

基于矩阵模式的局部子域最大间距判别分析

黄丽莉^{1,2}

(1. 安徽理工大学 电气与信息工程学院, 安徽 淮南 232001; 2. 盐城工学院 信息学院, 江苏 盐城 224051)

摘要:矩阵模式的 Fisher 线性判别准则(MatFLDA)作为近几年矩阵模式下的经典特征提取方法被广泛地加以研究和运用。然而 MatFLDA 方法作为全局判别准则一定程度上忽视了样本空间内在的局部结构和局部信息。在矩阵模式下,引入局部加权均值(LWM)并结合最大间距判别分析(MMC),提出一种具有局部学习能力的有监督的特征提取方法:基于矩阵模式的局部子域最大间距判别分析(Mat-LSMMC),提高了 MatFLDA 方法的局部学习能力,具有较强的特征提取能力。通过测试人造、真实数据集来表明所提方法的优势。

关键词:矩阵模式的 Fisher 线性判别准则;局部加权均值;最大间距判别分析

中图分类号:TP391 **文献标识码:**A **文章编号:**1671-5322(2014)01-0026-05

智能识别领域存在着大量需要处理的高维数据样本,为了便于从中找出我们最需要的信息,就需要进行特征降维,将一些冗余的信息进行处理。特征提取作为一种数据预处理方法得到了广泛的运用^[1]。一般经典的特征提取方法,如:PCA^[2,3], LDA^[4,5]方法都是在高维向量空间上进行特征提取。然而,在现实的空间中,将图像矩阵样本转换为向量的过程中,破坏了图像矩阵样本原有的几何结构,导致一些有用信息的损失。

近年来基于矩阵模式的特征提取算法(如文献[6-8])由于能有效地提高识别率、节省计算时间,被广泛的研究。如文献[6]提出的矩阵模式的 Fisher 线性判别准则(MatFLDA)、矩阵模式的主成份分析方法(MatPCA)不但在处理图像数据时使用矩阵模式,尽可能地保留原始的数据信息,而且,提出一种将矢量数据转换成矩阵模式的方法,大大扩展了矩阵模式的应用范围。

矩阵模式的 Fisher 线性判别准则(MatFLDA)虽然避免了基于向量方法的缺点,具有较好的特征提取能力,然而,MatFLDA 方法作为全局判别准则,算法中使用总体均值来替代方差,对于那些具有明显局部流形的数据,比如人脸数据、高维文

本数据等等,MatFLDA 方法一定程度上缺乏适应性。为了确保样本数据的局部信息,我们在本文的方法构造中引入局部保持投影(Locality Preserving Projections, LPP)^[9]流形学习方法中通过嵌入样本的最近邻图,提取样本空间的局部结构信息这一思想。

本文在 MatFLDA 基础上,引入局部加权均值(Local Weighted Mean, LWM)^[10]的概念,提出一种具有局部学习能力的特征提取方法:基于矩阵模式的局部子域最大间距判别分析(Local Sub-domains Maximum Margin Criterion Based Matrix Pattern, Mat-LSMMC)。该方法具有如下优势:(1)Mat-LSMMC 方法运用了矩阵模式,提高了特征提取能力,大大降低了空间复杂度,同时消除了一般行(列)向量所带来的限制;(2)Mat-LSMMC 方法使用了最大间距准则(Maximum Margin Criterion, MMC)^[11,12],避免小样本问题^[13];(3)Mat-LSMMC 方法中用 LWM 替代标准均值,一定程度上较好地反映不同样本在保持内在局部结构方面的贡献程度大小,实现了局部保持。

在实验中,通过测试传统的 UCI 数据集和人

收稿日期:2013-11-20

基金项目:国家自然科学基金(61272210);江苏省自然科学基金(BK20111417);苏州大学江苏省计算机信息处理技术重点实验室开放课题(KJS1126);江苏省新型环保重点实验室开放课题(AE201068)

作者简介:黄丽莉(1989-),女,江苏苏州人,硕士生,主要研究方向为模式识别、图像处理。

脸识别数据集,表明本文的 Mat - LSMMC 方法具有上述优势。

1 相关工作

为了便于描述,我们在本节简单回顾一下 LWM、MatFLDA 方法的概念。

1.1 局部加权均值:LWM

定义 1 (LWM) 假设 $X_{lq} = \{x_{lq}^i\}_{i=1}^{n_l}$ 表示一个局部子域,那么局部子域 X_{lq} LWM 可以写为:

$$\sum_{i=1}^{n_l} \frac{\beta_{\varphi} x_{lq}^i}{\sum_{p=1}^{n_l} \beta_{\varphi}}, \text{ 其中 } 0 \leq \beta_{\varphi} \leq 1 \text{ 且 } \beta_{\varphi} = \exp\left(-\frac{\|x_{lq} - x_{lq}^{(i)}\|^2}{h}\right)$$

是只与局部子域 X_{lq} 上样本相关的权值参数, h^1 是热核函数 $\exp\left(-\frac{d^2}{h}\right)$ 上的热核参数。

从定义 1 可以看出, LWM 是通过给在局部子域中的样本分配不同权值来表示不同样本在保持内在局部结构所反映的贡献程度的不同。

1.2 矩阵线性判别分析 MatFLDA

定义 2 (MatFLDA): 假设有 n 个样本组成的样本集 $A = \{A_1, \dots, A_n\}$, $A_i (i = 1, 2, \dots, n)$ 均是矩阵模式, 分别属于 C 个不同的类, 第 c 类大小为 n_c , 给定分类决策平面的法向量 ω 。则 MatFLDA 的类内、类间散度矩阵和目标函数为:

$$S_W^{\text{Mat}} = \sum_{c=1}^C \sum_{i=1}^{n_c} (\bar{A}_{ci} - \bar{A}_c)(\bar{A}_{ci} - \bar{A}_c)^T$$

$$S_B^{\text{Mat}} = \sum_{c=1}^C n_c (\bar{A}_c - \bar{A})(\bar{A}_c - \bar{A})^T$$

$$\arg\max_{\omega^T \omega = 1} J(\omega) = \arg\max_{\omega^T \omega = 1} \frac{\omega^T S_B^{\text{Mat}} \omega}{\omega^T S_W^{\text{Mat}} \omega}$$

其中: 样本总的平均值为 $\bar{A} = \frac{1}{n} \sum_{i=1}^n A_i$, 第 c 类样本的平均值为 $\bar{A}_c = \frac{1}{n_c} \sum_{i=1}^{n_c} A_{ci}$, $c = 1, 2, \dots, C, i = 1, 2, \dots, n_c$

2 基于矩阵模式的局部子域最大间距判别分析: Mat - LSMMC

在矩阵模式的基础上, 结合局部加权子域的概念, 本文提出了一种新的算法 Mat - LSMMC, 见图 1。我们用文献[6]的方法将每个矢量数据转换为矩阵模式, 在样本数据中结合 1.1 节中局部加权均值 LWM 的概念, 运用流形学习算法中的局部保存思想将数据有效划分为各个局部加权子域, 寻找每个局部加权子域的内最近邻局部子

域 (Within - class Nearest Neighbor Local Sub - domains, WNNLSD) 和类间最近邻局部子域 (Between - class Nearest Neighbor Local Sub - domains, BNNLSD), 构造对应的具有明显几何含义的矩阵模式类内散度和类间散度。

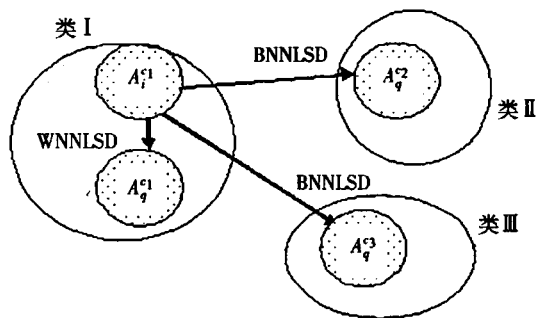


图 1 Mat - LSMMC 原理图

Fig. 1 The schematic diagram of Mat - LSMMC

下面定义本文 Mat - LSMMC 方法的类内散度矩阵和类间散度矩阵。

定义 3 假设有 n 个样本组成的样本集 $A = \{A_1, \dots, A_n\}$, $A_i (i = 1, 2, \dots, n)$ 均是矩阵模式, 分别属于 C 个不同的类, 其中大小为 n_c 样本子集 A^c 属于第 c 类, 且 $\forall A_i^c (i = 1, \dots, n_c)$ 为 A^c 的任意一个局部子域, 给定分类决策平面的法向量 ω , 则 Mat - LSMMC 方法对应的矩阵模式类内散度和类间散度分别为:

$$\alpha_W^{\text{Mat}} =$$

$$\sum_{c=1}^C \sum_{i=1}^{n_c} \sum_{j=1}^{n_c} r_{ij}^{\omega_c} \| \text{LWM}(\omega^T A_i^c) - \text{LWM}(\omega^T A_j^c) \|_F^2$$

(1)

$$\alpha_B^{\text{Mat}} =$$

$$\sum_{c_1=1}^C \sum_{c_2=1}^C \sum_{i=1}^{n_{c_1}} \sum_{j=1}^{n_{c_2}} r_{ij}^{b_{c_1 c_2}} \| \text{LWM}(\omega^T A_i^{c_1}) - \text{LWM}(\omega^T A_j^{c_2}) \|_F^2$$

(2)

其中: $r_{ij}^{\omega_c} = \begin{cases} 1 & A_i^c \text{ is the WNNLSD of the } A_j^c \\ 0 & \text{otherwise} \end{cases}$,

$r_{ij}^{b_{c_1 c_2}} = \begin{cases} 1 & A_i^{c_1} \text{ is the BNNLSD of the } A_j^{c_2} \\ 0 & \text{otherwise} \end{cases}$ 。

$\| \cdot \|_F$ 表示范数, 例如: $\| \cdot \|_F = \sqrt{\sum_i \sum_j a_{ij}^2}$ 。

则本文 Mat - LSMMC 方法的目标函数为:

$$\arg\max_{\omega^T \omega = 1} J(\omega) = (1 - \gamma) \alpha_B^{\text{Mat}} - \gamma \alpha_W^{\text{Mat}} \quad (3)$$

为了提高本文 Mat - LSMMC 方法求解目标函数的效率, 我们对式(1)、式(2)进行改写和简化, 由此, 我们得到如下定理。

定理 1 根据定义 3, 式(1)、式(2)分别可以简化为:

$$\alpha_{\mathbf{w}}^{\text{Mat}} = \text{tr}(\omega^T L_{\mathbf{w}} \omega) \quad (4)$$

$$\alpha_{\mathbf{B}}^{\text{Mat}} = \text{tr}(\omega^T L_{\mathbf{B}} \omega) \quad (5)$$

我们首先证明等式(4)中的 $\alpha_{\mathbf{w}}^{\text{Mat}}$ 。

根据定义 1, 等式(1)中的 $\text{LWM}(\omega^T A_i^c)$ 、 $\text{LWM}(\omega^T A_j^c)$ 分别可以表示为: $\beta_i^c A_i^{cT} \omega$ 、 $\beta_j^c A_j^{cT} \omega$, β_i^c 、 β_j^c 分别表示两个局部子域对应的权值。根据 $\|A\|_F^2 = \text{tr}(A^T A)$, 有:

$$\|\text{LWM}(\omega^T A_i^c) - \text{LWM}(\omega^T A_j^c)\|_F^2 =$$

$$\|\beta_i^c A_i^{cT} \omega - \beta_j^c A_j^{cT} \omega\|_F^2 = \text{tr}(\omega^T L_c^{\tilde{y}} \omega)$$

其中:

$$L_c^{\tilde{y}} = A_i^c \beta_i^c \beta_i^{cT} A_i^{cT} - A_i^c \beta_i^c \beta_j^{cT} A_j^{cT} - A_j^c \beta_j^c \beta_i^{cT} A_i^{cT} + A_j^c \beta_j^c \beta_j^{cT} A_j^{cT}$$

则式(1)可以表示为:

$$\alpha_{\mathbf{w}} = \sum_{c=1}^C \sum_{i=1}^{n_c} \sum_{j=1}^{n_c} r_{ij}^{\omega_c} \|\text{LWM}(\omega^T A_i^c) - \text{LWM}(\omega^T A_j^c)\|_F^2 =$$

$$\text{tr}(\omega^T (\sum_{c=1}^C \sum_{i=1}^{n_c} \sum_{j=1}^{n_c} r_{ij}^{\omega_c} L_c^{\tilde{y}}) \omega) = \text{tr}(\omega^T L_{\mathbf{w}} \omega)$$

则式(4)成立。

同理可以证明式(5)成立, 其中 $L_{\mathbf{B}} = \sum_{c_1=1}^C \sum_{c_2=1}^C$
 $\sum_{i=1}^{n_{c_1}} \sum_{j=1}^{n_{c_2}}, L_{c_1 c_2}^{\tilde{y}} = A_i^{c_1} \beta_i^{c_1} \beta_i^{c_1T} A_i^{c_1T} - A_i^{c_1} \beta_i^{c_1} \beta_j^{c_2T} A_j^{c_2T} -$
 $A_j^{c_2} \beta_j^{c_2} \beta_i^{c_1T} A_i^{c_1T} + A_j^{c_2} \beta_j^{c_2} \beta_j^{c_2T} A_j^{c_2T}$ 。定理得证。

算法: 基于矩阵模式的局部子域最大间距判别分析。

Algorithm: Mat - LSMMC

Input: 数据集 X , 并给 k_c, h 赋初值;

Output: 变换矢量 ω ;

Step1: 对矢量表示的样本根据文献[6]的方法转换成矩阵模式;

Step2: 根据给定的 k_c , 将数据集 A 划分成若干个局部子域;

Step3: 寻找每一个局部子域 $A_i^{c_1}$ 对应的类内最近邻局部子域 $A_q^{c_1}$ 和类间最近邻局部子域 $A_q^{c_2}$;

Step4: 根据定义 3, 构造相应的类内散度矩阵和类间散度矩阵;

Step5: 根据定理 1, 求解式(3), 得到满足式(3)最大特征值对应的矢量 ω 。

3 实验

通过测试 4 个 UCI 数据子集(即: IRIS、Ionosphere、Wine、Glass)、两个人脸识别数据集(ORL、Yale)来说明 Mat - LSMMC 方法具有较好的特征提取能力。在实验过程中都使用最近邻分类器。

3.1 测试 UCI 数据集

通过测试 4 个 UCI 数据子集(见表 1)来说明 Mat - LSMMC 方法具有较好的特征提取能力, 并将 Mat - LSMMC 方法同 MatPCA、MatFLDA、MatMMC 3 种特征提取方法进行比较。测试过程中 Mat - LSMMC 的参数: $K \in [2, 5]$, $h \in 2^t$, $t \in (-500, 500)$, $\gamma \in (0, 1)$, 使用 10 - 折交叉验证。测试结果见表 2。

表 1 Ionosphere、IRIS、Wine、Glass 数据集
Table 1 Ionosphere IRIS, Wine, Glass, datasets

Datasets	The Number of Training Samples	The Number of Test Samples	The Number of Features	The Number of Subjects	Matrix Model
Ionosphere	230	117	33	2	11 × 3
IRIS	100	50	4	3	2 × 2
Wine	118	60	12	3	10 × 6
Glass	142	72	9	6	3 × 3

从表 2, 我们可以得到以下结论:

(1) Mat - LSMMC 同其他 3 种方法相比, 其精度较大。说明 Mat - LSMMC 方法具有较好的特征提取。Mat - LSMMC 方法中运用了局部流形学习方法中构造近邻拉普拉斯图的方法, 实现了局部保持, 提高了特征提取效果和分类精度。

(2) Mat - LSMMC 由于是基于矩阵模式下的特征提取方法, 空间复杂度较低。

(3) Mat - LSMMC 方法由于使用了最大间距准则, 避免了 matFLDA 方法的小样本问题。

3.2 测试人脸数据

人脸图像数据集蕴含着内在的非线性流形结构。通过测试 ORL(32 × 32)(40 个类, 每类 10 个样本)和 Yale(32 × 32)(15 个类, 每类 11 个样本)两个人脸图像数据集来表明 Mat - LSMMC 方法在一定程度上能解决小样本问题, 具有较好的

特征提取能力。MatMMC 4 种方法来对比测试,Mat - LSMMC 参数的设定同 3.1。测试结果见表 3。

在测试过程中,为了有效说明 Mat - LSMMC 方法的优势,我们使用 PCA、MatPCA、MatFLDA、

表 2 对 Ionosphere、IRIS、Wine、Glass 数据集的识别效果比较

Table 2 Reognition performance comprison on Ionosphere IRIS, Wine, Glass datasets

Datasets	Algorithm	MatPCA	MatFLDA	MatMMC	Mat - LSMMC
Ionosphere	Accuracy	0.906 0	0.863 2	0.897 4	0.9316
					($K=3;h=2^5$; $\gamma=0.3$)
IRIS	Accuracy	0.94	0.94	0.96	1
					($K=5;h=2^{-3}$; $\gamma=0.5$)
Wine	Accuracy	0.716 7	0.733 3	0.766 7	0.8
					($K=5;h=2-10$; $\gamma=0.25$)
Glass	Accuracy	0.833 33	0.805 6	0.847 2	0.875
					($K=3;h=2-5$; $\gamma=0.1$)

表 3 对 ORL、Yale 数据集识别效果比较

Table 3 The comparison of ORL, Yale data set recognition effect

Datasets	ORL		Yale	
Number of Train Samples	4	6	4	6
Alorightm	Accuracy	Accuracy	Accuracy	Accuracy
PCA	0.854 17	0.887 5	0.638 1	0.666 67
MatPCA	0.875	0.9	0.6667	0.7067
MatFLDA	0.908 3	0.9	0.676 2	0.773 3
MatMMC	0.912 5	0.931 3	0.695 2	0.786 7
	0.9208	0.9437	0.7048	0.8133
Mat - LSMMC	($K=5;h=2300$; $\gamma=0.25$)	($K=3;h=2500$; $\gamma=0.875$)	($K=5;h=2200$; $\gamma=0.1$)	($K=5;h=2200$; $\gamma=0.875$)

为了提高测试结果的有效性,在测试过程中分别对 ORL、Yale 数据集选取 4 个、6 个样本作为训练样本来测试上述 4 种方法,并分别使用 4 - 折交叉验证和 6 - 折交叉验证。

根据上述结果,我们可以得到如下结论:

- (1)Mat - LSMMC 方法在处理人脸数据时,比其他方法具有更好的特征提取效果和分类精度,说明此方法由于引入了 LWM 来替代标准均值,较好地反映了不同样本在保持样本内在局部结构方面的贡献程度的差异,更好地保持样本内在蕴含的几何结构。
- (2)Mat - LSMMC 方法处理 32 × 32 图像数据集时,直接使用矩阵模式,如果用矢量形式表示,则需要存储 1 024 × 1 024 的散度矩阵,而本文的方法则只需要存储 32 × 32 散度矩阵,所占空间只是原方法的 1/1 024,大大降低了空间复杂度。

(3)Mat - LSMMC 方法使用了最大间距判别分析,不存在小样本问题。

4 总结

本文在矩阵模式的基础上,提出了具有一定局部学习能力的 Mat - LSMMC 方法。该方法结合局部加权均值概念,使用流形学习理论对原样本空间进行有效划分,一定程度上提高了算法的局部学习能力。同时,由于采用矩阵模式使得算法空间复杂度大大的降低。最后通过测试人造、真实数据集来表明本文 Mat - LSMMC 方法具有较强的特征提取能力。当然,Mat - LSMMC 方法也存在着一定的不足,比如如何更好地确定近邻参数和热核参数,更有效地提高该方法的执行效率,将是我们以后研究的方向。

参考文献:

- [1] Bian Z Q, Zhang X G. Pattern Recognition[M]. Beijing: TsingHua University Press, 2001.
- [2] Jolliffe I T. Principal Component Analysis[M]. New York: Springer – Verlag, 1986.
- [3] Todorov, Valentin, Filzmoser, et al. Comparing classical and robust sparse PCA[C]. Advances in Intelligent Systems and Computing. Berlin:Springer Verlag, 2013:1 283 – 1 291.
- [4] Li Ronghua, Liang Shuang, Baci George, et al. Equivalence between LDA/QR and direct LDA[J]. International Journal of Cognitive Informatics and Natural Intelligence, 2011, 5(1): 94 – 112.
- [5] Dhir Chandra Shekhar, Lee Sooyoung. Discriminant independent component analysis[J]. IEEE Transactions on Neural Networks, 2011, 22(6): 845 – 857.
- [6] Chen S C, Zhu Y L, Zhang D Q, et al. Feature extraction approaches based on matrix pattern: matPCA and matFLDA[J]. Pattern Recognition Letters, 2005, 26(8): 1 157 – 1 167.
- [7] He Xiaofei, Cai Deng, Niyogi Partha. Tensor Subspace Analysis[C]. Advances in Neural Information Processing Systems. 2005 Annual Conference on Neural Information Processing Systems. Canada: Neural information processing system foundation, 2005: 499 – 506.
- [8] Jun Gao, Chung Fulai, Wang Shitong. Matrix pattern based minimum within – class scatter support vector machines. Applied Soft Computing Journal, 2011, 11(8): 5 602 – 5 610.
- [9] Wong W K, Zhao H T. Supervised optimal locality preserving projection[J]. Pattern Recognition, 2012, 45(1): 186 – 197.
- [10] Atkeson, Christopher G, Moore, et al. Locally weighted learning[J]. Artificial Intelligence Review, 1997, 11(1 – 5): 75 – 113.
- [11] Yang Wankou, Sun Changyin, Du Helen S, et al. Feature extraction using Laplacian Maximum Margin Criterion[J]. Neural Processing Letters, 2011, 33(1): 99 – 110.
- [12] Cui Yan, Fan Liya. Feature extraction using fuzzy maximum margin criterion[J]. Neurocomputing, 2012, 86(1): 52 – 58.
- [13] Deng Weihong, Liu Yebin, Hu Jiani, et al. The small sample size problem of ICA: A comparative study and analysis[J]. Pattern Recognition, 2012, 45(12): 4 438 – 4 450.

Local Sub – domains Maximum Margin Criterion Based on Matrix Pattern

Huang Lili^{1,2}

(1. School of Electrical and Information Engineering, Anhui University of Science and Technology, Huainan Anhui 232001, China;)
(2. School of Information Engineering, Yancheng institute of Technology, Yancheng Jiangsu 224051, China)

Abstract: MatFLDA as a classic feature extraction method in recent years is widely studied and used. However, MatFLDA as a global criterion is neglected to some extent sample space inner local structure and local information. Therefore, based on the matrix pattern, by introducing the Local Weighted Mean (LWM) and combined with Maximum Margin Criterion (MMC), we put forward a certain local learning ability of supervised feature extraction method: Local Sub – domains Maximum Margin Criterion Based Matrix Pattern, Mat – LSMMC), the discrimination analysis with the strong ability of feature extraction can improve the local learning ability of MatFLDA method. Finally, the test on artificial and real datasets shows the above mentioned advantages of the Mat – LSMMC method.

Keywords: MatFLDA; Local Weighted Mean; Maximum Margin Criterion

(责任编辑: 张英健)