

最佳鉴别矢量集在无监督模式下的扩展

曹苏群^{1,2}, 王士同¹, 陈晓峰¹, 邓赵红¹

(1. 江南大学 信息学院, 江苏 无锡 214122; 2. 淮阴工学院 机械系, 江苏 淮安 223001)

摘要:基于 Fisher 准则函数的最佳鉴别矢量集是一种重要的有监督特征提取方法,在模式识别领域有着重要的影响. 提出一种将最佳鉴别矢量集扩展到无监督模式下的方法,其基本思想是通过定义的模糊 Fisher 准则函数将 Fisher 线性判别扩展成一种半模糊聚类算法,通过该算法求得最佳鉴别矢量和模糊散布矩阵,进而构造出最佳鉴别矢量集. 实验表明,在聚类有效性、分类准确率均优于无监督模式下常用的主成分分析特征提取算法.

关键词:最佳鉴别矢量集;无监督模式;Fisher 准则;半模糊聚类

中图分类号: TP181 **文献标识码:** A **文章编号:** 1673-4785(2008)06-0511-12

Extending the optimal set of discriminant vectors for an unsupervised pattern

CAO Su-qun^{1,2}, WANG Shi-tong¹, CHEN Xiao-feng¹, DENG Zhao-hong¹

(1. School of Information, Jiangnan University, Wuxi 214122, China; 2. Department of Mechanical Engineering, Huaiyin Institute of Technology, Huaian 223001, China)

Abstract: The optimal set of discriminant vectors, based on the Fisher criterion function, is an important supervised feature extraction method and has great influence in the area of pattern recognition. In this paper, an extension of the optimal set of discriminant vectors in unsupervised patterns is presented. The basic idea is to extend Fisher linear discriminants to a novel semi-fuzzy clustering algorithm through a predefined fuzzy Fisher criterion function. With the proposed algorithm, an optimal discriminant vector and fuzzy scatter matrixes can be figured out and then an unsupervised optimal set of discriminant vectors can be obtained. Experimental results for real datasets, testing clustering validity and correct classification recognition rates, demonstrated that this method is superior to the principal component analysis feature extraction algorithm in unsupervised patterns.

Keywords: optimal set of discriminant vectors; unsupervised pattern; Fisher criterion; semi-fuzzy clustering

在模式识别领域中, Fisher 线性判别^[1]方法应用广泛,其基本思想是在 Fisher 准则函数取极值条件下,求得一个最佳鉴别方向,然后将模式高维特征向量投影到该最佳鉴别方向上,构成一维鉴别特征空间,进而在此一维空间完成模式鉴别分析. Sammon 在 1970 年发展了 Fisher 线性判别方法,提出了最佳鉴别平面技术^[2],将它用于解决两类问题,实现了高维特征空间向二维空间的投影转换. Foly 和 Sammon 进一步对两类问题提出了最佳鉴别矢量集^[3]. Duchene 和 Leclercq 给出了多类问题的最佳鉴

别矢量计算公式^[4],实现了高维特征空间向低维空间的投影转换. Jin 和 Yang 等在文献[5]中给出了较 Duchene 和 Leclercq 更为简单的最佳鉴别矢量计算公式并提出了具有统计不相关性的最佳鉴别矢量集. 此后,文献[6-11]对此进行了进一步探讨. 由于在无监督模式下,通常无法使用 Fisher 线性判别方法求解最佳鉴别矢量,因此所有这些最佳鉴别矢量集计算方法均只适用于样本类别信息已知情况下.

本文提出一种将最佳鉴别矢量集扩展到无监督模式下的方法,其基本思想是:通过给出模糊 Fisher 准则函数定义,实现 Fisher 线性判别在无监督模式下的扩展,进而求得无监督模式下基于模糊 Fisher 准则的最佳鉴别矢量,在此基础上参照传统最佳鉴别矢量集计算方法完成其在无监督模式下的扩展.

收稿日期:2008-05-12.

基金项目:教育部优秀人才支持计划资助项目(NCET-04-0496);教育部重点科学研究资助项目(105087).

通信作者:曹苏群. E-mail: caosuqun@126.com.

按照文献[12]中对无监督情况下的特征降维方法的分类,从本质上讲,本文提出的方法属于聚类后降维^[13-14],即通过聚类得到样本类别信息,从而将有监督特征降维方法引入无监督模式。

1 最佳鉴别矢量集

设 X 包含 N 个 d 维样本 x_1, x_2, \dots, x_N , 可用 $d \times N$ 矩阵表示, 模式类别有 c 个, 各类别样本数分别为 $N_i (i=1, 2, \dots, c)$, $N_1 + N_2 + \dots + N_c = N$. 定义模式类间散布矩阵 S_b 和类内散布矩阵 S_w 分别为

$$S_b = \sum_{i=1}^c \frac{N_i}{N} (\mathbf{m}_i - \bar{\mathbf{x}}) (\mathbf{m}_i - \bar{\mathbf{x}})^T,$$

$$S_w = \frac{1}{N} \sum_{i=1}^c \sum_{j=1}^{N_i} (\mathbf{x}_j^i - \mathbf{m}_i) (\mathbf{x}_j^i - \mathbf{m}_i)^T.$$

式中: \mathbf{x}_j^i 表示属于第 i 类的样本, $\mathbf{m}_i (i=1, 2, \dots, c)$ 是第 i 类样本的均值向量, $\bar{\mathbf{x}}$ 是总体样本的均值向量。

定义 Fisher 准则函数为

$$J_F(\omega) = \frac{\omega^T S_b \omega}{\omega^T S_w \omega}, \omega \neq 0. \quad (1)$$

使 $J_F(\omega)$ 取得极大值的矢量 ω , 称为最佳鉴别矢量, 此时, 样本在 ω_1 方向投影将具有最大类间散布和最小类内散布. 使用 Lagrange 乘子法求解知, 当 S_w 非奇异时, 有

$$S_b \omega_1 = \lambda S_w \omega_1. \quad (2)$$

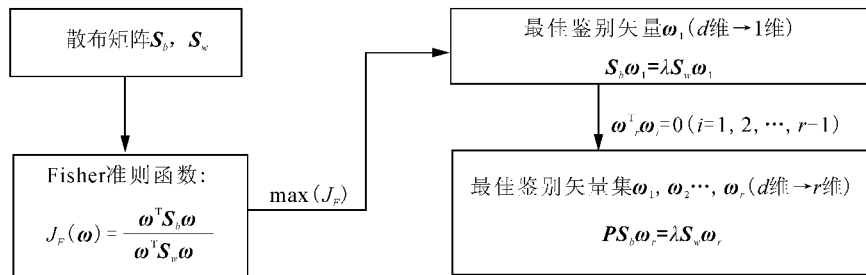


图1 Fisher 准则函数及最佳鉴别矢量、最佳鉴别矢量集求解关系

Fig. 1 Relationships between Fisher criterion function and optimal set of discriminant vectors

2 无监督模式下最佳鉴别矢量集

2.1 模糊 Fisher 准则函数

参照 Wu 等在文献[15]中模糊散布矩阵定义, 将模糊理论引入 Fisher 判别方法。

在样本空间, 定义模糊类内散布矩阵记为 S_{fw} ,

其中: λ 为 Lagrange 乘子, ω_1 即该广义特征方程中最大的特征值对应的特征向量. 对于两类问题, 可以求得 $\omega_1 = S_w^{-1}(\mathbf{m}_1 - \mathbf{m}_2)$. 通过 $Y = \omega_1^T X$, 实现模式 X 由 d 维向 1 维的投影转换。

若已求出 $k (k \geq 1)$ 个最佳鉴别矢量 $\omega_1, \omega_2, \dots, \omega_k$, 第 k 个最佳鉴别矢量 ω_{k+1} 是在满足下列正交条件下的使得式(1)取得最大值的向量:

$$\omega_{k+1}^T \omega_i = 0, i = 1, 2, \dots, k. \quad (3)$$

由文献[5]知, ω_{k+1} 是下列广义特征方程中最大的特征值对应的特征向量:

$$P S_b \omega_{k+1} = \lambda S_w \omega_{k+1} \quad (4)$$

其中 $P = I - D^T (D S_w^{-1} D^T)^{-1} D S_w^{-1}$, 这里记 $D = [\omega_1 \ \omega_2 \ \dots \ \omega_k]^T$.

求出的 $r (r \leq d)$ 个最佳鉴别矢量 $\omega_1, \omega_2, \dots, \omega_r$, 构成最佳鉴别矢量集, 通过

$$Y = \begin{bmatrix} \omega_1^T \\ \omega_2^T \\ \vdots \\ \omega_r^T \end{bmatrix} X$$

实现模式 X 由 d 维向 r 维的投影转换。

综上所述, 可以用下图表示 Fisher 准则函数、最佳鉴别矢量、最佳鉴别平面和最佳鉴别矢量集的求解关系。

$$S_{fw} = \sum_{i=1}^c \sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{m}_i) (\mathbf{x}_j - \mathbf{m}_i)^T \quad (5)$$

模糊类间散布矩阵记为 S_{fb} ,

$$S_{fb} = \sum_{i=1}^c \sum_{j=1}^N u_{ij}^m (\mathbf{m}_i - \bar{\mathbf{x}}) (\mathbf{m}_i - \bar{\mathbf{x}})^T \quad (6)$$

式中: $\bar{\mathbf{x}}$ 为所有样本均值向量, \mathbf{m}_i 为第 i 类样本均值向量, c 为类别数, N 为样本总数, u_{ij} 为 j 样本属于第

i 类的隶属度, $u_{ij} \in [0, 1]$ 且 $\sum_{i=1}^c u_{ij} = 1$, m 为模糊指数且 $m > 1$, 一般可取 2.

设 $y = \omega^T x$, 在该投影空间, 定义各类样本均值向量记为 \tilde{m}_i , 有

$$\tilde{m}_i = \omega^T m_i. \quad (7)$$

定义模糊类内散布矩阵记为 \tilde{S}_{fw} ,

$$\begin{aligned} \tilde{S}_{fw} &= \sum_{i=1}^c \sum_{j=1}^N u_{ij}^m (y_j - \tilde{m}_i)(y_j - \tilde{m}_i)^T = \\ &= \sum_{i=1}^c \sum_{j=1}^N u_{ij}^m (\omega^T x_j - \omega^T m_i)(\omega^T x_j - \omega^T m_i)^T = \\ &= \omega^T S_{fw} \omega \end{aligned}$$

模糊类间散布矩阵记为 \tilde{S}_{fb} .

$$\begin{aligned} \tilde{S}_{fb} &= \sum_{i=1}^c \sum_{j=1}^N u_{ij}^m (\tilde{m}_i - \bar{y})(\tilde{m}_i - \bar{y})^T = \\ &= \sum_{i=1}^c \sum_{j=1}^N u_{ij}^m (\omega^T m_i - \omega^T \bar{x})(\omega^T m_i - \omega^T \bar{x})^T = \\ &= \omega^T S_{fb} \omega \end{aligned}$$

定义模糊 Fisher 准则(fuzzy Fisher criterion)

$$u_{ij} = \frac{(\omega^T (x_j - m_i)(x_j - m_i)^T \omega - \frac{1}{\lambda} \omega^T (m_i - \bar{x})(m_i - \bar{x})^T \omega)^{-\frac{1}{m-1}}}{\sum_{k=1}^c (\omega^T (x_j - m_k)(x_j - m_k)^T \omega - \frac{1}{\lambda} \omega^T (m_k - \bar{x})(m_k - \bar{x})^T \omega)^{-\frac{1}{m-1}}}. \quad (12)$$

解式(10)为求一般矩阵 $S_{fw}^{-1} S_{fb}$ 的本征值问题, 取 λ 为该矩阵的最大特征值, 而 ω 为对应的特征向量.

在模糊聚类中, 通常限定 $u_{ij} \in [0, 1]$; 因此, 对式(12)给出如下限定条件, 若

$$\begin{aligned} \omega^T (x_j - m_i)(x_j - m_i)^T \omega &\leq \\ \frac{1}{\lambda} \omega^T (m_i - \bar{x})(m_i - \bar{x})^T \omega, \end{aligned} \quad (13)$$

则: $u_{ij} = 1$, 且对所有 $i' \neq i$, 有 $u_{i'j} = 0$. 也就是说, 当该条件满足时, 样本点 x_j 将完全隶属于第 i 类, 即在此范围内采用硬划分. 根据此条件几何意义知: 将样本点、聚类中心以及所有样本中心向 ω 方向投影, 若样本投影点与聚类中心投影点欧氏距离小于或等于聚类中心投影点与所有样本中心投影点间欧氏距离的 $\frac{1}{\sqrt{\lambda}}$, 则对该点采用硬划分. 可以使用图 2, 对此作出直观解释.

图中: “□”为第 1 类样本点, “○”为第 2 类样本点, 样本点中带“·”标记的为满足式(13)而采用硬划分的点, 将这些点按虚线方向向最佳鉴别矢量方向线投影.

根据求解结果, 本算法描述如下:

函数:

$$J_{ffo} = \frac{\tilde{S}_{fb}}{\tilde{S}_{fw}} = \frac{\omega^T S_{fb} \omega}{\omega^T S_{fw} \omega}. \quad (8)$$

2.2 基于模糊 Fisher 准则的算法

使用 Lagrange 乘子法求解 J_{ffo} 取得极大值时, 取 ω, m_i 和 u_{ij} 的值.

定义 Lagrange 函数为

$$L = \omega^T S_{fb} \omega - \lambda \omega^T S_{fw} \omega + \sum_{j=1}^N \lambda_j (\sum_{i=1}^c u_{ij} - 1) \quad (9)$$

式中: λ 和 $\lambda_j (j=1, 2, \dots, n)$ 为 Lagrange 乘子.

将 L 分别对 ω, m_i 及 u_{ij} 求偏导数, 并令偏导数为零, 经附录 A 求解步骤可求得

$$S_{fb} \omega = \lambda S_{fw} \omega, \quad (10)$$

$$m_i = \frac{\sum_{j=1}^N u_{ij}^m (x_j - \frac{1}{\lambda} \bar{x})}{\sum_{j=1}^N u_{ij}^m (1 - \frac{1}{\lambda})}, \quad (11)$$

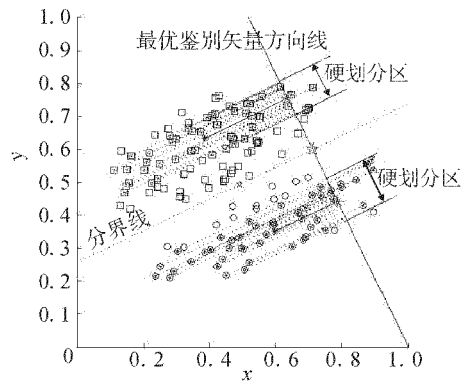


图2 FFC 聚类硬划分示意图

Fig. 2 Fuzzy Fisher criterion based clustering hard partition regions

1) 使用 K-means 算法初始化隶属矩阵 U 以及聚类中心 m_i ;

2) 使用式(5)、(6)分别计算 S_{fw} 和 S_{fb} , 根据式(10)求得广义特征方程的最大特征值 λ , 并取 ω 为属于 λ 的模为 1 的特征向量;

3) 使用式(8)计算模糊 Fisher 准则函数 J_{ffo} , 若它相对上次准则函数数值的改变量小于某个给定阈值或迭代次数超过设定次数, 则算法停止;

4)使用式(12)、(11)分别计算新的隶属矩阵 U 以及聚类中心 m_i , 返回步骤2)。

2.3 无监督模式下最佳鉴别矢量集

通过上述算法,由式(10)求得最佳鉴别矢量,设为 ω_1 ,并根据求得的隶属矩阵 U 和聚类中心 m_i ,由式(5)和(6)求得 S_{fw} 和 S_{fb} 。

若已求出 $k(k \geq 1)$ 个最佳鉴别矢量 $\omega_1, \omega_2, \dots, \omega_k$, 第 $k+1$ 个最佳鉴别矢量 ω_{k+1} 是在满足下列正交条件下的使得式(1)取得最大值的向量。

$$\omega_{k+1}^T \omega_i = 0, i = 1, 2, \dots, k.$$

参照式(4)知, ω_{k+1} 是下列广义特征方程中最大的特征值对应的特征向量:

$$PS_{fb}\omega_{k+1} = \lambda S_{fw}\omega_{k+1}.$$

其中 $P = I - D^T(DS_{fw}^{-1}D^T)^{-1}DS_{fw}^{-1}$, 这里记 $D = [\omega_1 \ \omega_2 \ \dots \ \omega_k]^T$ 。

求出的 $r(r \leq d)$ 个最佳鉴别矢量 $\omega_1, \omega_2, \dots, \omega_r$ 构成最佳鉴别矢量集, 同样通过

$$Y = \begin{bmatrix} \omega_1^T \\ \omega_2^T \\ \vdots \\ \omega_r^T \end{bmatrix} X$$

实现模式 X 由 d 维向 r 维的投影转换。

综上所述,对照图1,可以用图3表示最佳鉴别矢量集在无监督模式下的扩展方法。

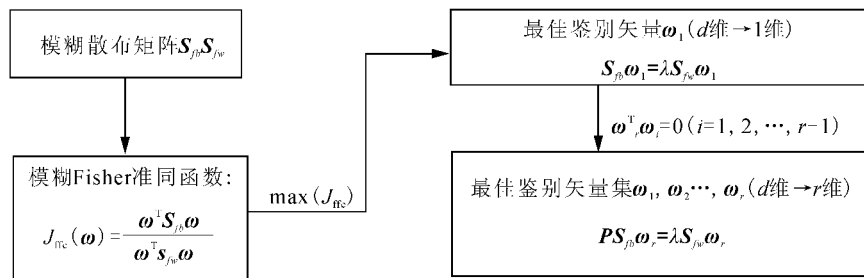


图3 模糊 Fisher 准则函数及最佳鉴别矢量、最佳鉴别矢量集求解关系

Fig. 3 Relationships between fuzzy Fisher criterion function and unsupervised optimal set of discriminant vectors

3 实验结果及其分析

采用 UCI^[16] 中 Iris, Wdbc 等 4 个数据集作为实验数据,这 4 个数据集基本信息如表 1 所示。

表 1 实验所用数据集的基本信息

Table 1 Basic information of experimental dataset

数据集	类别数	样本数	特征数
Iris	3	150	4
Pima_indians_diabetes	2	768	8
Wine	3	178	13
Wdbc	2	569	30

3.1 聚类有效性

使用 FCM 和 K-means 算法分别对使用主成分分析(principal component analysis, PCA)、最佳鉴别矢量集(optimal set of discriminant vectors, OSDV)和无监督最佳鉴别矢量集(unsupervised optimal set of discriminant vectors, UOSDV)降维后的数据进行聚类,将聚类结果与原数据集类别标记相比较,通过著

名的约当指数(Rand index, RI)^[17]作为聚类有效性评价指标。

RI 具体解释如下:定义 P_1, P_2 分别为原数据集 D 的类别标记和变换后数据集聚类划分结果,通过公式

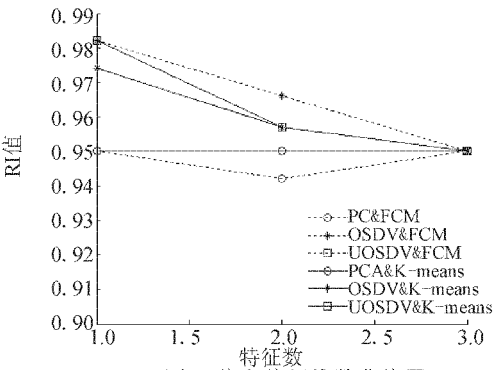
$$R(P_1, P_2) = \frac{a + b}{n \times (n - 1) / 2}$$

来计算这两种划分的一致性。其中: a 表示 D 中任意两个样本 d_i, d_j 在 P_1, P_2 中同属于一类的个数; b 表示 d_i, d_j 都不属于同一类的个数; n 表示数据集 D 的样本个数。显然, RI 的范围为 $[0, 1]$, 只有在 P_1, P_2 完全一致的情况下, RI 值为 1。RI 统计的值越小,两种划分差距越大。为了保证比较的准确性和公平性,两种算法分别进行 100 次,计算测得的 RI 均值,并绘制 FCM 和 K-means 分别对降维数据进行聚类所得 RI 值和特征维数曲线图以及分别采用 PCA, OSDV 和 UOSDV 降维后 2 维数据散布图如图 4~7 和表 2~5 所示。

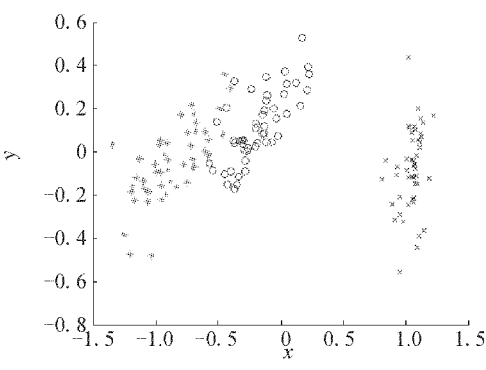
表 2 采用 FCM 和 K-means 分别对经 PCA, OSDV 和 UOSDV 降维的 Iris 数据集进行聚类所得 RI 值

Table 2 Rand index of FCM and K-means for reduced-dimension Iris dataset extracted by PCA, OSDV and UOSDV

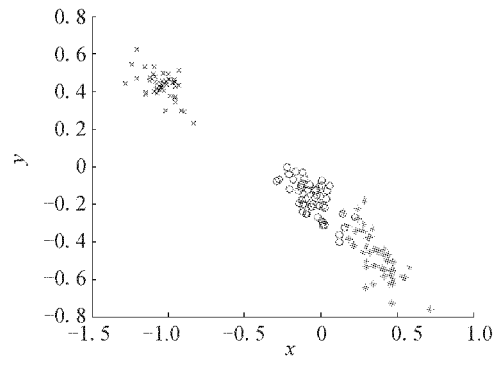
Number of features	PCA		OSDV		UOSDV	
	RI with FCM	RI with K-means	RI with FCM	RI with K-means	RI with FCM	RI with K-means
1	0.950	0.950	0.982	0.974	0.982	0.982
2	0.942	0.950	0.966	0.957	0.957	0.957
3	0.950	0.950	0.950	0.950	0.950	0.950



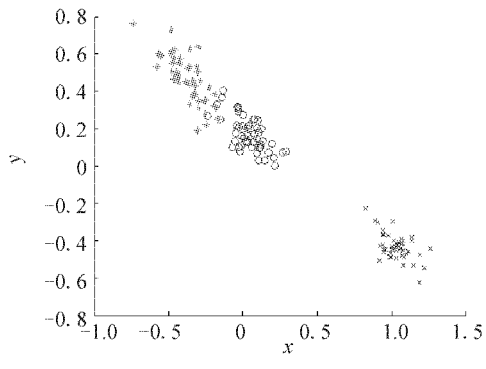
(a) RI值和特征维数曲线图



(b) 采用PCA降为2维的数据散布图



(c) 采用OSDV降为2维的数据散布图



(d) 采用UOSDV降为2维的数据散布图

图 4 对 Iris 数据集分别使用 PCA, OSDV 和 UOSDV 降维后分析结果

Fig. 4 Results for the analysis of reduced-dimension Iris dataset extracted by PCA, OSDV and UOSDV

表 3 采用 FCM 和 K-means 分别对经 PCA, OSDV 和 UOSDV 降维的 Pima_indians_diabetes 数据集进行聚类所得 RI 值

Table 3 Rand index of FCM and K-means for reduced-dimension Pima_indians_diabetes dataset extracted by PCA, OSDV and UOSDV

Number of features	PCA		OSDV		UOSDV	
	RI with FCM	RI with K-means	RI with FCM	RI with K-means	RI with FCM	RI with K-means
1	0.549	0.551	0.643	0.643	0.553	0.555
2	0.550	0.551	0.626	0.625	0.549	0.552
3	0.550	0.551	0.625	0.625	0.551	0.552
4	0.550	0.551	0.616	0.617	0.554	0.553
5	0.550	0.551	0.606	0.617	0.552	0.556
6	0.550	0.551	0.591	0.596	0.556	0.553
7	0.550	0.551	0.551	0.552	0.549	0.549

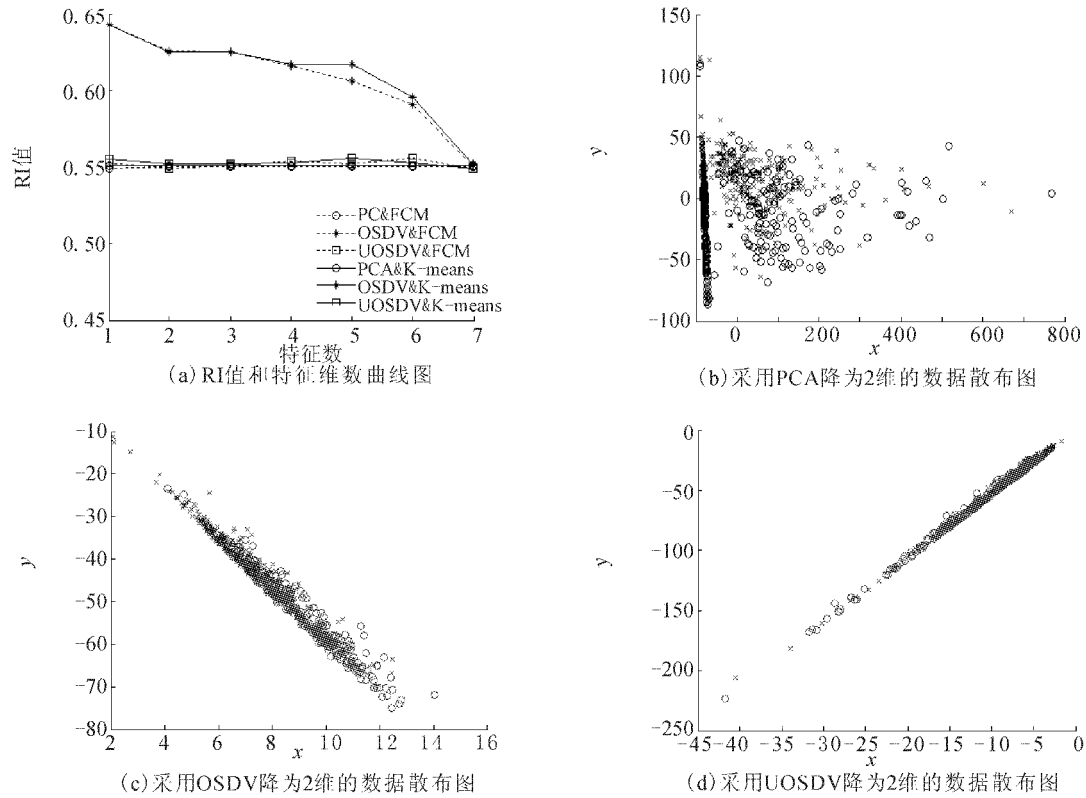


图5 对 Pima_indians_diabetes 数据集分别使用 PCA, OSDV 和 UOSDV 降维后分析结果

Fig.5 Results for the analysis of reduced-dimension Pima_indians_diabetes dataset extracted by PCA, OSDV and UOSDV

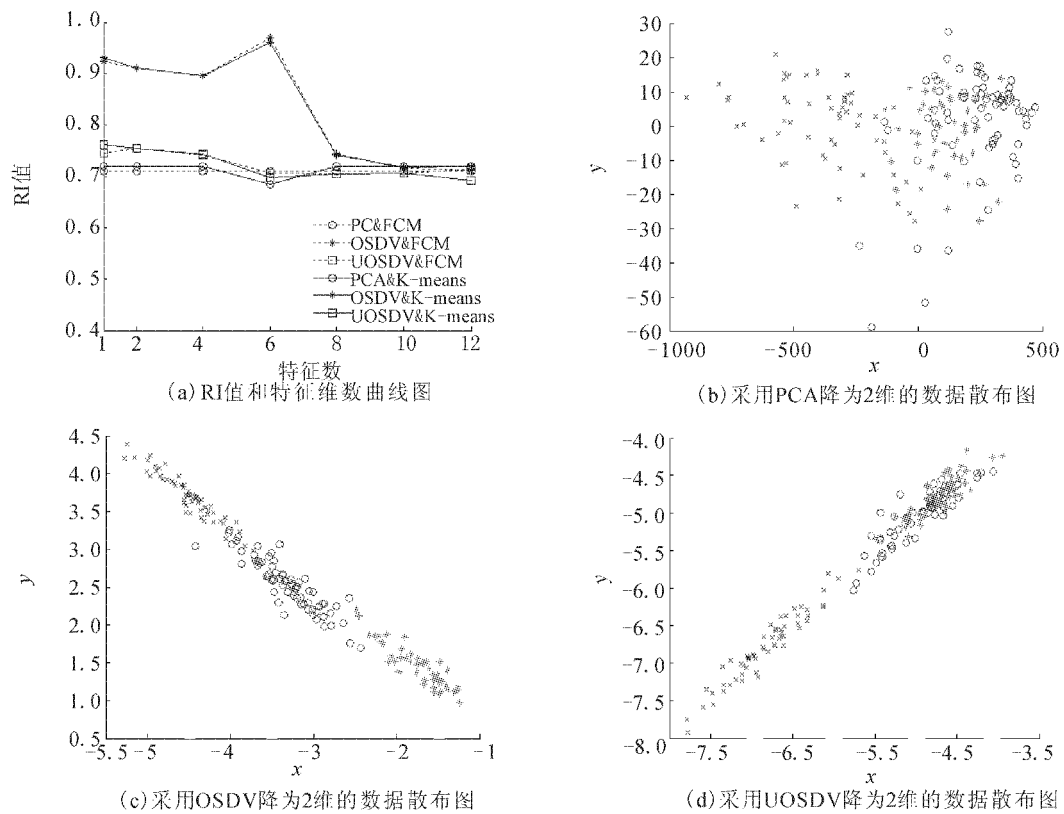


图6 对 Wine 数据集分别使用 PCA, OSDV 和 UOSDV 降维后分析结果

Fig.6 Results for the analysis of reduced-dimension Wine dataset extracted by PCA, OSDV and UOSDV

表 4 采用 FCM 和 K-means 分别对经 PCA, OSDV 和 UOSDV 降维的 Wine 数据集进行聚类所得 RI 值
Table 4 Rand index of FCM and K-means for reduced-dimension Wine dataset extracted by PCA, OSDV

Number of features	PCA		OSDV		UOSDV	
	RI with FCM	RI with K-means	RI with FCM	RI with K-means	RI with FCM	RI with K-means
1	0.711	0.719	0.924	0.931	0.745	0.761
2	0.711	0.719	0.910	0.910	0.754	0.754
4	0.711	0.719	0.896	0.896	0.742	0.744
6	0.711	0.685	0.969	0.961	0.707	0.698
8	0.711	0.719	0.744	0.741	0.704	0.704
10	0.711	0.719	0.715	0.717	0.707	0.707
12	0.711	0.719	0.712	0.719	0.714	0.692

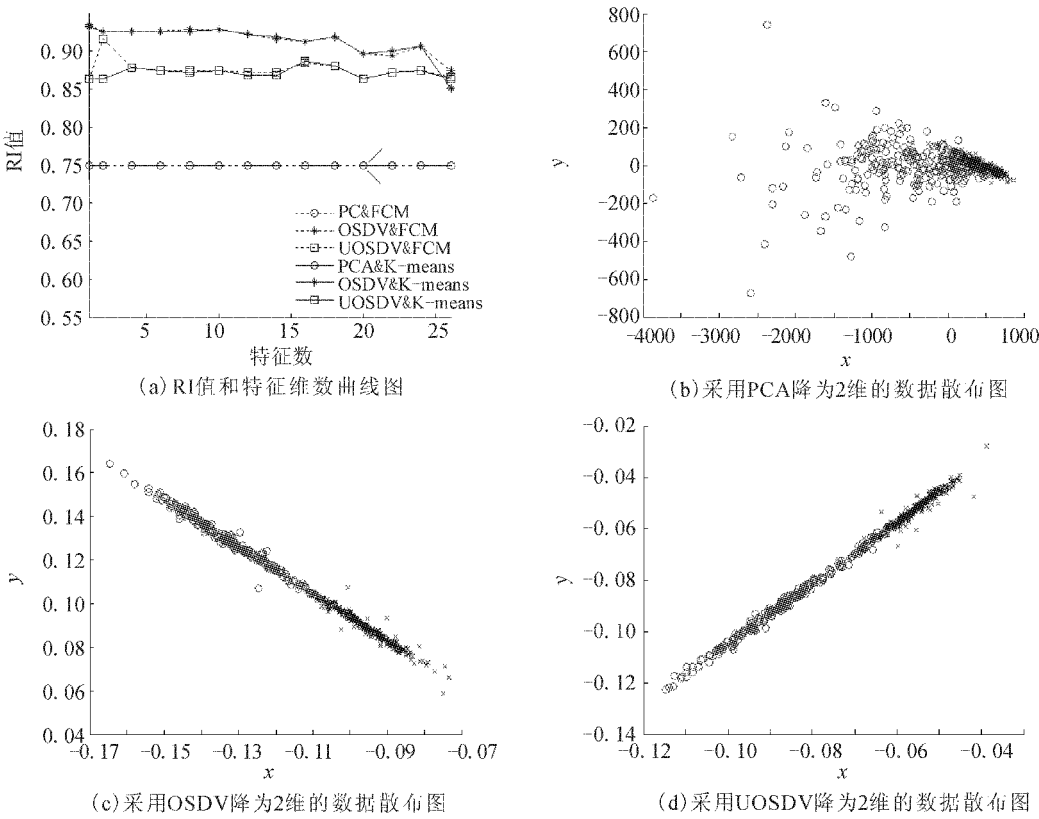


图 7 对 Wdbc 数据集分别使用 PCA, OSDV 和 UOSDV 降维后分析结果
Fig. 7 Results for the analysis of reduced-dimension Wdbc dataset extracted by PCA, OSDV and UOSDV

表 5 采用 FCM 和 K-means 分别对经 PCA, OSDV 和 UOSDV 降维的 Wdbc 数据集进行聚类所得 RI 值
Table 5 Rand index of FCM and K-means for reduced-dimension Wdbc dataset extracted by PCA, OSDV and UOSDV

Number of features	PCA		OSDV		UOSDV	
	RI with FCM	RI with K-means	RI with FCM	RI with K-means	RI with FCM	RI with K-means
1	0.750	0.750	0.935	0.932	0.863	0.863

续表 5

Number of features	PCA		OSDV		UOSDV	
	RI with FCM	RI with K-means	RI with FCM	RI with K-means	RI with FCM	RI with K-means
2	0.750	0.750	0.926	0.926	0.916	0.863
4	0.750	0.750	0.926	0.926	0.878	0.878
6	0.750	0.750	0.926	0.926	0.875	0.875
8	0.750	0.750	0.929	0.926	0.875	0.872
10	0.750	0.750	0.929	0.929	0.875	0.875
12	0.750	0.750	0.922	0.922	0.872	0.869
14	0.750	0.750	0.916	0.919	0.872	0.869
16	0.750	0.750	0.913	0.913	0.884	0.887
18	0.750	0.750	0.919	0.919	0.881	0.881
20	0.750	0.750	0.897	0.897	0.863	0.863
22	0.750	0.750	0.894	0.900	0.872	0.872
24	0.750	0.750	0.906	0.906	0.875	0.875
26	0.750	0.750	0.875	0.851	0.866	0.863

从以上图表中看出,与 OSDV 有监督特征提取方法相比,无监督模式下样本信息的缺失,使得 UOSDV 无监督特征提取方法性能有一定下降;但是 UOSDV 与同属于无监督特征提取方法的 PCA 相比,不论是采用 FCM 还是 K-means,其降维数据的聚类结果均具有较大的 RI 值,说明了其降维数据聚类结果与实际数据划分有较好的一致性.从降维数据散布图中,可以直观看出 UOSDV 降维数据与 PCA 降维数据相比,具有更好的可分性.

3.2 分类准确率

分别使用主成分分析(PCA)、最佳鉴别矢量集(OSDV)和无监督最佳鉴别矢量集(UOSDV)方法进行降维,通过最近邻分类器(1-nearest neighbor, 1-NN)和 LibSVM^[18]分别对降维后数据进行留一法交叉验证.即依次取降维后数据中的一个作为测试集,用其他所有的数据作为训练集,通过训练集训练分类器后测试该含一个数据的测试集,将测试结果与已知类别信息进行比较,计算分类准确率.表 6~9 为 1-NN 和 LIBSVM 对这 4 个数据集分别进行留一法交叉验证的分类准确率实验结果.

表 6 采用 1-NN 和 LibSVM 分别对经 PCA, OSDV 和 UOSDV 降维的 Iris 数据集进行留一法交叉验证所得分类准确率
Table 6 Leave-one-out rate with 1-NN and LibSVM for reduced-dimension Iris dataset extracted by PCA, OSDV and UOSDV

Number of features	PCA		OSDV		UOSDV	
	Leave-one-out rate with 1-NN	Leave-one-out rate with LibSVM	Leave-one-out rate with 1-NN	Leave-one-out rate with LibSVM	Leave-one-out rate with 1-NN	Leave-one-out rate with LibSVM
1	0.940	0.953	0.953	0.980	0.973	0.980
2	0.953	0.960	0.960	0.967	0.960	0.967
3	0.947	0.953	0.953	0.967	0.953	0.973

表7 采用1-NN和LibSVM分别对经PCA, OSDV和UOSDV降维的Pima_indians_diabetes数据集进行留一法交叉验证所得分类准确率

Table 7 Leave-one-out rate with 1-NN and LibSVM for reduced-dimension Pima_indians_diabetes dataset extracted by PCA, OSDV and UOSDV

Number of features	PCA		OSDV		UOSDV	
	Leave-one-out	Leave-one-out	Leave-one-out	Leave-one-out	Leave-one-out	Leave-one-out
	rate with 1-NN	rate with LibSVM	rate with 1-NN	rate with LibSVM	rate with 1-NN	rate with LibSVM
1	0.637	0.656	0.682	0.777	0.583	0.660
2	0.664	0.667	0.667	0.772	0.578	0.643
3	0.639	0.618	0.689	0.707	0.565	0.620
4	0.673	0.637	0.695	0.680	0.615	0.600
5	0.686	0.655	0.685	0.638	0.672	0.645
6	0.682	0.651	0.664	0.642	0.643	0.658
7	0.680	0.654	0.678	0.655	0.685	0.658

表8 采用1-NN和LibSVM分别对经PCA, OSDV和UOSDV降维的Wine数据集进行留一法交叉验证所得分类准确率

Table 8 Leave-one-out rate with 1-NN and LibSVM for reduced-dimension Wine dataset extracted by PCA, OSDV and UOSDV

Number of features	PCA		OSDV		UOSDV	
	Leave-one-out	Leave-one-out	Leave-one-out	Leave-one-out	Leave-one-out	Leave-one-out
	rate with 1-NN	rate with LibSVM	rate with 1-NN	rate with LibSVM	rate with 1-NN	rate with LibSVM
1	0.708	0.640	0.944	0.938	0.764	0.702
2	0.719	0.472	0.944	0.961	0.691	0.730
4	0.758	0.461	0.972	0.983	0.736	0.742
6	0.770	0.449	0.983	0.983	0.663	0.657
8	0.770	0.444	0.972	0.978	0.736	0.831
10	0.770	0.444	0.944	0.966	0.888	0.899
12	0.770	0.444	0.803	0.584	0.770	0.685

表9 采用1-NN和LibSVM分别对经PCA, OSDV和UOSDV降维的Wdbc数据集进行留一法交叉验证所得分类准确率

Table 9 Leave-one-out rate with 1-NN and LibSVM for reduced-dimension Wdbc dataset extracted by PCA, OSDV and UOSDV

Number of features	PCA		OSDV		UOSDV	
	Leave-one-out	Leave-one-out	Leave-one-out	Leave-one-out	Leave-one-out	Leave-one-out
	rate with 1-NN	rate with LibSVM	rate with 1-NN	rate with LibSVM	rate with 1-NN	rate with LibSVM
1	0.859	0.852	0.963	0.956	0.931	0.924
2	0.914	0.627	0.968	0.963	0.933	0.935
4	0.916	0.627	0.961	0.961	0.935	0.942
6	0.914	0.627	0.965	0.970	0.937	0.947

续表9

Number of features	PCA		OSDV		UOSDV	
	Leave-one-out rate with	Leave-one-out rate with	Leave-one-out rate with	Leave-one-out rate with	Leave-one-out rate with	Leave-one-out rate with
	1-NN	LibSVM	1-NN	LibSVM	1-NN	LibSVM
8	0.916	0.627	0.963	0.972	0.930	0.954
10	0.916	0.627	0.963	0.972	0.938	0.954
12	0.916	0.627	0.963	0.972	0.935	0.953
14	0.916	0.627	0.972	0.972	0.940	0.958
16	0.916	0.627	0.956	0.974	0.944	0.960
18	0.916	0.627	0.956	0.970	0.938	0.960
20	0.916	0.627	0.953	0.965	0.938	0.958
22	0.916	0.627	0.946	0.963	0.930	0.960
24	0.916	0.627	0.951	0.944	0.942	0.944
26	0.916	0.627	0.951	0.938	0.944	0.940

从以上表中看出,采用无监督特征提取方法 UOSDV 降维的数据留一法交叉验证所得分类准确率均低于采用有监督特征提取方法 OSDV 降维的数据,而却不低于同属于无监督特征提取方法的 PCA 降维数据.

4 结束语

本文针对无监督特征降维问题,提出了一种无监督模式下最佳鉴别矢量集方法.实验表明,尽管由于样本类别信息的缺失,使得该方法无法优于有监督特征提取的最佳鉴别矢量集技术.但是不论是在聚类有效性还是分类准确率实验中均优于同属于无监督特征提取算法的主成分分析方法.该算法的局限性在于只适用于线性可分的数据集,对于线性不可分的数据集,如何在核化空间运用该方法将是下一步研究的重点.

附录:

模糊 Fisher 准则(fuzzy Fisher criterion)函数为

$$J_{\text{ffc}} = \frac{\tilde{S}_{fb}}{\tilde{S}_{fw}} = \frac{\omega^T S_{fb} \omega}{\omega^T S_{fw} \omega}, \quad (\text{A1})$$

使用 Lagrange 乘子法求解 J_{ffc} 取得极大值时,取 ω , m_i 和 u_{ij} 的值.

定义 Lagrange 函数为

$$L = \omega^T S_{fb} \omega - \lambda \omega^T S_{fw} \omega + \sum_{j=1}^N \lambda_j \left(\sum_{i=1}^c u_{ij} - 1 \right). \quad (\text{A2})$$

式中: λ 和 $\lambda_j (j = 1, 2, \dots, n)$ 为 Lagrange 乘子.

将 L 分别对 ω, m_i 及 u_{ij} 求偏导数,并令偏导数为零,即

$$\frac{\partial L}{\partial \omega} = 0, \quad (\text{A3})$$

$$\frac{\partial L}{\partial m_i} = 0, \quad (\text{A4})$$

$$\frac{\partial L}{\partial u_{ij}} = 0. \quad (\text{A5})$$

对于式(A3),由于 S_{fb} 和 S_{fw} 均为对称半正定矩阵,因此有

$$\frac{\partial (\omega^T S_{fb} \omega)}{\partial \omega} = 2 S_{fb} \omega,$$

$$\frac{\partial (\omega^T S_{fw} \omega)}{\partial \omega} = 2 S_{fw} \omega.$$

由此, $\frac{\partial L}{\partial \omega} = 0 \Rightarrow 2 S_{fb} \omega - 2 \lambda S_{fw} \omega = 0$, 因此有

$$S_{fb} \omega = \lambda S_{fw} \omega.$$

对于式(A4),由于

$$\frac{\partial S_{fb}}{\partial m_i} = 2 \sum_{j=1}^N u_{ij}^m (m_i - \bar{x}),$$

$$\frac{\partial S_{fw}}{\partial m_i} = -2 \sum_{j=1}^N u_{ij}^m (x_j - \bar{m}_i).$$

因此

$$\frac{\partial L}{\partial m_i} = 0 \Rightarrow 2 \omega^T \sum_{j=1}^N u_{ij}^m (m_i - \bar{x}) \omega +$$

$$2 \lambda \omega^T \sum_{j=1}^N u_{ij}^m (x_j - \bar{m}_i) \omega = 0$$

$$\Rightarrow \omega^T \sum_{j=1}^N u_{ij}^m (m_i - \bar{x}) \omega = -\lambda \omega^T \sum_{j=1}^N u_{ij}^m (x_j - \bar{m}_i) \omega$$

$$\Rightarrow \sum_{j=1}^N u_{ij}^m (\mathbf{m}_i - \bar{\mathbf{x}}) = -\lambda \sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \mathbf{m}_i)$$

$$\Rightarrow \mathbf{m}_i = \frac{\sum_{j=1}^N u_{ij}^m (\mathbf{x}_j - \frac{1}{\lambda} \bar{\mathbf{x}})}{\sum_{j=1}^N u_{ij}^m (1 - \frac{1}{\lambda})}$$

对于式(A5), 由于

$$\frac{\partial S_{fb}}{\partial u_{ij}} = m u_{ij}^{m-1} (\mathbf{m}_i - \bar{\mathbf{x}}) (\mathbf{m}_i - \bar{\mathbf{x}})^T,$$

$$\frac{\partial S_{fw}}{\partial u_{ij}} = m u_{ij}^{m-1} (\mathbf{x}_j - \mathbf{m}_i) (\mathbf{x}_j - \mathbf{m}_i)^T.$$

因此,有

$$\frac{\partial L}{\partial u_{ij}} = 0$$

$$\Rightarrow \omega^T m u_{ij}^{m-1} (\mathbf{m}_i - \bar{\mathbf{x}}) (\mathbf{m}_i - \bar{\mathbf{x}})^T \omega - \lambda \omega^T m u_{ij}^{m-1} (\mathbf{x}_j - \mathbf{m}_i) (\mathbf{x}_j - \mathbf{m}_i)^T \omega + \lambda_j = 0$$

$$\Rightarrow u_{ij} = \left(\frac{\lambda_j}{m (\lambda \omega^T (\mathbf{x}_j - \mathbf{m}_i) (\mathbf{x}_j - \mathbf{m}_i)^T \omega - \omega^T (\mathbf{m}_i - \bar{\mathbf{x}}) (\mathbf{m}_i - \bar{\mathbf{x}})^T \omega)} \right)^{\frac{1}{m-1}}. \quad (\text{A6})$$

又 $\sum_{k=1}^c u_{kj} = 1, j = 1, 2, \dots, N$, 所以有

$$1 = \sum_{k=1}^c u_{kj} = \sum_{k=1}^c \left(\frac{\lambda_j}{m (\lambda \omega^T (\mathbf{x}_j - \mathbf{m}_k) (\mathbf{x}_j - \mathbf{m}_k)^T \omega - \omega^T (\mathbf{m}_k - \bar{\mathbf{x}}) (\mathbf{m}_k - \bar{\mathbf{x}})^T \omega)} \right)^{\frac{1}{m-1}}. \quad (\text{A7})$$

式(A6)和(A7)相除,得

$$u_{ij} = \frac{(\omega^T (\mathbf{x}_j - \mathbf{m}_i) (\mathbf{x}_j - \mathbf{m}_i)^T \omega - \frac{1}{\lambda} \omega^T (\mathbf{m}_i - \bar{\mathbf{x}}) (\mathbf{m}_i - \bar{\mathbf{x}})^T \omega)^{-\frac{1}{m-1}}}{\sum_{k=1}^c (\omega^T (\mathbf{x}_j - \mathbf{m}_k) (\mathbf{x}_j - \mathbf{m}_k)^T \omega - \frac{1}{\lambda} \omega^T (\mathbf{m}_k - \bar{\mathbf{x}}) (\mathbf{m}_k - \bar{\mathbf{x}})^T \omega)^{-\frac{1}{m-1}}}.$$

参考文献:

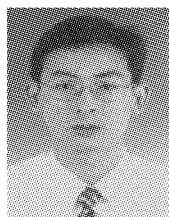
- [1] FISHER R A. The use of multiple measurements in taxonomic problems [J]. *Annals of Eugenics*, 1936, 7: 179-188.
- [2] SAMMON J W. An optimal discriminant plane [J]. *IEEE Trans on Computers*, 1970, 19(9): 826-829.
- [3] FOLEY D H, SAMMON J W. An optimal set of discriminant vectors [J]. *IEEE Trans on Computers*, 1975, 24(3): 281-289.
- [4] DUCHENE J, LECLERCQ S. An optimal transformation for discriminant and principal component analysis [J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 1988, 10(6): 978-983.
- [5] 金忠, 杨静宇, 陆建峰. 一种具有统计不相关性的最佳鉴别矢量集 [J]. *计算机学报*, 1999, 22(10): 1105-1108.
JIN Zhong, YANG Jingyu, LU Jianfeng. An optimal set of uncorrelated discriminant features [J]. *Chinese J Computers*, 1999, 22(10): 1105-1108.
- [6] JIN Zhong, YANG Jingyu, HU Zhongshan, et al. Face recognition based on the uncorrelated discriminant transformation [J]. *Pattern Recognition*, 2001, 34(7): 1405-1416.
- [7] 束婷婷, 甘岚, 杨静宇. 求解统计不相关的最佳鉴别矢量的统一算法 [J]. *南京理工大学学报: 自然科学版*, 2002, 26(3): 290-294.
- [8] 吴小俊, 杨静宇, 王士同, 等. 改进的统计不相关最优鉴别矢量集 [J]. *电子与信息学报*, 2005, 27(1): 47-50.
WU Xiaojun, YANG Jingyu, WANG Shitong, et al. An improved optimal set of statistical uncorrelated discriminant vectors [J]. *Journal of Electronics & Information Technology*, 2005, 27(1): 47-50.
- [9] ZHENG Wenming. A note on kernel uncorrelated discriminant analysis [J]. *Pattern Recognition*, 2005, 38(11): 2185-2187.
- [10] LIANG Yixiong, LI Chengrong, GONG Weiguo, et al. Uncorrelated linear discriminant analysis based on weighted pairwise Fisher criterion [J]. *Pattern Recognition*, 2007, 40(12): 3606-3615.
- [11] WU Xiaojun, LU Jieping, YANG Jingyu, et al. An extreme case of the generalized optimal discriminant transformation and its application to face recognition [J]. *Neurocomputing*, 2007, 70(4): 828-834.
- [12] DY J G, BRODLEY C E. Feature selection for supervised learning [J]. *Journal of Machine Learning Research*,

2004, 5: 845-889.

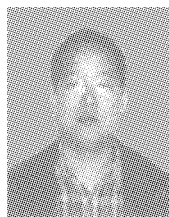
- [13] MIRKIN B. Concept learning and feature selection based on square-error clustering[J]. Machine Learning, 1999, 35(1): 25-39.
- [14] 刘涛, 吴功宜, 陈正. 一种高效的用于文本聚类的无监督特征选择算法[J]. 计算机研究与发展, 2005, 42(3): 381-386.
- LIU Tao, WU Gongyi, CHEN Zheng. An effective unsupervised feature selection method for text clustering[J]. Journal of Computer Research and Development, 2005, 42(3): 381-386.
- [15] WU Kuolung, YU Jian, YANG Minshen. A novel fuzzy clustering algorithm based on a fuzzy scatter matrix with optimality tests[J]. Pattern Recognition Letters, 2005, 26(5): 639-652.
- [16] BLAKE C L, MERZ C J. UCI repository of machine learning databases[EB/OL]. (2008-03-24) [2008-04-20]. <http://www.ics.uci.edu/~mlearn/MLRepository.html>.
- [17] RAND W M. Objective criteria for the evaluation of clustering methods[J]. Journal of the American Statistical Association, 1971, 66(336): 846-850.
- [18] CHANG Chihchung, LIN Chihjen. LIBSVM: a library for

support vector machines[EB/OL]. [2008-04-15]. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.

作者简介:



曹苏群, 男, 1976年生, 博士研究生, 主要研究方向为模式识别、图像处理.



王士同, 男, 1964年生, 教授, 博士生导师, 主要研究方向为模糊人工智能、模式识别、图像处理和生物信息学等, 先后十多次留学英国、日本和香港地区, 发表学术论文数十篇.



陈晓峰, 男, 1977年生, 博士研究生, 主要研究方向为机器学习、模式识别.

第2届全国智能信息处理学术会议

The 2nd National Conference on Intelligent Information Processing

由中国人工智能学会知识工程与分布式智能专业委员会与中国计算机学会人工智能与模式识别专业委员会联合主办, 中国矿业大学与中国科学院智能信息处理重点实验室联合承办的“第二届全国智能信息处理学术会议”(The 2nd National Conference on Intelligent Information Processing, NCIIP2009)定于2009年5月22-24日在江苏省徐州市召开.

征稿范围(包括但不限于)

- | | | | |
|-----------|--------------|-------------|-----------|
| ● 机器学习 | ● 机器挖掘 | ● 语义计算 | ● 服务计算 |
| ● 多媒体信息检索 | ● 多 Agent 系统 | ● 信息粒度计算 | ● 神经信息处理 |
| ● 模糊信息处理 | ● 粗糙集信息处理 | ● 知识获取与知识发现 | ● 生物信息处理 |
| ● 智能计算 | ● 模式识别 | ● 自然语言理解 | ● 图像处理与理解 |
| ● 语音识别 | ● 智能决策 | ● 智能控制 | ● 智能系统 |
| ● 其他 | | | |

重要日期

投稿截止: 2009年3月1日

录用通知: 2009年4月15日

最终稿: 2009年4月30日

会议网址: <http://nciip2009.cumt.edu.cn>

电子邮件: nciip2009@cumt.edu.cn