计算机研究与发展
Journal of Computer Research and Development

ISSN 1000-1239/CN 11-1777/TP
42(3): 367～374, 2005

一种基于核的快速非线性鉴别分析方法

徐 勇 杨静宇 金 忠 娄 震
(南京理工大学计算机科学与技术系 南京 210094)
(laterfall@sohu.com)

A Fast Kernel-Based Nonlinear Discriminant Analysis Method

Xu Yong, Yang Jingyu, Jin Zhong, and Lou Zhen
(*Department of Computer Science & Technology, Nanjing University of Science & Technology, Nanjing* 210094)

Abstract The least squares solution of novel discriminant analysis method, based on kernel trick, is equivalent to kernel-based Fisher discriminant analysis. The discriminant vector of the novel method is efficiently solved from linear equations. Moreover, corresponding classifying strategy is very simple. The most striking advantage of the novel method is that only a few original training samples are sorted as "significant" nodes for constructing discriminant vector. As a result, corresponding testing is much more efficient than the naïve kernel Fisher discriminant analysis. In addition an appropriative, optimized algorithm is developed to improve the efficiency of selecting "significant" nodes. Experiments on benchmarks and face databases show that the performance of the novel method is comparative to kernel-based Fisher discriminant analysis, with superiority in efficiency.

Key words kernel-based Fisher discriminant analysis; kernel-based nonlinear discriminant analysis; least squares solution; feature extraction

摘 要 基于"核技巧"提出的新的非线性鉴别分析方法在最小二乘意义上与基于核的Fisher鉴别分析方法等效,相应鉴别方向通过一个线性方程组得出,计算代价较小,相应分类实现极其简便.该方法的最大优点是,对训练数据进行筛选,可使构造鉴别矢量的"显著"训练模式数大大低于总训练模式数,从而使得测试集的分类非常高效;同时,设计出专门的优化算法以加速"显著"训练模式的选取.实验表明,这种非线性方法不仅具有明显的效率上的优势,且具有不低于基于核的Fisher鉴别分析方法的性能.

关键词 基于核的Fisher鉴别分析;基于核的快速非线性鉴别分析;最小二乘解;特征抽取

中图法分类号 TP391.4

1 引 言

Fisher线性鉴别分析是比较有效的特征抽取与鉴别分析方法之一,在处理线性可分性较好的问题时,其性能与效率都是较优的,在实际应用中取得了较好结果[1~3].然而实际应用中有的模式识别问题的复杂度较高,以致线性鉴别分析手段难以取得较好的分类效果.研究表明,基于核函数的鉴别分析方法有着较优异的分类性能.其基本思想是将原特征空间通过某种形式的非线性映射变换到一个高维空间,并借助于"核技巧"在新的空间中应用鉴别分析方法.由于新空间中的线性方向对应于原特征空间的非线性方向,所以基于核的鉴别分析得出的鉴别方向也对应原特征空间的非线性方向.因此,基于核的鉴别分析是一种非线性鉴别分析方法.相对

收稿日期:2003-09-08;修回日期:2004-02-27
基金项目:国家自然科学基金项目(60072034)

于其他非线性方法,这种方法的独特和关键之处在于它巧妙地借助于“核函数”,而不需要对原特征空间进行任何直接的非线性映射,从而使计算鉴别矢量和进行鉴别分析的工作变得相对容易.“核技巧”最早应用于SVM方法中[4],后来又陆续发展出了基于核的主分量分析方法[5,6]与基于核的Fisher鉴别分析方法[7~9].基于核的Fisher鉴别分析方法是“核”方法中研究得较多的方法之一.

不过,在基于核的Fisher鉴别分析方法中,鉴别矢量的计算代价会随着训练样本数的增多而增大.另外,基于核的Fisher鉴别分析方法对每一个待测模式进行分类时均需计算它与所有训练模式间对应的核函数,所以测试效率较低.

本文分析了对原特征空间进行非线性映射得到的新空间中,线性判别方法对应的鉴别方向在最小二乘意义上与Fisher鉴别分析方法的等效性.新空间中的线性判别方法对应原特征空间的非线性方法.借助“核技巧”,这种非线性方法非常容易实现,而且模式分类的实现非常简便.不仅如此,本文提出的选择“显著”训练模式的方法可以使构造鉴别矢量的模式数大大低于总训练模式数,相应地,测试集的分类效率得到大大提升.实验表明,本文设计的基于核的快速非线性鉴别分析方法不仅有显著的效率优势,而且其性能不低于基于核的Fisher鉴别分析方法.

2 基于核的非线性鉴别分析

2.1 高维空间 F 中线性判别方法与Fisher鉴别分析的一致性

为了描述的简便,本文只分析两类问题。可用两个不同的数字来标示这两个类别,按照通常习惯,我们分别用1与-1作为两类的标示.我们在由 ϕ 映射得到的高维空间 F 中考虑分类问题.令训练模式数为 l,其中对应模式类1(也记做 c_1)的训练模式共 l_1 个:$\boldsymbol{\phi}(x_1),\boldsymbol{\phi}(x_2),\cdots,\boldsymbol{\phi}(x_{l_1})$,对应模式类-1(也记做 c_2)的训练模式共 l_2 个($l_1+l_2=l$):$\boldsymbol{\phi}(x_{l_1+1}),\boldsymbol{\phi}(x_{l_1+2}),\cdots,\boldsymbol{\phi}(l)$.下文将说明空间 F 中线性判别方法与Fisher鉴别分析是等效的.空间 F 中线性鉴别分析的任务就是寻找最好的鉴别方向,以实现对模式的最好分类(即空间 F 中的模式数据在鉴别方向的投影值应最好地接近其类别标示1或-1).因此,在空间 F 中根据训练模式设计线性鉴别方向可表述为如下问题:

$$\boldsymbol{\Phi W} = \boldsymbol{B}, \tag{1}$$

其中,

$$\boldsymbol{W} = \begin{bmatrix} w_0 \\ \boldsymbol{w} \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} 1 \\ \vdots \\ -1 \end{bmatrix}, \boldsymbol{\Phi} = \begin{bmatrix} 1 & \boldsymbol{\phi}(x_1)' \\ 1 & \boldsymbol{\phi}(x_2)' \\ \vdots & \vdots \\ 1 & \boldsymbol{\phi}(x_l)' \end{bmatrix}, \tag{2}$$

我们将 $\boldsymbol{W}$ 称为鉴别方向;将 $\boldsymbol{w}$ 称为鉴别矢量;将 w_0 称为阈值权.若能对方程组(1)进行求解,就能得出鉴别方向 $\boldsymbol{W}$.方程组(1)通常不存在惟一解,但容易求出其最小二乘解.式(1)的最小二乘解即为

$$\boldsymbol{\Phi}'\boldsymbol{\Phi W} = \boldsymbol{\Phi}'\boldsymbol{B} \tag{3}$$

的解.通过矩阵运算,方程组(3)可改写成

$$\begin{bmatrix} l & \left(\sum_{i=1}^{l_1}\boldsymbol{\phi}(x_i) + \sum_{i=l_1+1}^{l}\boldsymbol{\phi}(x_i)\right)' \\ \sum_{i=1}^{l_1}\boldsymbol{\phi}(x_i) + \sum_{i=l_1+1}^{l}\boldsymbol{\phi}(x_i) & \sum_{i=1}^{l}\boldsymbol{\phi}(x_i)\boldsymbol{\phi}(x_i)' \end{bmatrix}$$

$$\begin{bmatrix} w_0 \\ \boldsymbol{w} \end{bmatrix} = \begin{bmatrix} l_1 - l_2 \\ \sum_{i=1}^{l_1}\boldsymbol{\phi}(x_i) - \sum_{i=l_1+1}^{l}\boldsymbol{\phi}(x_i) \end{bmatrix}. \tag{4}$$

而式(4)中:

$$\sum_{i=1}^{l}\boldsymbol{\phi}(x_i)\boldsymbol{\phi}(x_i)' = \boldsymbol{S}_w^{\phi} + \frac{1}{l_1}\sum_{i=1}^{l_1}\boldsymbol{\phi}(x_i)\left(\sum_{i=1}^{l_1}\boldsymbol{\phi}(x_i)\right)' + \frac{1}{l_2}\sum_{i=l_1+1}^{l}\boldsymbol{\phi}(x_i)\left(\sum_{i=l_1+1}^{l}\boldsymbol{\phi}(x_i)\right)', \tag{5}$$

因此,由式(4)与式(5)可得出:

$$l\boldsymbol{S}_w^{\phi}\boldsymbol{w} + \left[\frac{l_1}{l_2}\sum_{i=l_1+1}^{l}\boldsymbol{\phi}(x_i) - \sum_{i=1}^{l_1}\boldsymbol{\phi}(x_i)\right]\left(\sum_{i=l_1+1}^{l}\boldsymbol{\phi}(x_i)\right)'\boldsymbol{w} + \left[\frac{l_2}{l_1}\sum_{i=1}^{l_1}\boldsymbol{\phi}(x_i) - \sum_{i=l_1+1}^{l}\boldsymbol{\phi}(x_i)\right]\left(\sum_{i=1}^{l_1}\boldsymbol{\phi}(x_i)\right)'\boldsymbol{w} = 2l_2\sum_{i=1}^{l_1}\boldsymbol{\phi}(x_i) - 2l_1\sum_{i=l_1+1}^{l}\boldsymbol{\phi}(x_i). \tag{6}$$

由于:

$$\sum_{i=1}^{l_1}\boldsymbol{\phi}(x_i) = l_1\boldsymbol{m}_1^{\phi}, \sum_{i=l_1+1}^{l}\boldsymbol{\phi}(x_i) = l_2\boldsymbol{m}_2^{\phi}, \tag{7}$$

$\boldsymbol{m}_1^{\phi}$ 与 $\boldsymbol{m}_2^{\phi}$ 分别为空间 $\boldsymbol{F}$ 中模式类1与-1的均值向量,且 $\left(\sum_{i=l_1+1}^{l}\boldsymbol{\phi}(x_i)\right)'\boldsymbol{w}$,$\left(\sum_{i=1}^{l_1}\boldsymbol{\phi}(x_i)\right)'\boldsymbol{w}$ 应各自对应一个数.因此,可得出:

$$\boldsymbol{w} = a(\boldsymbol{S}_w^{\phi})^{-1}(\boldsymbol{m}_1^{\phi} - \boldsymbol{m}_2^{\phi}), \tag{8}$$

$$w_0 = \frac{1}{l}[l_1 - l_2 - (l_1\boldsymbol{m}_1^{\phi} + l_2\boldsymbol{m}_2^{\phi})'\boldsymbol{w}, \tag{9}$$

其中,a 为一系数,S_w^{ϕ} 为模式的类内散布矩阵. 由此看出,由方程组(1)的最小二乘解得出的鉴别矢量与 $(S_w^{\phi})^{-1}(m_1^{\phi}-m_2^{\phi})$ 同方向. 而 $(S_w^{\phi})^{-1}(m_1^{\phi}-m_2^{\phi})$ 就是空间 F 中 Fisher 鉴别分析得出的最佳鉴别矢量[10]. 至此,可以看出,我们给出的空间 F 中的鉴别矢量就是最佳鉴别矢量. 由于空间 F 由非线性映射 ϕ 映射得到,因此,这个空间中的线性方法对应原特征空间中的非线性方法. 根据上述分析,若不考虑阈值权 w_0,可认为本文的非线性鉴别分析方法在最小二乘意义上与非线性 Fisher 鉴别分析是等效的. 事实上,除与 Fisher 鉴别分析等效外,还可以证明,当样本数趋于无穷时,本文非线性鉴别分析渐近逼近于贝叶斯判别函数[10].

2.2 鉴别方向的求解及待测模式的分类

依据再生核的理论,鉴别方向 W 可改写为

$$W=\begin{bmatrix} w_0 \\ \sum_{i=1}^{l}\alpha_i^l\phi(x_i)\end{bmatrix}. \tag{10}$$

为了与基于核的 Fisher 鉴别分析有所区别,本节中使用符号 α_i^l 而非 α_i. 方程组(1)可变形为

$$KA=B, \tag{11}$$

$$A=\begin{bmatrix} w_0 \\ \alpha_1^l \\ \vdots \\ \alpha_l^l\end{bmatrix}, K=\begin{bmatrix} 1 & k(x_1,x_1) & \cdots & k(x_1,x_l) \\ 1 & k(x_2,x_1) & \cdots & k(x_2,x_l) \\ \vdots & \vdots & & \vdots \\ 1 & k(x_l,x_1) & \cdots & k(x_l,x_l)\end{bmatrix}, \tag{12}$$

B 同式(2). 形同式(11)的方程组最小二乘解的一般形式为

$$A=(K'K)^{-1}K'B. \tag{13}$$

但是,由于 $K'K$ 是病态矩阵,所以可以引入一个正系数 μ,根据下式进行求解:

$$A=(K'K+\mu I)^{-1}K'B, \tag{14}$$

其中,I 为单位阵. 实际上,还可以证明,式(14)得出的 A 即为将 $\min(\mu A'A+(B-KA)'(B-KA))$ 作为目标函数得出的解,而 $\min(\mu A'A+(B-KA)'(B-KA))$ 即为考虑了正规化的最小平方错误代价函数,且 μ 为相应的正规化系数.

需要说明的是,文献[11]得出了与本节的分析相近的 A 的求解形式.

值得注意的是,求出 A 后并不需要设计专门的分类器就可实现分类. 具体方法如下,依据式(15)计算待测模式 x 在鉴别方向的投影值:

$$l_p(x)=w_0+\sum_{i=1}^{l}\alpha_i^l k(x,x_i). \tag{15}$$

若 $l_p(x)>0$,将 x 分类到模式类 1,否则,将 x 分类到模式类 -1. 因此,这种基于核的非线性鉴别分析方法的分类实现是非常简便的.

2.3 "显著"训练模式的选取

按照再生核的理论,鉴别矢量应为所有训练模式的线性组合,因此,对测试集的分类效率受到训练集大小的制约. 训练集越大则分类效率越低. 可喜的是,相关研究发现,实际鉴别分析中,只利用一部分训练模式来构造鉴别矢量以提高测试效率是可行的.

我们也基于这样的思路考虑问题. 假设只需使用训练模式中的一部分模式就可以较好地"描述"鉴别矢量,我们将这些模式称为"显著"训练模式. 此时,可将 w 表示如下:

$$w=\sum_{i=1}^{r}\alpha_i^0\phi(x_i^0), \tag{16}$$

其中,$x_1^0,x_2^0,\cdots,x_r^0$ 为"显著"训练模式. 由此,方程组(11)仍可表示为 $KA=B$,其中,B 同式(2),

$$A=\begin{bmatrix} w_0 \\ \alpha_1^0 \\ \vdots \\ \alpha_r^0\end{bmatrix}, K=\begin{bmatrix} 1 & k(x_1,x_1^0) & \cdots & k(x_1,x_r^0) \\ 1 & k(x_2,x_1^0) & \cdots & k(x_2,x_r^0) \\ \vdots & \vdots & & \vdots \\ 1 & k(x_l,x_1^0) & \cdots & k(x_l,x_r^0)\end{bmatrix}. \tag{17}$$

求出 A,即可计算待测模式 x 在鉴别方向的投影值:

$$l_p(x)=w_0+\sum_{i=1}^{r}\alpha_i^0 k(x,x_i^0). \tag{18}$$

但是,问题的关键是如何选择"显著"训练模式. 根据上述问题模型,我们提出下述选择合适的训练模式构造鉴别矢量的原则:使得构造鉴别矢量的"显著"训练模式数尽可能的少,且 A 对应的 $\mu\|A\|_2^2+\|KA-B\|_2^2$ 尽可能的小. $\mu\|A\|_2^2+\|KA-B\|_2^2$ 越小,说明方程组的求解误差越小,训练越成功,即训练模式在鉴别方向的投影值越接近其类别标示. 训练模式数越小,相应测试速度就越高. 因此,按照这样的原则选择"显著"训练模式是合理的. 从选择"显著"训练模式的方法来讲,可以从所有训练模式中逐一剔除不"显著"的模式,剔除过程完毕后,余下的模式即为所需的"显著"训练模式,我们将其称为"后退法". 也可以先假定所有模式均是"不显著"的,然后按重要程度将"显著"训练模式依次选出,我们将其称为"前进法". 训练集中模式数较多时,"后退法"将对应较大的计算量,因此,本文采用"前进

法”选择“显著”训练模式,其具体步骤如下:

(1) 算法第1步:选择第1个“显著”训练模式.

考察训练模式集中的每个 x_j, $j=1,2,\cdots,l$,其列矢量为 $[k(x_1,x_j) \quad k(x_2,x_j) \quad \cdots \quad k(x_l,x_j)]'$,对应该列矢量的矩阵为

$$K=\begin{bmatrix}1 & k(x_1,x_j)\\ 1 & k(x_2,x_j)\\ \vdots & \vdots\\ 1 & k(x_l,x_j)\end{bmatrix}. \tag{19}$$

根据式(14)求解 A,并将其记为 A_j. 计算出相应 $R_j=[\mu\|A_j\|_2^2+\|KA_j-B\|_2^2]^{\frac{1}{2}}$ 之值,将对应 R_j 值最小的 x_j 选做第1个“显著”训练模式,记为 x_1^0.

(2) 算法第 s 步:选择第 s 个“显著”训练模式 s.

在已选出 $s-1$ 个“显著”训练模式 $x_1^0,x_2^0,\cdots,x_{s-1}^0$ 后,相应矩阵 K 记为 K_{s-1}:

$$K_{s-1}=\begin{bmatrix}1 & k(x_1,x_1^0) & \cdots & k(x_1,x_{s-1}^0)\\ 1 & k(x_2,x_1^0) & \cdots & k(x_2,x_{s-1}^0)\\ \vdots & \vdots & & \vdots\\ 1 & k(x_l,x_1^0) & \cdots & k(x_l,x_{s-1}^0)\end{bmatrix}. \tag{20}$$

将 x_j, $j=1,2,\cdots,l(x_j\neq x_1^0,x_2^0,\cdots,x_{s-1}^0)$ 对应的列矢量 $k_j=[k(x_1,x_j) \quad k(x_2,x_j) \quad \cdots \quad k(x_l,x_j)]'$ 增加进矩阵 K_{s-1},作为其最后一列,将新得到的矩阵记为 K. 根据式(14)求解 A,并将其记为 A_j.

式(14)包含矩阵逆运算,如果直接根据矩阵求逆算法进行计算,计算量较大,将影响选择“显著”训练模式的效率. 可喜的是,由于 $K'K$ 为对称矩阵,可采用分块矩阵的思想设计出如下优化算法:

$$K'K+\mu I=\begin{bmatrix}K'_{s-1}\\ k'_j\end{bmatrix}[K_{s-1} \quad k_j]+\mu I=$$

$$\begin{bmatrix}K'_{s-1}K_{s-1}+\mu I_s & K'_{s-1}k_j\\ k'_jK_{s-1} & k'_jk_j+\mu\end{bmatrix}, \tag{21}$$

其中, I_s 为 s 阶单位矩阵. 令 $D=K'K+\mu I$, $D_{s-1}=K'_{s-1}K_{s-1}+\mu I_s$, $u=K'_{s-1}k_j$, $\gamma=k'_jk_j+\mu$,则式(21)可写为

$$D=\begin{bmatrix}D_{s-1} & u\\ u' & \gamma\end{bmatrix}. \tag{22}$$

由于行列式运算存在如下关系[12]:

$$\det(D)=\rho\cdot\det(D_{s-1}),$$
$$\rho=\gamma-u'(D_{s-1})^{-1}u. \tag{23}$$

可以证明:

$$(D)^{-1}=$$
$$\frac{1}{\rho}\begin{bmatrix}\rho(D_{s-1})^{-1}+(D_{s-1})^{-1}uu'(D_{s-1})^{-1} & -(D_{s-1})^{-1}u\\ -u'(D_{s-1})^{-1} & 1\end{bmatrix}. \tag{24}$$

令 $z=(D_{s-1})^{-1}u$,则根据式(14),有

$$A_j=\frac{1}{\rho}\begin{bmatrix}\rho(D_{s-1})^{-1}+zz' & -z\\ -z' & 1\end{bmatrix}\begin{bmatrix}K'_{s-1}B\\ k'_jB\end{bmatrix},$$
$$\rho=\gamma-u'z. \tag{25}$$

由于 $(D_{s-1})^{-1}$ 与 $K'_{s-1}B$ 在选出第 $s-1$ 个“显著”训练模式时已得出,使用式(25),可在选择“显著”训练模式的过程中以较少的计算量求出 A_j. 我们将该算法称为 A_j 的优化算法. 该算法的关键之处是在选择第 $s(s>1)$ 个“显著”训练模式的过程中使用已经计算出的第 $s-1$ 个“显著”训练模式的计算结果,在不直接进行矩阵逆运算的情况下巧妙地求出 $(K'K+\mu I)^{-1}$.

计算 $R_j=[\mu\|A_j\|_2^2+\|KA_j-B\|_2^2]^{\frac{1}{2}}$,将 R_j 值最小的 x_j 选做第 s 个“显著”训练模式,并将相应 R_j 记为 R_s. 重复该步骤(即步骤2)直至 $|R_s-R_{s-1}|<\varepsilon$ 成立(ε 为一个小量)才结束“显著”模式的选择. 最后一个被选做“显著”训练模式的模式对应的 A_j 即为 A 的最优解. 据此最优解,即可利用所有”显著”训练模式构造出形同式(16)的鉴别矢量 w 的表示式.

3 多类非线性鉴别分析

虽然上文的讨论都是针对两类模式展开的,但是,上文阐述的非线性鉴别分析方法可以方便地扩展到多类分类问题. 假设有 L 个类别,则可用 L 个向量来分别加以标示,联合使用 L 个非线性鉴别分析模型即可实现对待测模式的分类. 以 $L=6$ 为例来说明,第1类至第6类可分别用如下6个向量进行标示:

$$[+1 \; -1 \; -1 \; -1 \; -1 \; -1],$$
$$[-1 \; +1 \; -1 \; -1 \; -1 \; -1],$$
$$[-1 \; -1 \; +1 \; -1 \; -1 \; -1],$$
$$[-1 \; -1 \; -1 \; +1 \; -1 \; -1],$$
$$[-1 \; -1 \; -1 \; -1 \; +1 \; -1],$$
$$[-1 \; -1 \; -1 \; -1 \; -1 \; +1].$$

称这6个向量为各类的类别标示向量. 使用6个如同第3节中的非线性鉴别分析模型,可以容易

地实现6类问题的分类.6个不同的非线性鉴别分析模型设计如下:第1个非线性鉴别分析模型中将属于第1类的训练模式标示为1,而将其他所有训练模式标示为-1;第2个非线性鉴别分析模型中将属于第2类的训练模式标示为1,而将其他所有训练模式标示为-1;以此类推,…….在第6个非线性鉴别分析模型中,将属于第6类的训练模式标示为1,而将其他所有训练模式标示为-1.根据第3节的结论,所有这6个非线性鉴别分析模型都在最小二乘意义上与Fisher鉴别分析等效.同样根据第3节中的方法选择"显著"训练模式来构造各个非线性鉴别分析对应的鉴别矢量的表示式,这样的鉴别矢量共6个.在测试集的分类阶段,分别求出每个待测模式在6个鉴别方向的投影值,并将其组成一个包含6个元素的行向量;计算该行向量与上述各个类别标示向量间的距离,将该待测模式分类到与其距离最近的类别.不同类别数的分类问题均可依据上述办法分别加以解决.

实际上,这样一种分类方法中,每个非线性鉴别分析模型都采用"一对多"的分类策略,其中,第1个非线性鉴别分析模型将第1类与其他各类加以区分,第2个鉴别分析模型区分第2类与其他各类,……,第6个鉴别分析模型区分第6类与其他各类.由于每个训练模式都要在训练过程中被使用L次(L为类别数),一般的"一对多"策略会对应较大的计算量.本文虽然也使用"一对多"策略,但本文提出的多类非线性鉴别分析有如下两方面特点:首先,此处的每个非线性鉴别分析模型在最小二乘意义上都与鉴别性能较优的Fisher鉴别分析等效.其次,虽然多个非线性模型对应着一定的计算量增加,但由于选择"显著"训练模式方法的有效性,待测模式的分类仍然具有较高的效率,因此,在类别数不是很大的情况下,整个方法的实现仍然是高效的.

4　实　　验

在基于核的Fisher鉴别分析方法中,由于矩阵$\boldsymbol{N}$一般是奇异的,可对$\boldsymbol{N}$实施正则化后求解$\boldsymbol{\alpha}$:

$$\boldsymbol{\alpha} = (\boldsymbol{N} + \mu \boldsymbol{I})^{-1}(\boldsymbol{M}_1 - \boldsymbol{M}_2), \qquad (26)$$

其中,$l \times 1$阶矩阵$\boldsymbol{M}_i$的元素$(M_{ij}) = \frac{1}{l_i}\sum_{x \in c_i} k(x_j, \boldsymbol{x})$,$j = 1, 2, \cdots, l$,$\boldsymbol{N} = \sum_{i=1}^{2} \boldsymbol{K}_i(\boldsymbol{I} - \boldsymbol{I}_{l_i})\boldsymbol{K}_i'$,且$\boldsymbol{I}$为单位矩阵,$\boldsymbol{I}_{l_i}$为$l_i \times l_i$阶矩阵且所有元素都为$1/l_i$,$\boldsymbol{K}_i$为$l \times l_i$阶矩阵,且$\boldsymbol{K}_1$为$c_1$类的核矩阵$(\boldsymbol{K}_1)_{i,j} = k(x_i, x_j)$,$x_i = x_1, x_2, \cdots, x_l$;$x_j = x_1, x_2, \cdots, x_{l_1}$,$\boldsymbol{K}_2$为$c_2$类的核矩阵,$(\boldsymbol{K}_2)_{i,j} = k(x_i, x_j)$,$x_i = x_1, x_2, \cdots, x_l$;$x_j = x_{l_1+1}, x_{l_1+2}, \cdots, x_l$.本文实验中将$\mu$取为1.0E-3.我们使用最小距离分类器进行识别,其具体实现如下:

分别计算出训练模式中所有模式在最佳鉴别矢量上的投影值$p(x_i)$:

$$p(x_i) = \sum_{j=1}^{l} \alpha_j k(x_i, x_j), i = 1, 2, \cdots, l, \qquad (27)$$

并由此估计新特征空间中各类的类中心.计算出各待测模式在最佳鉴别矢量上的投影值后,根据类中心与该投影值距离最近的那个类别即是该测试模式所属类别的原则进行分类.

在非线性鉴别分析中,式(14)中的系数也取值为1.0E-3.

4.1　"基准"模式集实验结果

实验在10个"基准"模式集(http://ida.first.gmd.de/~raetsch/data/)上进行.每个模式集被随机地分成了100个部分(Image与Splice例外,这两个模式集只包含20个部分).这10个"基准"模式集中既有人工模式集也有实测模式集.该实验采用高斯型核函数

$$k(\boldsymbol{x}, \boldsymbol{y}) = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{y}\|^2}{2\sigma^2}\right) \qquad (28)$$

进行计算.每次实验中将σ^2取为第1个训练模式集的方差.在第1个训练模式集上进行训练,然后对100个测试集进行分类."显著"模式的选择中将ε取为0.04.

表1与表2分别给出了基于核的Fisher鉴别分析方法与本文非线性鉴别分析方法在10个"基准"模式集上的实验结果.表1与表2中第2列为每个模式集上的百分比平均分类错误率与百分比分类错误率标准差,表中第3列为对测试集进行分类所需的CPU时间,表中第4列为整个方法实现所需的CPU时间,时间单位为秒.表中第5列为"显著"训练模式数(对基于核的Fisher鉴别分析而言,即为训练集的总模式数)以及"显著"训练模式数与总训练模式数之比.从表中看出,本文非线性鉴别分析方法筛选出的各个训练集中的"显著"训练模式数只占总训练模式数的一小部分."显著"训练模式数与总训练模式数之比最大值为11%,最小值仅为1%.相应地,本文非线性鉴别分析方法对测试集的分类效率大大高于基于核的Fisher鉴别分析方法.不仅

如此,本文非线性鉴别分析整个方法实现所需的总时间(即训练和测试所需的总时间)也小于基于核的 Fisher 鉴别分析方法. 而本文非线性鉴别分析方法也取得了较低的分类错误率,在 10 个模式集的分类中,非线性鉴别分析在 6 个模式集上的分类结果优于使用最小距离分类器进行分类的基于核的 Fisher 鉴别分析方法. 实验说明了非线性鉴别分析的高效与高性能.

Table 1 Classification Error Rates and Consumed Time in Kernel Fisher Discriminant Analysis

表 1 基于核的 Fisher 鉴别分析分类错误率与时间

Dataset	Error Rate, Deviation	Classification Time(s)	Total Time(s)	Number of Total Training Patterns
Banana	13.7±0.1	3515	3526	400
B.Cancer	22.7±4.4	34	37	200
Diabetis	22.1±1.9	281	298	468
German	21.3±2.1	457	501	700
Heart	11.5±2.8	35	37	170
Image	9.0±0.5	542	766	1300
F.Solar	32.2±1.6	47	94	666
Splice	11.0±0.5	1008	1023	1000
Thyroid	1.8±1.1	23	25	140
Titanic	25.5±0.3	567	569	150

Table 2 Classification Error Rates and Consumed Time in Nonlinear Discriminant Analysis

表 2 非线性鉴别分析分类错误率与时间

Dataset	Error Rate, Deviation	Classification Time(s)	Total Time(s)	Number of Significant Training Patterns (The Ratio Between "Significant Nodes" and Total Training Patterns)
Banana	13.5±0.1	125	161	14(4%)
B.Cancer	19.5±3.9	4	26	22(11%)
Diabetis	21.3±1.5	9	48	12(3%)
German	22.6±2.1	12	70	8(1%)
Heart	13.9±3.0	4	10	14(8%)
Image	5.3±0.5	14	727	54(4%)
F.Solar	32.1±1.8	7	41	6(1%)
Splice	12.9±0.4	43	660	41(4%)
Thyroid	3.3±1.7	7	12	15(11%)
Titanic	22.7±0.3	18	110	3(2%)

4.2 CMU-Pittsburgh 动作单元编码人脸表情数据库实验结果

对 CMU-Pittsburgh 动作单元编码人脸表情数据库[15]采用适当的局部 Gabor 滤波处理[16],可得到 463 幅人脸表情"模板"图像. 每幅图像均为灰度图像,分辨率为 60×70. 这些图像分别包含惊讶、愤怒、悲伤、喜悦、恐惧和厌恶共 6 种表情. 我们将前 210 幅图像作为训练模式,后 253 幅图像作为测试模式.

这是一个多类别识别问题,6 种表情对应 6 个类别,可分别用第 4 节的 6 个行向量来加以标示. 该实验中采用多项式核函数

$$k(x,y)=(x\cdot y)^d. \tag{29}$$

计算中将 d 取为 2."显著"模式的选择中将 ε 取为 0.1. 表 3 给出了联合使用 6 个非线性鉴别分析模型得出的分类结果以及每个非线性鉴别分析模型中的"显著"训练模式数. 表 3 中用 $C1,C2,C3,C4,C5,C6$ 分别表示第 1~6 个鉴别分析模型,其下方对应数据为各鉴别分析模型选择出的"显著"训练模式数和"显著"训练模式数占总训练模式数的百分比(百分比为括号中给出的数据). 我们看到,6 个鉴别分析模型中,"显著"训练模式数均占总训练模式数的一小部分. 联合使用这 6 个模型得到的测试集的分类错误率为 11%.

Table 3　Experimental Result on CMU-Pittsburgh Face Database
表 3　CMU-Pittsburgh 动作单元编码人脸表情数据库实验结果

Discriminant Analysis Model	Number of Significant of Nodes (The Ratio Betteen"Significant Nodes" and Total Training Patterns)
$C1$	8(4%)
$C2$	13(6%)
$C3$	15(7%)
$C4$	9(4%)
$C5$	10(5%)
$C6$	15(7%)
Classification Error Rate	11%

5　结　　论

基于核的 Fisher 鉴别分析借助于"核技巧",巧妙地实现了从原特征空间到高维特征空间的非线性映射,以求在高维空间中增大模式可分性. 这种基于核的方法在研究中取得了较好的效果,但它也对应较大的计算代价,当训练模式数较大时尤其如此. 因此,设计计算量相对较小的基于核的方法是非常重要的. 本文分析了基于核的非线性鉴别分析在最小二乘意义上与核 Fisher 鉴别分析的等效性,并给出其求解方法,这种方法求解鉴别矢量所需计算量较小;而且可以简便地实现模式的分类. 本文设计出选择"显著"训练模式的算法,以求由较少的训练模式构造鉴别矢量,从而提高测试集的分类效率,并给出了该算法的优化方案. 进一步,应用"一对多"的分类策略,本文将该非线性鉴别分析方法扩展到多类问题的分类. 本文非线性鉴别分析方法与基于核的 Fisher 鉴别分析在 10 个"基准"模式集上的对比实验表明,前者有不低于基于核的 Fisher 鉴别分析的性能,且其选择出的"显著"训练模式只为总训练模式中的一小部分模式,计算效率大大高于基于核的 Fisher 鉴别分析方法. 而本文方法在 CMU-Pittsburgh 动作单元编码人脸表情数据库上的实验表明,由于选择出的"显著"训练模式数较小,本文方法解决类别数不太大的多类问题时仍然具有较高的效率.

参 考 文 献

1　Y. Xu, J. Yang, Z. Jin. Theory analysis on FSLDA and ULDA. Pattern Recognition, 2003, 36(12): 3031～3033

2　Y. Xu, J. Yang, Z. Jin. A novel method for Fisher discriminant analysis. Pattern Recognition, 2004, 37(2): 381～384

3　Z. Jin, J. Yang, Z. Hu, *et al*.. Face recognition based on the uncorrelated discriminant transformation. Pattern Recognition, 2001, 34(7): 1405～1416

4　V. N. Vapnik. The Nature of Statistical Learning Theory. Beijing: Tsinghua University Press, 2002 (in Chinese)
(V. N. Vapnik. 统计学习理论的本质. 北京: 清华大学出版社, 2002)

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

6　K. R. Muller, S. Mika, G. Rätsch, *et al*.. An introduction to kernel-based learning algorithms. IEEE Trans. on Neural Network, 2001, 12(2): 181～201

7　S. Mika, G. Rätsch, J. Weston, *et al*.. Fisher discriminant analysis with kernels. In: Y. H. Hu, J. Larsen, E. Wilson, eds.. Neural Networks for Signal Processing IX. Piscataway, NJ: IEEE Press, 1999. 41～48

8　S. Mika, A. J. Smola, B. Schölkopf. An improved training algorithm for kernel Fisher discriminants. In: T. Jaakkola, T. Richardson eds.. Proc. of AISTATS. San Mateo, CA: Morgan Kaufmann, 2001. 98～104

9　G. C. Cawley, N. L. C. Talbot. Efficient leave-one-out cross-validation of kernel Fisher discriminant classifiers. Pattern Recognition, 2003, 36(11): 2585～2592

10　Bian Zhaoqi, Zhang Xuego. Pattern Recognition. Beijing: Tsinghua University Press, 2000 (in Chinese)
(边肇祺, 张学工. 模式识别. 北京: 清华大学出版社, 2000)

11　J. Xu, X. Zhang, Y. Li. Kernel MSE algorithm: A unified framework for KFD, LS-SVM and KRR. The Int'l Joint Conf. on Neural Networks(IJCNN-2001), Washington, D. C., 2001

12　Zhang Mingchun. Matrix Theory. Nanjing: Southeast University Press, 1995 (in Chinese)
(张明淳. 工程矩阵理论. 南京: 东南大学出版社, 1995)

13　D. H. Wolpert. Stacked generalization. Neural Networks, 1992, 5(2): 241～260

14　T. Hastie, R. Tibshirani. Classification by pairewise coupling. In: M. I. Jordan, M. J. Kearns, S. A. Solla eds(in Chinese). Advances in Neural Information Processing Systems, Vol. 10. MA: MIT Press, 1998

15　T. Kanade, J. F. Cohn, Y. Tian. Comprehensive database for facial expression analysis. The 4th Int'l Conf. of Face and Gesture Recognition, Grenoble, France, 2000

16　S. Dubuisson, F. Davoine, M. Masson. A solution for facial expression representation and recognition. Signal Processing: Image Communication, 2002, 17(9): 657～673

Xu Yong, born in 1972. He received his Ph. D.'s degree in the Pattern Recognition and Intelligence System at Nanjing University of Science & Technology. His current research interests include face recognition and authentication, character recognition, image

retrieval and kernel method. He is the author of over 10 scientific papers in these areas.

徐勇,1972年生,博士,讲师,主要研究方向为模式识别与图像处理,目前研究兴趣为人脸识别与认证、字符识别、图像检索与核方法,已在该领域发表学术论文十余篇.

Yang Jingyu, born in 1941. He works in the Department of Computer Science NUST as professor. His research areas include computer vision, information fusion, pattern recognition, and intelligent robot. He is the author of over 200 scientific papers in these areas.

杨静宇,1941年生,教授,博士生导师,主要研究方向为计算机视觉、信息融合、模式识别、智能机器人,已在该领域发表学术论文两百多篇.

Jin Zhong, born in 1961. Currently a professor in the Department of Computer Science at NUST. His main research interests are face recognition, character recognition, image retrieval and image processing. He is the author of over 20 scientific papers in these areas.

金忠,1961年生,教授,主要研究方向为人脸与字符识别、图像检索与图像处理,已在该领域发表学术论文二十余篇.

Lou Zhen, born in 1974. Currently a vice-professor in the Department of Computer Science at NUST. His main research interests are character recognition and image processing. He is the author of over 10 scientific papers in these areas.

娄震,1974年生,博士后、副教授,主要研究方向为字符识别、图像处理,已在该领域发表学术论文十余篇.

Research Background

Kernel Fisher discriminant analysis has attracted many attention and has been applied in many recognition problems because it owns conceptual elegance and the state-of-the-art performance. The key of kernel Fisher method is to classify test patterns in some high dimensional space using kernel trick. However, it is well known that kernel Fisher discriminant analysis is based on the theory of reproducing kernels, as a result the classification efficiency of the naive kernel Fisher discriminant analysis is in inverse proportion to the number of training patterns. Consequently if the number of training patterns is large enough, the method may become impractical, especially for some real time classification application with strong require on efficiency. So it is very important to improve the classification efficiency of the naive kernel Fisher discriminant analysis if suitable approach is available.

In this paper, a nonlinear discriminant analysis scheme, based on one nonlinear mapping, is discussed. In space feature, the least squares solution of the discriminant analysis method is equivalent to Fisher discriminant analysis. Using kernel trick, it can derive a kernel classification method. As a result, it can be expected that the kernel method is comparative to kernel Fisher discriminant analysis. The discriminant vector of the novel method may be efficiently solved from linear equations. Obviously, its efficiency for solving discriminant vector is great higher than kernel Fisher discriminant analysis. Moreover, corresponding classifying strategy is very simple. In order to achieve more efficient classification, we develop a novel scheme. The most striking advantage of the novel scheme is that only a few original training samples are taken as "significant" nodes for constructing discriminant vector in feature space. As a result, corresponding testing is much more efficient than the naïve nonlinear discriminant analysis. In addition an appropriative, optimized algorithm is developed to assure that selecting "significant" nodes is not too time consuming. On the other hand, the algorithm is dependable and feasible with reasonable computing cost and is a near optimal. Experiments on benchmarks and face databases show that the performance of the novel scheme is comparative to kernel-based Fisher discriminant analysis, with superiority in efficiency.

The above idea enlightens us that maybe other approach, for improving the above kernel method in efficiency, is available. Actually, we do intend to explore the possibilities for it and expect better result on this subject.

This paper is supported by the National Natural Science Foundation of China under grant No. 60072034.