_1$，$\boldsymbol{K}_2$。根据上面相应公式可知，在该步骤中，$\boldsymbol{K}_2$为一数量值，$\boldsymbol{K}_1 (\boldsymbol{K}_1)^{\mathrm{T}}$也为一数量值。计算$\lambda_i = \boldsymbol{K}_1 (\boldsymbol{K}_1)^{\mathrm{T}} / \boldsymbol{K}_2$。将对应最大$\lambda_i$的样本作为第一个节点，并记为$x_1^{(0)}$。将对应$x_1^{(0)}$的$\boldsymbol{K}_1$，$\boldsymbol{K}_2$分别记为$\boldsymbol{K}_1^{(0)}$，$\boldsymbol{K}_2^{(0)}$。

Step 2 选出第二个节点：

称

$$k_j^{(1)} = [k(x_j, x_1), k(x_j, x_2), \cdots, k(x_j, x_N)] \tag{21}$$

为样本x_j（$x_j \neq x_1^{(0)}$）的核向量，且与$x_1^{(0)}$，x_j对应的矩阵$\boldsymbol{K}_1$，$\boldsymbol{K}_2$分别为

$$\boldsymbol{K}_1 = \begin{bmatrix} K_1^{(0)} \\ k_j^{(1)} \end{bmatrix}, \quad \boldsymbol{K}_2 = \begin{bmatrix} K_2^{(0)} & k(x_1^{(0)}, x_j) \\ k(x_1^{(0)}, x_j) & k(x_j, x_j) \end{bmatrix}。$$

计算相应特征方程式（19）的特征值λ_1，λ_2，令$v = \lambda_1 + \lambda_2$。考察完所有满足条件的样本后，将对应最大$v$值的样本选作第二个节点，并记为$x_2^{(0)}$。将$x_1^{(0)}$，$x_2^{(0)}$对应的矩阵$\boldsymbol{K}_1$，$\boldsymbol{K}_2$分别记为$\boldsymbol{K}_1^{(0)}$，$\boldsymbol{K}_2^{(0)}$。

Step 3 选出第三个节点：

令样本$x_1^{(0)}$，$x_2^{(0)}$，x_j（$x_j \neq x_1^{(0)}$，$x_2^{(0)}$）对应矩阵$\boldsymbol{K}_1$，$\boldsymbol{K}_2$为

$$\boldsymbol{K}_1 = \begin{bmatrix} K_1^{(0)} \\ k_j^{(1)} \end{bmatrix}, \quad \boldsymbol{K}_2 = \begin{bmatrix} K_2^{(0)} & (k_j^{(2)})^{\mathrm{T}} \\ k(x_j^{(2)}) & k(x_j, x_j) \end{bmatrix},$$

其中$(k_j^{(2)}) = [k(x_j, x_1^{(0)}) \; k(x_j, x_2^{(0)})]$，$k_j^{(1)}$同式（21）。计算相应特征方程式（19）的特征值$\lambda_1$，$\lambda_2$，$\lambda_3$，令$v = \lambda_1 + \lambda_2 + \lambda_3$。考察完所有满足条件的样本后，将对应最大$v$值的样本选作第三个节点，并记为$x_3^{(0)}$。将$x_1^{(0)}$，$x_2^{(0)}$，$x_3^{(0)}$对应的矩阵$\boldsymbol{K}_1$，$\boldsymbol{K}_2$分别记为$\boldsymbol{K}_1^{(0)}$，$\boldsymbol{K}_2^{(0)}$。

Step s 选出第s个节点：

假设已有$s-1$个节点$x_1^{(0)}$，$x_2^{(0)}$，…，$x_{s-1}^{(0)}$被选出，与其对应的矩阵$\boldsymbol{K}_1$，$\boldsymbol{K}_2$分别记为$\boldsymbol{K}_1^{(0)}$，$\boldsymbol{K}_2^{(0)}$。对样本x_j（$x_j \neq x_1^{(0)}$，$x_2^{(0)}$，…，$x_{s-1}^{(0)}$）仍令

$$\boldsymbol{K}_1 = \begin{bmatrix} K_1^{(0)} \\ k_j^{(1)} \end{bmatrix}, \quad \boldsymbol{K}_2 = \begin{bmatrix} K_2^{(0)} & (k_j^{(2)})^{\mathrm{T}} \\ k(x_j^{(2)}) & k(x_j, x_j) \end{bmatrix},$$

其中$(k_j^{(2)})=[k(x_j, x_1^{(0)})\ k(x_j, x_2^{(0)})\ \cdots\ k(x_j, x_{s-1}^{(0)})]$，$k_j^{(1)}$同式（21）。计算相应特征方程式（19）的所有特征值 $\lambda_1, \lambda_2, \cdots, \lambda_s$。若$s \leqslant p$，令 $v=\lambda_1+\lambda_2+\cdots+\lambda_s$；若 $s>p$，令 $v=\lambda_1+\lambda_2+\cdots+\lambda_p$（$p$ 为主分量分析中选取的特征向量数目）。考察完所有满足条件的样本后，将对应最大 v 值（记为 v_s）的样本选作第 s 个节点，并记为 $x_s^{(0)}$。将 $x_1^{(0)}, x_2^{(0)}, \cdots, x_s^{(0)}$ 对应的矩阵 $\boldsymbol{K}_1$，$\boldsymbol{K}_2$ 分别记为 $\boldsymbol{K}_1^{(0)}$，$\boldsymbol{K}_2^{(0)}$。重复该步骤，当条件 $s \geqslant Nr$（r 为小于 1 的系数，N 为训练样本总数）满足时终止节点的选择过程。

节点选择完毕后，特征空间中样本$\phi(x)$的特征抽取可按下式进行

$$y=\left[\lambda_1^{-1/2}\sum_{j=1}^{s}\alpha_j^{(1)}k(x_j^{(0)},x),\ \lambda_2^{-1/2}\sum_{j=1}^{s}\alpha_j^{(2)}k(x_j^{(0)},x),\cdots,\ \lambda_m^{-1/2}\sum_{j=1}^{s}\alpha_j^{(m)}k(x_j^{(0)},x)\right]^{\mathrm{T}} \tag{22}$$

4　实验

实验在 4 个基准数据集上进行。每个数据集被随机地分成了 100 部分（S 除了 plice 除了数据集只包含 20 部分外），每部分又分别包含训练样本子集与测试样本子集。实验采用高斯型核函数 $k(x, y)=\exp(-\|x-y\|^2/2\sigma^2)$。每次实验中，将 σ^2 取为第一个训练样本子集协方差矩阵 $\boldsymbol{F}$ - 范数的二次方。在第一个训练样本子集上进行训练，然后对所有测试样本子集进行分类。除了关于数据集 Splice 的实验中 r 取 0.25 外，其他数据集的 $r=0.5$。对测试样本的分类使用最小距离分类器。

表 1 与表 2 分别给出了 KPCA 与 IPKCA 在 4 个基准数据集上的实验结果。实验得出的特征抽取时间显示，改进的 KPCA 方法的特征抽取效率大大高于 KPCA 方法。对 4 个数据集的分类结果显示，在 Diabetis 与 Cancer 两个数据集上，基于两类方法的分类错误率相当；在 Banana 数据集上，基于改进的 KPCA 方法的分类错误率略高于基于 KPCA 方法的分类错误率；在训练数据多达1 000 的数据集 Splice 上，基于改进的 KPCA 方法的分类错误率明显低于 KPCA 方法。需要说明的是，错误率一项中的 2 个数据分别为平均分类错误率与分类错误率标准差。

表 1　KPCA 在基准数据集上的实验结果

Table 1　KPCA experimental result on benchmark datasets

	主分量数	错误率	训练样本总数	特征抽取时间/s
Splice	100	25.5±2.6	1 000	864
	90	24.7±2.5		832
	80	24.0±2.4		801
	70	21.8±2.2		778
Diabetis	100	11.5±2.8	468	238
	90	11.7±2.8		212
	80	11.5±2.8		202
	70	11.8±2.9		191
Banana	100	13.8±0.2	400	3 128
	90	13.8±0.2		2 983
	80	13.8±0.2		2 908
	70	13.8±0.2		2 825
Cancer	70	9.0±3.2	200	25.4
	60	8.5±3.0		23.2
	50	8.5±3.0		22.2
	40	9.8±3.3		21.8

表 2　改进的 KPCA 在基准数据集上的实验

Table 2　Improved PCA experimental result on benchmark datasets

	主分量数	错误率	节点数	特征抽取时间/s
Splice	100	17.8±1.8	250	260
	90	17.8±1.8		247
	80	17.5±1.8		230
	70	17.6±1.7		214
Diabetis	100	11.4±2.8	234	140
	90	11.9±2.9		125
	80	11.6±2.8		121
	70	12.0±2.9		114
Banana	100	14.1±0.2	200	1 807
	90	14.2±0.2		1 799
	80	14.2±0.2		1 734
	70	14.2±0.2		1 688
Cancer	70	9.2±3.3	100	15.3
	60	8.6±2.9		13.9
	50	8.6±2.9		13.2
	40	8.1±2.9		12.7

5　结论

作为一类核方法，KPCA 方法在特征抽取中得到了较多的应用。由于 KPCA 抽取一个样本的特征时，需计算训练集中所有样本与该样本间的核函数，特征抽取的效率会随着训练集的增大而减小。另一方面，实际应用中往往要求系统有较高的特征

抽取效率。假定特征空间中 KPCA 对应的变换轴可由一部分训练样本（节点）线性表出，并设计了一个改进的 KPCA（IKPCA）算法。IKPCA 算法只基于所有节点与某样本间的核函数，即可抽取该样本特征。因此，IKPCA 抽取特征的效率与节点数的多少直接相关，节点数越少，特征抽取效率越高。在基准数据集上进行的 KPCA 与 IKPCA 的对比实验显示，IKPCA 方法对应较高的特征抽取效率，而且在此基础上的分类正确率与基于 KPCA 方法所抽取特征的分类正确率相当。

参考文献

[1] 边肇祺，张学工. 模式识别[M]. 北京：清华大学出版社，2000

[2] Duda R O, Hart P E, Stork D G. 模式分类[M]. 北京:机械工业出版社，中信出版社，2003

[3] Scholkopf B, Smola A, Müller K R. Nonlinear component analysis as a kernel eigenvalue problem [J]. Neural Computation, 1998, 10(5): 1299～1319

[4] Mika S, Rätsch G, Weston J, Schölkopf B, Müller K R. Fisher discriminate analysis with kernels [A]. In: Hu Y H, Larsen J, Wilson E, Douglas S, eds. Neural Networks for Signal Processing IX, IEEE [C]. 1999. 41～48

[5] Mika S, Smola A J, Schölkopf B. An improved training algorithm for kernel fisher discriminants [A]. In: Jaakkola T, Richardson T, eds. Proceedings AISTATS [C]. Morgan Kaufmann, 2001. 98～104

[6] 徐 勇，杨静宇，金 忠，娄 震. 一种基于核的快速非线性鉴别分析方法[J]. 计算机研究与发展，2004，(1)

[7] Billings S A, Lee K L. Nonlinear fisher discriminant analysis using a minimum squared error cost function and the orthogonal least squares algorithm [J]. Neural Networks, 2002, 15(2): 263～270

[8] Xu J, Zhang X, Li Y. Kernel MSE algorithm: a unified framework for KFD, LS-SVM and KRR [A]. In: Proceedings of the International Joint Conference on Neural Networks (IJCNN-2001) [C]. Washington, DC, 2001. 1486～1491

[9] 福永圭之介. 统计图形识别导论[M]. 陶笃纯译. 北京:科学出版社，1978

[10] 金 忠. 人脸图像特征抽取与维数研究[D]. 南京：南京理工大学，1999

[11] Xu Yong, Yang Jingyu, Yang Jian, A reformative kernel Fisher discriminant analysis [J]. Pattern Recognition, 2004, 37: 1299～1302

[12] Xu Yong, Yang Jingyu, Lu Jianfeng, Yu Dongjun, An efficient renovation on kernel Fisher discriminant analysis and face recognition experiments [J]. Pattern Recognition, 2004, 37: 2091～2094

Algorithm Design for Improving Feature Extraction Efficiency Based on KPCA

Xu Yong, Yang Jingyu, Lu Jianfeng

(*Department of Computer Science & Technology, Nanjing University of Science & Technology, Nanjing* 210094, *China*)

[**Abstract**] KPCA (kernel PCA) is derived from PCA. It can extract nonlinear feature components of samples. However, feature extraction for one sample requires that kernel functions between training samples and the sample be calculated in advance. So, the size of training sample set affects the efficiency of feature extraction. It is supposed that in feature space the eigenvectors may be linearly expressed by a part of training samples, called nodes. According to the supposition, an improved KPCA (IKPCA) algorithm is developed. IKPCA extracts feature components of one sample efficiently, only based on kernel functions between nodes and the sample. Experimental results show that IKPCA is very close to KPCA in performance, while with higher efficiency.

[**Key words**] KPCA(Kernel PCA); IKPCA(Improved KPCA); feature extraction; feature space