

# 一种基于 QR 分解的增量式核判别分析法

王万良 陈宇 邱虹 郑建炜

(浙江工业大学计算机学院 杭州 310023)

**摘要** 针对非线性系统在线学习的效率问题,提出了一种基于 QR 分解的增量式核判别分析法。该算法充分利用基于 QR 分解的核判别分析法的先降维后提取特征的思想,将核空间映射到低维空间进行计算,减少了构造核矩阵的计算量,降低了核矩阵的存储空间。同时引入增量计算的思想,有效地解决了在线学习中冗余计算的问题。在 TE 过程数据和 ORL 人脸库上的仿真实验证明了该算法在特征提取上的有效性,其相比批量式算法有更高的效率优势。

**关键词** 非线性系统,在线学习,QR 分解,增量计算,特征降维

中图分类号 TP277 文献标识码 A

## Incremental Kernel Discriminant Analysis Method via QR Decomposition

WANG Wan-liang CHEN Yu QIU Hong ZHENG Jian-wei

(College of Computer Science, Zhejiang University of Technology, Hangzhou 310023, China)

**Abstract** To improve the online learning efficiency of nonlinear system, an incremental kernel decomposition analysis method via QR decomposition was developed. This algorithm maps the kernel space to a lower dimension space before performing feature decomposition to reduce the amount of calculation and the storage space of the kernel matrix. Then this algorithm solves the problem of redundant computation by combining the idea of incremental computation. The experiments on TE process and ORL data set show this algorithm is effective on feature decomposition as well as more efficient than the batch one.

**Keywords** Nonlinear system, Online learning, QR decomposition, Incremental computation, Dimension reduction

## 1 引言

日益复杂的工业系统和人脸识别系统中普遍存在着非线性问题。近年来,非线性问题逐渐成为研究热点。核方法的出现大大推动了非线性过程研究的发展<sup>[1-3]</sup>。核 Fisher 判别分析法(Kernel Fisher Discriminant Analysis, KFDA)是将核学习的思想引入到经典降维算法 FDA 中所产生的一种非线性特征提取方法,其基本思想是通过一个非线性映射,将原空间中的数据映射到一个高维的核空间中,然后在这个高维核空间中进行线性 Fisher 判别分析<sup>[4,5]</sup>。KFDA 方法虽然在非线性特征提取方面有着不错的效果,但是存在两个主要问题,第一是构造核矩阵  $K$  的计算量非常大,第二是当训练样本数量很大时,核矩阵  $K$  的存储空间会变得非常大<sup>[6]</sup>。

复杂的非线性过程往往还面临在线学习问题,在实际应用中,采集一组完整且有代表性的数据经常是昂贵和费时的<sup>[7]</sup>。在很多应用的初始阶段只有少量可供使用的样本数据,而随着时间的推移,新增样本会陆续呈现,需根据新的训练数据及时更新最佳投影矩阵。如果采用传统批量式的非线性特征提取方法,每次新样本到来都需要重新构造核矩阵  $K$ ,并且需要将新样本与原有样本一起重新进行一次完整的投影

矩阵计算,这严重影响了特征提取的效率。因此本文提出了一种专门针对非线性系统在线学习效率问题的增量式非线性特征提取方法。这种方法有效地提高了更新投影矩阵的效率,同时减少了核矩阵  $K$  的存储空间,更适用于非线性过程在线监控。

基于 QR 分解的核判别分析(Kernel Discriminant Analysis via QR Decomposition, KDR/QR)是由 Xiong 等人提出的<sup>[8]</sup>。它是一种利用核思想对非线性数据进行特征降维的方法。该方法的主要思想是,首先根据核方法将样本原始空间中的类间散度矩阵  $S_b$  和总体散度矩阵  $S_t$  映射到高维核空间中,然后在高维核空间中以类间散度矩阵  $S_b^*$  最大为目标再将  $S_b^*$  和  $S_t^*$  映射到一个低维空间中,并在低维空间中应用 Fisher 判别分析求取低维空间的最佳投影矩阵,再利用 QR 分解求得原始空间的最佳投影矩阵。

KDR/QR 的优势是将核空间  $H$  映射到一个低维空间  $L$  后再进行判别分析,减小了特征分解的规模从而提升了特征提取效率。本文提出的基于 QR 分解的增量式核判别分析法(Incremental Kernel Discriminant Analysis via QR Decomposition, IKDR/QR)利用了这种先降维后特征提取的思想,在增量过程中充分利用低维空间  $L$  降低构造核矩阵  $K$  的规模,

到稿日期:2013-06-21 返修日期:2013-11-12 本文受国家科技支撑计划课题(2012BAD10B01),浙江省自然科学基金(LQ12F03011)资助。

王万良(1957-),男,教授,博士生导师,主要研究方向为计算机控制与智能自动化等;陈宇(1989-),男,硕士生,主要研究方向为模式识别、计算机视觉等;邱虹(1989-),女,硕士生,主要研究方向为模式识别、计算机视觉等;郑建炜(1982-),男,博士,讲师,主要研究方向为模式识别、计算机视觉等,E-mail:zjw@zjut.edu.cn.

减少核矩阵  $K$  的存储空间。

本文第 2 节概述了小波去噪的原理以及基于 QR 分解的核判别分析法 KDR/QR;第 3 节提出了一种基于 QR 分解的增量式核判别分析法 IKDR/QR,并分析了增量式方法与批量式方法的算法复杂度,从理论上证明了 IKDR/QR 的性能优势;第 4 节通过在 TE 过程数据和 ORL 人脸库上的仿真实验从识别率和计算性能两个角度验证了算法的有效性;最后对本文工作进行总结和展望。

## 2 相关工作

基于 QR 分解的核判别分析法 KDR/QR 首先以核空间  $H$  中类间散度矩阵  $S_b^\circ$  最大为目标,将  $S_b^\circ$  和  $S_t^\circ$  映射到一个低维空间<sup>[8]</sup>。设映射  $\Phi$  将数据  $a$  从原始空间映射到更高维的核空间  $H$  中,记高维空间中的  $a$  为  $\Phi(a)$ ,于是有核空间  $H$  中的质心矩阵  $C^\circ$  为:

$$C^\circ = [m_1^\circ, \dots, m_c^\circ] = \left[ \frac{1}{n_1} \sum_{i \in c_1} \Phi(a_i), \dots, \frac{1}{n_c} \sum_{i \in c_c} \Phi(a_i) \right] \quad (1)$$

于是有,

$$\begin{aligned} S_b^\circ &= C^\circ E E^T (C^\circ)^T = Q^\circ (R^\circ E E^T (R^\circ)^T) (Q^\circ)^T \\ &= Q^\circ \tilde{E}^\circ (Q^\circ)^T \end{aligned} \quad (2)$$

其中,  $Q^\circ$  为对质心矩阵  $C^\circ$  进行 QR 分解后的  $Q$  矩阵。  $\tilde{E}^\circ = R^\circ E E^T (R^\circ)^T$ ,  $E$  是一个  $c \times c$  维的矩阵,其第  $i$  列为:

$$\sqrt{n_i} (0, \dots, 0, 1, 0, \dots, 0)^T + \frac{\sqrt{n_i}}{n} (n_1, n_2, \dots, n_c)^T \quad (3)$$

其中,  $c$  为样本类别数,  $n_i$  为第  $i$  类样本的样本个数,  $n$  为样本总数。对于任意正交矩阵  $Y$  有:

$$\begin{aligned} \text{trace}(Y^T S_b^\circ Y) &\leq \text{trace}(S_b^\circ) = \text{trace}(Q \tilde{E}^\circ Q^T) \\ &= \text{trace}(\tilde{E}^\circ (Q^\circ)^T Q^\circ) = \text{trace}(\tilde{E}^\circ) \end{aligned} \quad (4)$$

由此可见  $\text{trace}(\tilde{E}^\circ)$  是  $\text{trace}(Y^T S_b^\circ Y)$  的上界,根据迹的特性以及矩阵  $Q$  中列的两两正交性有:

$$\begin{aligned} \text{trace}((Q^\circ)^T S_b^\circ Q^\circ) &= \text{trace}((Q^\circ)^T Q^\circ \tilde{E}^\circ (Q^\circ)^T Q^\circ) \\ &= \text{trace}(\tilde{E}^\circ) \end{aligned} \quad (5)$$

由式(5)可以看出,当到矩阵  $Y$  取  $Q^\circ$  时,  $\text{trace}(Y^T S_b^\circ Y)$  取得上界。因此可以以  $Q^\circ$  作为投影矩阵将  $S_b^\circ$  和  $S_t^\circ$  映射到一个以类间散度矩阵  $S_b^\circ$  最大为目标的低维空间,则有低维空间类间散度矩阵  $B^\circ = (Q^\circ)^T S_b^\circ Q^\circ$  和低维空间总体散度矩阵  $T^\circ = (Q^\circ)^T S_t^\circ Q^\circ$ 。

设矩阵  $A^\circ = [\Phi(a_1), \Phi(a_2), \dots, \Phi(a_n)]$ , 则有  $C^\circ = A^\circ M$ , 其中  $M$  是一个  $n \times c$  矩阵,其第  $i$  列为  $(0, \dots, 0, \frac{1}{n_i}, \dots, \frac{1}{n_i}, 0, \dots, 0)^T$ 。设  $K$  为核矩阵,其中  $K(i, j) = \langle \Phi(a_i), \Phi(a_j) \rangle$ , 则有

$$(C^\circ)^T C^\circ = M^T K M \quad (6)$$

对于核空间  $H$  中的类间散度矩阵  $S_b^\circ$  有  $S_b^\circ = (H_b^\circ)^T H_b^\circ = (C^\circ N)^T (C^\circ N)$ , 其中  $N$  的第  $i$  列为  $((0, \dots, \sqrt{n_i}, \dots, 0)^T - \frac{\sqrt{n_i}}{n} (n_1, \dots, n_c)^T)$ , 因此有,

$$\begin{aligned} B^\circ &= (Q^\circ)^T S_b^\circ Q^\circ = (Q^\circ)^T H_b^\circ (H_b^\circ)^T Q^\circ \\ &= N^T (C^\circ)^T C^\circ (R^\circ)^{-1} = N^T M^T K M (R^\circ)^{-1} \end{aligned} \quad (7)$$

其中,  $R^\circ$  为对  $C^\circ$  进行 QR 分解的  $R$  矩阵,有  $(C^\circ)^T C^\circ =$

$(R^\circ)^T R^\circ$ ,  $R^\circ$  可以通过对  $(C^\circ)^T C^\circ$  进行 Cholesky 分解得到<sup>[9]</sup>。同样地,对于总体散度矩阵有  $S_t^\circ = (H_t^\circ)^T H_t^\circ = (A^\circ E)^T A^\circ E$ , 其中  $E = I - \frac{1}{n} e e^T$ , 则有:

$$\begin{aligned} T^\circ &= (Q^\circ)^T S_t^\circ Q^\circ \\ &= (Q^\circ)^T H_t^\circ (H_t^\circ)^T Q^\circ \\ &= E^T (A^\circ)^T C^\circ (R^\circ)^{-1} \\ &= E^T (A^\circ)^T A^\circ M (R^\circ)^{-1} \\ &= E^T K M (R^\circ)^{-1} \end{aligned} \quad (8)$$

对  $(T^\circ + \mu I)^{-1} B^\circ$  特征分解得到  $c$  个特征向量  $\{\phi_i^\circ\}_{i=1}^c$ 。

设  $V^\circ = [\phi_1^\circ, \phi_2^\circ, \dots, \phi_c^\circ]$ , 最终的投影矩阵  $G$  可以按如下公式计算:

$$G^\circ = Q^\circ V^\circ = C^\circ (R^\circ)^{-1} V^\circ \quad (9)$$

对于一个给定的样本  $z$ , 其进行  $G^\circ$  投影降维后有:

$$\begin{aligned} (G^\circ)^T \Phi(z) &= (V^\circ)^T ((R^\circ)^{-1})^T (C^\circ)^T \Phi(z) \\ &= (V^\circ)^T ((R^\circ)^{-1})^T M^T K_\pi \end{aligned} \quad (10)$$

其中,  $K_\pi \in \mathbb{R}^n$ ,  $K_\pi(i) = \langle \Phi(a_i), \Phi(z) \rangle$ 。

## 3 基于 QR 分解的增量式核判别分析法

对于在线学习问题,传统降维方法局限于每次训练都需要对所有样本数据进行一次完整的降维操作,该过程包含很多冗余的计算,影响求取最佳投影矩阵的效率。增量式计算的思想是直接根据新的样本,对前一次的降维结果进行更新<sup>[10]</sup>。依据式(7)和式(8)可以看到,矩阵  $B^\circ$ 、 $T^\circ$  都是  $c \times c$  的矩阵,而  $S_b^\circ$  和  $S_t^\circ$  都是  $d \times d$  的矩阵,  $d$  是样本在核空间中的维度,一般  $d$  远大于  $c$ 。因此在增量式训练过程中反复地更新  $S_b^\circ$  和  $S_t^\circ$  是费时的,而本文提出的增量式方法将直接选择在低维空间  $L$  中对矩阵  $B^\circ$ 、 $T^\circ$  进行更新。

IKDR/QR 算法过程就是在新数据样本插入原有样本空间后采用更高效的方法更新  $R^\circ$ 、核矩阵  $K$  以及低维空间中的类间散度矩阵  $B^\circ$  和总体散度矩阵  $T^\circ$ 。当新的样本数据插入后,对于任意变量  $X$ ,更新后的  $X$  记为  $\tilde{X}$ 。

### 3.1 更新类间散度矩阵 $B^\circ$

根据式(7)有  $\tilde{B}^\circ = \tilde{N}^T \tilde{M}^T \tilde{K} \tilde{M} (\tilde{R}^\circ)^{-1}$ , 我们选择直接更新其中  $\tilde{M}^T \tilde{K} \tilde{M}$  部分,而不是重新构造核矩阵  $K$ 。记  $K(i, j) = \langle \Phi(a_i), \Phi(a_j) \rangle$  为  $\Phi(a_{ij})$ , 对于矩阵  $M^T K M$  有,

$$M^T K M = \begin{pmatrix} \frac{1}{n_1^2} \sum_{a_{i,j} \in c_1} \Phi(a_{ij}) & \dots & \frac{1}{n_1 n_c} \sum_{a_i \in c_1, a_j \in c_c} \Phi(a_{ij}) \\ \vdots & \ddots & \vdots \\ \frac{1}{n_c n_1} \sum_{a_{i,j} \in c_1} \Phi(a_{ij}) & \dots & \frac{1}{n_c^2} \sum_{a_{i,j} \in c_c} \Phi(a_{ij}) \end{pmatrix} \quad (11)$$

可以看到  $M^T K M$  是一个  $c \times c$  矩阵,设  $A_{ij}$  为矩阵  $M^T K M$  中的第  $i$  行第  $j$  列的对应项,则  $A_{ij}$  为第  $i$  类样本与第  $j$  类样本核关系的和。设新插入的样本数据  $x$  属于第  $t$  类,  $t \leq c$ 。随着新样本数据的到来,原始样本空间结构产生了两种改变:1)原始空间中的样本与新的样本形成新的核空间关系  $K_\pi = \langle \Phi(a_i), \Phi(x) \rangle$ ;2)影响了第  $t$  类样本的类内空间结构。因此求取更新后的  $\tilde{M}^T \tilde{K} \tilde{M}$  包括两个步骤:

1)更新  $M^T K M$  中非  $t$  类样本与  $t$  类样本的核关系

$$\begin{aligned}\tilde{A}_p &= \frac{1}{n_p \bar{n}_t} \sum_{a_i \in c_p, a_j \in c_t} \Phi(a_{ij}) \\ &= \frac{1}{n_p(n_t+1)} \left( \sum_{a_i \in c_p, a_j \in c_t} \Phi(a_{ij}) + \sum_{a_i \in c_p} \Phi(a_{ix}) \right) \\ &= \frac{n_t}{(n_t+1)} A_p + \frac{1}{n_p(n_t+1)} \sum_{a_i \in c_p} \Phi(a_{ix})\end{aligned}\quad (12)$$

其中,  $p \leq c, p \neq t$ 。

2) 更新  $M^T KM$  中  $t$  类样本的类内关系

$$\begin{aligned}\tilde{A}_t &= \frac{1}{n_t^2} \sum_{a_i, j \in c_t} \Phi(a_{ij}) \\ &= \frac{1}{(n_t+1)^2} \left( \sum_{a_i, j \in c_t} \Phi(a_{ij}) + 2 \sum_{a_i \in c_t} \Phi(a_{ix}) + \Phi(a_{xx}) \right) \\ &= \frac{n_t^2}{(n_t+1)^2} A_t + \frac{2}{(n_t+1)^2} \sum_{a_i \in c_t} \Phi(a_{ix}) + \frac{1}{(n_t+1)^2} \Phi(a_{xx})\end{aligned}\quad (13)$$

由式(12)、式(13)可以看到  $A_p$  和  $A_t$  是前一次增量计算的结果, 所以更新  $M^T KM$  需要重新计算的部分只有新插入的样本  $x$  与原始样本之间的核空间关系, 这就避免了大量的重复计算。同时根据式(7), 对  $\tilde{M}^T \tilde{K} \tilde{M}$  运用 Cholesky 分解即可求得更新后的  $\tilde{R}^\circ$ 。

### 3.2 更新总体散度矩阵 $T^\circ$

根据式(8)有  $\tilde{T}^\circ = \tilde{E}^T \tilde{K} \tilde{M} (\tilde{R}^\circ)^{-1}$ , 其中  $\tilde{E} = I - \frac{1}{n} ee^T$ , 同

样我们也选择直接更新  $\tilde{K} \tilde{M}$ , 对于  $KM$  有:

$$KM = \begin{pmatrix} \frac{1}{n_1} \sum_{a_i \in c_1} \Phi(a_{i1}) & \cdots & \frac{1}{n_c} \sum_{a_i \in c_c} \Phi(a_{i1}) \\ \vdots & \ddots & \vdots \\ \frac{1}{n_1} \sum_{a_i \in c_1} \Phi(a_{in}) & \cdots & \frac{1}{n_c} \sum_{a_i \in c_c} \Phi(a_{in}) \end{pmatrix}\quad (14)$$

$KM$  是一个  $n \times c$  矩阵, 其中  $A_{ij}$  为第  $i$  个样本与第  $j$  类样本核关系的和。因此随着新的样本  $x$  的插入, 矩阵  $KM$  也产生了两种改变: 1) 每个样本与  $t$  类样本的核空间关系; 2) 插入代表新样本  $x$  与每类样本核空间关系的向量。

1) 更新每个样本与  $t$  类样本的核关系

根据公式, 需要更新  $KM$  中的  $A_p, p \leq c, p \neq t$ 。

$$\begin{aligned}\tilde{A}_p &= \frac{1}{\bar{n}_t} \sum_{a_j \in c_t} \Phi(a_{pj}) \\ &= \frac{1}{n_t+1} \left( \sum_{a_j \in c_t} \Phi(a_{pj}) + \Phi(a_{px}) \right) \\ &= \frac{n_t}{n_t+1} A_p + \frac{1}{n_t+1} \Phi(a_{px})\end{aligned}\quad (15)$$

2) 插入代表新样本  $x$  核空间关系向量

计算新样本  $x$  与每类样本的核关系, 并插入  $KM$  矩阵中, 则有:

$$\tilde{K} \tilde{M} = \left( \cdots, \frac{1}{\bar{n}_1} \sum_{a_i \in c_1} \Phi(a_{ix}), \cdots, \frac{1}{\bar{n}_c} \sum_{a_i \in c_c} \Phi(a_{ix}), \cdots \right)\quad (16)$$

可见更新后的  $\tilde{K} \tilde{M}$  是一个  $(n+1) \times c$  的矩阵, 插入的向

量  $\left[ \frac{1}{\bar{n}_1} \sum_{a_i \in c_1} \Phi(a_{ix}), \cdots, \frac{1}{\bar{n}_c} \sum_{a_i \in c_c} \Phi(a_{ix}) \right]$  即为  $x$  样本与每类样本的核空间关系。

通常情况下核方法构造核矩阵  $K$  计算量非常大, 从式(11)、式(14)中可以看到, IKDR/QR 算法只需要在第一次计算投影矩阵时构造完整的核矩阵  $K$ , 而在之后的增量计算中

只需更新两个规模更小的矩阵, 且在第一次计算中训练样本数较少, 需要构造的核矩阵  $K$  规模并不大。表 1 总结了 IKDR/QR 算法的完整流程。

表 1 增量式核判别分析算法 IKDR/QR 的流程

输入: 前一次增量计算中的矩阵 $M^T KM, KM$ , 每类样本的数量 $n_i (i=1, 2, \dots, c)$ , 新插入的样本 $x$
输出: 最佳投影矩阵 $G$
1. 更新矩阵 $M^T KM; \tilde{M}^T \tilde{K} \tilde{M} \leftarrow M^T KM$
2. 更新矩阵 $R^\circ; \tilde{R}^\circ \leftarrow \text{Cholesky}(\tilde{M}^T \tilde{K} \tilde{M})$
3. 更新低维类间散度矩阵 $B^\circ; \tilde{B}^\circ \leftarrow \tilde{M}^T \tilde{M}^T \tilde{K} \tilde{M} (\tilde{R}^\circ)^{-1}$
4. 更新矩阵 $KM; \tilde{K} \tilde{M} \leftarrow KM$
5. 更新低维总体散度矩阵 $T^\circ; \tilde{T}^\circ \leftarrow \tilde{E}^T \tilde{K} \tilde{M} (\tilde{R}^\circ)^{-1}$
6. 计算低维最佳投影矩阵 $V^\circ; J(V_{\text{opt}}^\circ) = \arg \max_{V^\circ} \frac{ \text{PT} \tilde{B} P }{ \text{PT} T P }$
7. 计算更新后的最佳投影矩阵 $G; G \leftarrow C^\circ (R^\circ)^{-1} V^\circ$

### 3.3 算法复杂度比较

IKDR/QR 与 KDR/QR 在算法性能上的差距主要体现在更新核矩阵  $K$  的方式上, KDR/QR 直接构造核矩阵  $K$  后再计算矩阵  $KM$  以及矩阵  $\tilde{M}^T \tilde{K} \tilde{M}$ , 其算法复杂度为  $O(n^2 + 2nc)$ , 其中  $c$  为样本类别数,  $n$  为样本总数。IKDR/QR 直接根据新样本调整矩阵  $KM$  以及矩阵  $\tilde{M}^T \tilde{K} \tilde{M}$ , 其算法复杂度为  $O(2n+c)$ 。因为增量式计算中与新插入样本不同类的数据对之间的核空间关系无需重复计算, 故计算量小于批量式方法。

对于空间复杂度, KDR/QR 要存储完整的核矩阵  $K$ , 需要一个  $n \times n$  的矩阵。而 IKDR/QR 需要存储矩阵  $KM$  以及矩阵  $M^T KM$ , 需要一个  $c \times (n+1)$  的矩阵。通常样本总数  $n$  要远大于样本的类别数  $c$ , 因此, 在减少存储空间方面, 增量式方法较批量式方法更有优势。

## 4 仿真实验

### 4.1 在 TE 过程上的实验

#### 4.1.1 数据预处理

Eastman 化学公司开发的 Tennessee Esatman 过程模拟器提供了一个实际工业过程的仿真平台<sup>[11]</sup>, 近年来被广泛用于验证各种故障检测与诊断方法, 成为一种国际上通用的标准仿真模型。TE 过程包括 5 个主要操作单元、4 种气体进料、2 个气液放热反应生成的 2 种主产品、2 个衍生放热反应生成的 2 种副产品等, 过程机理复杂, 变量较多, 包括阶跃、随机变化、慢漂移、粘滞和恒定位置等多种数据故障类型<sup>[12]</sup>。

TE 过程的仿真数据可以从 <http://depts.washington.edu/control/> 下载得到。仿真数据包括 15 组不同类别的故障数据, 每组故障数据包含 301 个样本和 51 种故障测量值, 构成  $4515 \times 51$  的数据集<sup>[13]</sup>。

在仿真实验中, 首先利用小波变换软阈值法<sup>[14]</sup>对样本数据进行小波去噪<sup>[15]</sup>, 这里采用三层小波分解, 并将小波重构后的信号作为原始样本输入算法。

#### 4.1.2 分类性能实验

分类性能实验选择 KFDA 算法以及 KDR/QR 算法在不同数量的训练数据下比较降维效果以反映在线学习问题中训练样本数量的变化对算法性能的影响。从每类样本中选择 20、50、100、200 个样本作为 4 组实验的训练样本, 并选择每类 50, 共 750 个样本作为测试数据。核函数分别选用高斯核

函数  $k(x, y) = \exp(-\|x - y\|^2 / \sigma)$ , 核参数  $\sigma$  取 400, 以及多项式核函数  $k(x, y) = (\langle x, y \rangle + 1)^d$ ,  $d$  取 2。通过 5-fold 交叉验证法选择适合的降维维度, 选为 30。选用欧式近邻法作为分类算法, 每组实验各进行 10 次, 取平均值作为结果。实验结果如图 1、图 2 所示。

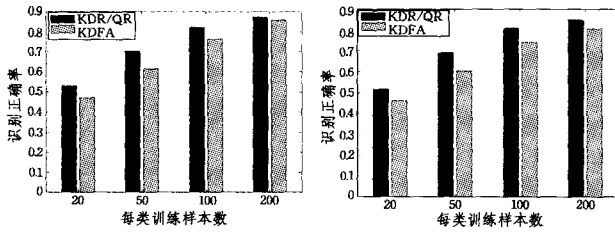


图 1 高斯核分类性能实验结果 图 2 多项式核分类性能实验结果

从图 1、图 2 中可以看到, 对于 2 种不同的核函数, KDR/QR 相比 KFDA 有着更好的降维效果。随着训练样本的增多, KFDA 的降维效果慢慢接近 KDR/QR, 但是在训练样本较少时 KDR/QR 的优势明显。这一优势特别适用于在线学习中初始训练样本较少的情况, 由于本文提出的 IKDR/QR 每一次更新投影矩阵的计算都建立在前一次计算上, 初始的训练效果会影响到后续的增量结果, 因此 KDR/QR 适合增量化。

#### 4.1.3 在线学习实验

这一部分将分析比较 IKDR/QR 与 KDR/QR 在线学习问题上的算法性能, 包括特征提取效果和计算速度两方面。在线学习问题的特点是初始训练数据较少, 随着训练的进行会有新的样本到来, 因此我们从每类样本中选取 20 个, 共 300 个作为初始训练样本, 同时选取每类 80 个样本, 共 1200 个作为待插入样本。核函数选择高斯核函数, 其他数值与 4.1.2 节实验数值一致, 将 1200 个样本依次插入训练样本, 实验结果如图 3、图 4 所示。

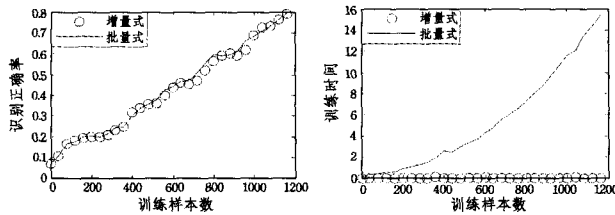


图 3 特征提取效果比较

图 4 训练时间比较

从图 3 可知, 增量式算法 IKDR/QR 与批量式算法 KDR/QR 的识别正确率相近, 这证明了 IKDR/QR 算法在非线系统在线学习过程中对样本空间特性的有效保持, 其在逐次更新的过程中相比进行完整降维计算的批量式方法没有损失较多的信息。另一方面, 从图 4 中可以看出随着新插入的样本数量增多, IKDR/QR 算法的时间开销一直维持在一个较低水平, 而批量式方法每次计算投影矩阵的耗时快速递增, 并且随着训练样本的增加, 增量式方法在速度上的优势越来越显著。为了更好地反映出 IKDR/QR 在计算效率上的优越性, 表 2 记录了插入新的训练样本后训练时间开销的变化。由表可知, 当插入 150 个样本时, 两者训练时间的差距只有 0.5 秒, 而当插入 1200 个样本时, 两者训练时间的差距达到了 16.6 秒。

表 2 训练时间开销(秒)

已插入样本数	150	300	450	600	750	900	1050	1200
增量式	0.008	0.015	0.017	0.023	0.031	0.042	0.044	0.062
批量式	0.593	1.342	2.574	4.306	6.521	9.142	12.309	16.677

#### 4.2 在 ORL 人脸数据库上的实验

ORL 人脸数据库<sup>[16]</sup>由剑桥大学 AT&T 实验室创建。该人脸库中共有 40 个人的不同姿态和表情的 400 个人脸图像, 每个人 10 幅, 图片的规格是  $112 \times 92$  的灰度图像。ORL 数据库中的实验数据经过预处理后分为 40 个类, 每类包含 10 个样本, 共 400 个样本, 每个样本大小为  $1 \times 1024$ 。

KFDA 算法与 KDR/QR 算法的比较: 本实验以每类的 5 个样本作为训练样本, 其余的 5 个样本作为测试样本, 这样训练样本和测试样本总数均为 200。选用的核函数为高斯核函数, 核参数  $\sigma$  取 400。实验结果如表 3 所列。

表 3 KFDA 算法与 KDR/QR 算法在 ORL 人脸库上不同降维维度下的识别率比较

降维维度	20	25	30	35	40
KFDA	0.705	0.725	0.765	0.830	0.865
KDR/QR	0.730	0.755	0.805	0.865	0.870

从表 3 可知, 降维维度一致的情况下, KDR/QR 较 KFDA 有更好的降维效果。当降维维度为 35 时, KDR/QR 的识别率已经趋于稳定且高于任何降维维度下 KFDA 的识别率。

IKDR/QR 算法与 KDR/QR 算法的比较: 本实验主要比较这两种算法在人脸识别增量学习问题上的算法性能。实验先随机选择每类 3 个样本作为首次训练样本, 然后每类增加 1 个样本作为新增训练样本, 将余下的作为测试样本。实验结果如图 5、图 6 所示。

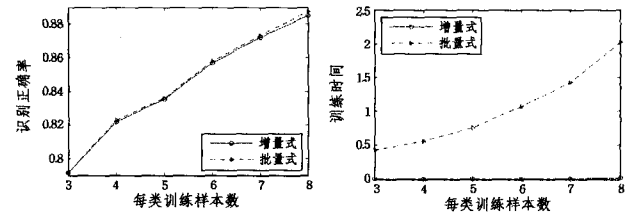


图 5 ORL 识别性能比较

图 6 ORL 时间开销比较

从图 5 中可以看出, IKDR/QR 和它的批量形式有着同样的识别正确率, 这证明了 IKDR/QR 算法在增量式计算过程中对样本空间特性的有效保持, 其在逐次更新的过程中相比进行完整降维计算的批量式方法没有损失较多的信息。此外, 从图 6 中可以看出随着新插入的样本数量增多, IKDR/QR 算法的时间开销一直维持在一个较低水平, 而批量式方法每次计算投影矩阵的耗时快速递增。

结束语 本文提出了一种基于 QR 分解的增量式核判别分析法, 其充分利用了基于 QR 分解的核判别分析法中的先降维后提取特征的思想, 将核空间映射到低维空间进行计算, 有效地解决了在线学习中冗余计算的问题, 减少了构造核矩阵的计算量, 降低了核矩阵的存储空间, 使得算法更适用于复杂的非线性工业设备系统在线学习。在 TE 过程数据和 ORL 人脸库上的仿真实验证明了该算法在特征提取上的有效性, 其相比批量式算法有较大的效率优势。由于本算法是在每类样本的训练样本数一致的前提下实现的, 而实际应用中很难

保证这一前提,因此算法存在一些问题有待进一步研究,即一些类别的多样性,一些类别的无样本性。这些问题导致空间结构信息的不平衡,影响特征提取效果。

### 参考文献

- [1] Peng H, Li B, Luo X, et al. A learning-based audio watermarking scheme using kernel Fisher discriminant analysis [J]. *Digital Signal Processing*, 2013, 23(1): 382-389
- [2] Meng Y, Zou J, Gan X, et al. Adaptive WNN aerodynamic modeling based on subset KPCA feature extraction [J]. *Journal of Central South University*, 2013, 20: 931-941
- [3] 胡益, 王丽, 马贺贺, 等. 基于核 PLS 方法的非线性过程在线监控[J]. *化工学报*, 2011, 62(9): 2555-2561
- [4] Gan J Y, Liu J F. Fusion and recognition of face and iris feature based on wavelet feature and KFPCA [C]//International Conference on Wavelet Analysis and Pattern Recognition, 2009. IC-WAPR 2009. IEEE, 2009: 47-50
- [5] Zhang X, Yan W, Zhao X, et al. Nonlinear biological batch process monitoring and fault identification based on kernel fisher discriminant analysis [J]. *Process biochemistry*, 2007, 42(8): 1200-1210
- [6] 郑建伟, 王万良, 姚晓敏, 等. 张量局部 Fisher 判别分析的人脸识别[J]. *自动化学报*, 2012, 38(9): 1485-1495
- [7] Jin L W, Ding K, Huang Z B. Incremental learning of FDA model for Chinese writer adaptation [J]. *Neurocomputing*, 2010, 73(10): 1614-1623
- [8] Xiong T, Ye J, Li Q, et al. Efficient kernel discriminant analysis via QR decomposition [J]. *Advances in neural information processing systems*, 2005, 17: 1529-1536
- [9] Golub G H, Van Loan C F. *Matrix computations* [M]. Johns Hopkins University Press, 1996
- [10] Zhang X, Yan W W, Zhao X, et al. Performance monitoring and fault identification based on DKPCA and PSVM [J]. *Journal of Shanghai Jiaotong University*, 2008, 42(2): 181-185
- [11] Chiang L H, Russell E, Braatz R D. *Fault detection and diagnosis in industrial systems* [M]. Springer Verlag, 2001
- [12] Bo C M, Zhang S, Zhang G M, et al. Fault identification of Tennessee Eastman process based on FS-KPCA [J]. *Journal of Chemical Industry and Engineering*, 2008, 59(7): 1783-1789
- [13] Ricker N L, Lee J H. Nonlinear model predictive control of the Tennessee Eastman challenge process [J]. *Computers & chemical engineering*, 1995, 19(9): 961-981
- [14] 张熠卓, 徐光华, 梁霖, 等. 利用增量式非线性流形学习的状态监测方法[J]. *西安交通大学学报*, 2011, 45(001): 64-68
- [15] 王迎, 王新明, 赵小强. 基于小波去噪与 KPCA 的 TE 过程故障检测研究[J]. *化工机械*, 2011, 38(1): 49-53
- [16] [http://blog.sina.com.cn/s/blog\\_02b6f23d0100eq88.html](http://blog.sina.com.cn/s/blog_02b6f23d0100eq88.html)

(上接第 279 页)

- [18] Li H, Jiang T, Zhang K. Efficient and robust feature extraction by maximum margin criterion [J]. *IEEE Transactions on Neural Networks*, 2006, 17(1-3): 157-165
- [19] Duchon J. Splines minimizing rotation-invariant semi-norms in Sobolev spaces [C]//Constructive theory of functions of several variables. 1977: 85-100
- [20] Meinguet J. Multivariate interpolation at arbitrary points made simple [J]. *Journal of Applied Mathematics and Physics*, 1979, 30: 292-304
- [21] Wahba G. *Spline models for observational data* [M]. SIAM Press, 1990
- [22] Söderkvist O. Computer vision classification of leaves from Swedish trees [D]. Master's Thesis, Linköping University, 2001
- [23] Ye J. Characterization of a family of algorithms for generalized discriminant analysis on undersampled problems [J]. *Journal of Machine Learning Research*, 2006, 6: 483-502
- [24] Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces vs. fisherfaces; Recognition using class specific linear projection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, 19(7): 711-720
- [25] Cai D, He X, Han J. Using graph model for face analysis [R]. Univ. Illinois Urbana-Champaign, Urbana, IL, Department of Computer Science, Technical Report, 2005
- [26] Zhang T, Yang J, Zhao D, et al. Linear local tangent space alignment and application to face recognition [J]. *Neurocomputing*, 2007, 70(7-9): 1547-1553
- [27] De Silva V, Tenenbaum J B. Global versus local methods in nonlinear dimensionality reduction [C]//Proceedings of Advances in Neural Information Processing Systems. 2003, 15: 721-728

(上接第 286 页)

- [13] Cai Guo-liang, Hu Pei, Li Yu-xiu. Modified function lag projective synchronization of a financial hyperchaotic system [J]. *Nonlinear Dynamics*, 2012, 69(3): 1457-1464
- [14] 刘恒, 余海军, 向伟. 带未知扰动的多涡卷混沌系统修正函数时滞投影同步 [J]. *物理学报*, 2012, 61(18): 180503
- [15] Gao Yan-bo, Sun Bing-hua, Lu Guo-ping. Modified function projective lag synchronization of chaotic systems with disturbance estimations [J]. *Applied Mathematical Modelling*, 2013, 37(7): 4993-5000
- [16] Chai Xiu-li, Gan Zhi-hua, Shi Chun-xiao. Adaptive modified function projective lag synchronization of uncertain hyperchaotic dynamical systems with the same or different dimension and structure [J]. *Mathematical Problems in Engineering*, 2013, 2013: 282064
- [17] Fu Gui-yuan. Robust adaptive modified function projective synchronization of different hyperchaotic systems subject to external disturbance [J]. *Communications in Nonlinear Science and Numerical Simulation*, 2012, 17(6): 2602-2608
- [18] Yu Fei, Wang Chun-hua, Hu Yan, et al. Antisynchronization of a novel hyperchaotic system with parameter mismatch and external disturbances [J]. *Pramana-Journal of physics*, 2012, 79(1): 81-93
- [19] 邓玮, 方洁, 吴振军, 等. 含有不确定项的混沌系统自适应修正函数投影同步 [J]. *物理学报*, 2012, 61(14): 140503