

基于语境距离度量的拉普拉斯最大间距判别准则

皋 军^{1,2} 王士同¹ 王晓明³

摘要 线性拉普拉斯判别准则 (Linear Laplacian discrimination, LLD) 作为一种非线性特征提取方法得到了较为成功的运用. 然而通过分析得知在具体使用 LLD 方法的过程中还会面临小样本以及如何确定原始样本空间类型的问题. 因此, 本文引入语境距离度量并结合最大间距判别准则的基本原理提出一种基于语境距离度量的拉普拉斯最大间距判别准则 (Contextual-distance metric based Laplacian maximum margin criterion, CLMMC). 该准则不但在一定程度上避免小样本问题, 而且由于语境距离度量更关注输入样本簇内在的本质结构而不是原始样本空间的类型, 从而降低了该准则对特定样本空间的依赖程度. 同时通过引入计算语境距离度量的新算法并结合 QR 分解的基本原理, 使得 CLMMC 在处理高维矢量模式数据时更具适应性和效率. 并从理论上讨论 CLMMC 准则具有的基本性质以及与 LLD 准则的内在联系. 实验证明 CLMMC 准则具有上述优势.

关键词 线性拉普拉斯判别准则, 语境距离度量, 最大间距判别准则, QR 分解

DOI 10.3724/SP.J.1004.2010.01661

Contextual-distance Metric Based Laplacian Maximum Margin Criterion

GAO Jun^{1,2} WANG Shi-Tong¹ WANG Xiao-Ming³

Abstract Linear Laplacian discrimination (LLD) as a non-linear feature extraction method has obtained very extensive applications. However, LLD suffers from the small sample size problem (SSS) and/or the type of the sample space when it is used. In order to circumvent such shortcomings, a contextual-distance metric based Laplacian maximum margin criterion (CLMMC) is proposed in this paper by using contextual-distance metric and integrating maximum margin criterion (MMC) into the LLD. The proposed criterion can obviously decrease the dependence on the sample space since the contextual-distance metric focuses more on intrinsic structure of a cluster of samples than on its type. And it is of higher adaptability and efficiency to use the new algorithm to compute contextual-distance metric and applying QR-decomposition when high-dimensional vector data are dealt with. The basic properties of CLMMC and its relation to LLD are also discussed. The experimental results indicate the above advantages of the CLMMC.

Key words Linear Laplacian discrimination (LLD), contextual-distance metric, maximum margin criterion (MMC), QR-decomposition

特征提取作为一种有效的降维方法在模式识别、计算机视觉和数据挖掘中得到了成功的运用. 所谓特征提取就是将高维特征空间通过变换转化为相应的低维表示空间^[1], 而该低维空间要尽可能地保持原有空间的判别信息. 目前主成分分析 (Principal component analysis, PCA)^[2-5] 和线性判别分析 (Linear discrimination analysis, LDA)^[6-8] 作为两种比较经典的特征提取方法被广

泛地加以研究. PCA 方法作为一种无监督的方法主要是依据所构造的协方差矩阵去选取能保持样本全局信息的主元 (即一组相互正交的基底), 并通过这些主元实现原始样本空间的低维投影. 比如: 在人脸识别领域基于 PCA 基本原理的特征脸方法 (Eignfaces method)^[9-10] 就可以有效地解决高维空间低维嵌入的问题. 而 LDA 作为一种有监督的特征提取方法在充分使用类标号的前提下, 依据最大化类间离散度同时最小化类内离散度原则来得到原始高维特征空间的低维投影, 从而 LDA 方法与 PCA 方法相比更有利于模型化不同类别样本所具有的潜在结构和特征. 然而, LDA 方法在处理高维小样本数据时会导致类内离散度矩阵发生奇异, 即所谓的小样本问题 (Small sample size problem, SSS), 从而影响了 LDA 方法的特征提取能力. 为了解决这一问题, 提高 LDA 方法的泛化能力, 目前已有大量的解决方法, 比如: Fisher 脸方法 (Fisher faces method)^[11-12]、P-LDA (Perturbation LDA)^[13]、BDMSD (Biometric dispersion matcher with simultaneously diagonal-

收稿日期 2009-09-25 录用日期 2010-09-15
Manuscript received September 25, 2009; accepted September 15, 2010

国家自然科学基金 (60773206, 60903100, 90820002), 江苏省普通高校研究生科研创新计划 (CX09B-175Z), 盐城工学院应用基础研究项目 (XKY2009070) 资助

Supported by National Natural Science Foundation of China (60773206, 60903100, 90820002), Jiangsu Provincial Innovation Project of Graduate Students of China (CX09B-175Z), and Applied Basic Research Programs of Yancheng Institute of Technology (XKY2009070)

1. 江南大学物联网学院 无锡 214122 2. 盐城工学院信息工程学院 盐城 224001 3. 西华大学数学与计算机学院 成都 610039

1. School of Internet of Things, Jiangnan University, Wuxi 214122 2. School of Information Engineering, Yancheng Institute of Technology, Yancheng 224001 3. School of Mathematics and Computer Engineer, Xihua University, Chengdu 610039

ization)^[14]、PCA+LDA^[15]、2DFLD^[16]、DLA^[17]，特别是一些基于最大间距原则的方法：最大散度差判别准则^[18]、最大间距判别准则 (Maximum margin criterion, MMC)^[19] 和规范化的最大间距判别准则 (Regularized maximum margin criterion, RMMC)^[20] 等，该类方法则是通过将传统的 LDA 方法的比率形式转换成差的形式，这样就很容易解决小样本问题，同时与其他方法相比具有相对较低的时间复杂度^[21]。

相对于非线性特征提取方法，上述线性方法并不能发现样本空间蕴含的非线性结构。目前非线性方法主要存在两类：一类是基于核的方法，比如 Kernel PCA (KPCA)^[22-23]、局部非线性鉴别准则 (Localized kernel Fisher criterion, LKFC)^[24]、Kernel RMMC^[20] 等；另一类基于流形 (Manifold) 的学习方法，比如 Tenenbaum 等提出的等距特征映射 (Isometric feature mapping, isomap)^[25]、Roweis 等提出的局部线性嵌入 (Locally linear embedding, LLE)^[26]、Belkin 等提出的拉普拉斯特征映射 (Laplacian eigenmap, LE)^[27] 等。特别是 He 等提出的局部保持投影 (Locality preserving projections, LPP)^[28] 流形学习方法。LPP 可以看成 LE 的线性形式，该方法不但能有效保持样本内在的局部几何结构，还能在低维的表示空间提供显式的嵌入映射函数。不幸的是 LPP 方法同样面临着小样本问题，为此，在基于 LPP、MMC 基本原理的前提下，文献 [29-30] 分别提出了无监督的局部保持最大信息投影 (Locality-preserved maximum information projection, LPMIP) 和有监督的拉普拉斯双向最大间距判别 (Laplacian bidirectional maximum margin criterion, LBMMC)。

虽然以上各种子空间学习方法都表现出一定的有效性，但通过研究发现上述方法都是使用传统的 l_2 范数作为距离度量，从而在一定程度上假设原始样本空间是欧几里德空间 (即欧氏空间)。然而并不是所有的原始样本空间都是欧几里德空间，比如在计算机视觉中使用直方图表示的样本空间和非线性黎曼流形空间就属于非欧几里德空间。因此在非欧几里德空间还使用 l_2 范数作为距离度量显然是不合适的。为了在一定程度上降低距离度量对原始样本空间的依赖，Zhao 等结合 LE、LPP 和 LDA 基本原理提出了一种新颖的特征提取方法：线性拉普拉斯判别准则 (Linear Laplacian discrimination, LLD)^[31]。LLD 作为一种非线性的特征提取方法^[32]，则是通过对类内散度和类间散度分别加权的方式来弱化距离度量对原始样本空间的依赖，从而提高了特征提取方法的泛化能力。然而 LLD 准则中距离度量虽然具有较强的泛化能力，但还属于样本间标准

的距离函数，也就是说在实际使用 LLD 方法时还需要首先确定原始样本空间类型，并根据相应空间类型去寻找适合的距离度量，但 LLD 并没有提供如何确定原始样本空间类型的方法。同时 LLD 方法同 LDA 方法一样也存在着小样本问题，虽然 Zhang 等提出的张量线性拉普拉斯判别准则 (Tensor linear Laplacian discrimination, TLLD)^[32] 可以在一定程度上避免 LLD 的小样本问题，但 TLLD 属于张量方法，其特征提取的效果往往要依据相应的张量模式，特别是对于一些矢量表示的高维数据 (比如基因数据) 如何张量化以及张量方法与相应的矢量方法相比是否具有更好的特征提取能力都是有待讨论的问题^[33]。近来语境距离度量 (Contextual-distance metric) 作为一种新的距离度量得到了较为成功的应用^[32-36]，其实语境距离并不是真正意义上的距离概念，而反映的是样本间的不相似性，同时由于语境集 (Contextual set) 的存在将导致样本间的语境距离是不对称的。语境距离度量关注的是样本对数据簇结构的贡献程度，如果两个样本相对于同一个簇结构的贡献程度相当，则说明样本间的不相似性较小，即语境距离较小，反之则较大，而且数据的簇结构一般是不依赖于特定输入样本空间类型的^[34-35]，因此语境距离度量更具有广泛性。

因此，本文依据 LLD 方法提出基于语境距离度量的拉普拉斯最大间距准则 (Contextual-distance metric based Laplacian maximum margin criterion, CLMMC)，该方法具有以下几点优势：

- 1) CLMMC 方法不但继承了 LLD 方法具有的非线性特征提取能力，同时还能有效地解决 LLD 方法的小样本问题，而且还实现了 MMC 方法的非线性学习能力；
- 2) CLMMC 方法由于引入了语境距离度量使得该方法相对独立于原始空间的空间类型，提高了特征提取的泛化能力，同时引入新的语境距离形式使得在处理高维矢量数据时更具适应性；
- 3) CLMMC 方法结合 QR 分解技术从而提高了处理高维矢量数据的效率；
- 4) CLMMC 方法还从理论上讨论了与 LLD 方法内在的联系，

本文的内容安排如下：第 1 节简单介绍 LLD 方法；第 2 节重点讨论本文的 CLMMC 方法；第 3 节讨论 CLMMC 方法的权值设定；第 4 节对 CLMMC 进行实验研究；第 5 节总结全文以及提出未来期待解决的问题。

1 线性拉普拉斯判别准则：LLD

对于任意 D 维的原始样本空间 M^D 包含有 n 个样本的样本集 $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$ ，其中，该

样本集被分成 c 类, 第 $s = 1, \dots, c$ 类记为 $X_s = [\mathbf{x}_1^s, \mathbf{x}_2^s, \dots, \mathbf{x}_{c_s}^s]$, LLD 准则的目的就是通过寻找相应的变换矩阵 U 使得样本集被嵌入到低维的欧氏子空间 $Y = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n]$, 即

$$\mathbf{y}_i = U^T \mathbf{x}_i \quad (1)$$

因此, 为了得到更有效的变换矩阵 U , LLD 方法使用泛化的距离度量作为权值来定义新的类内散度和类间散度.

定义 1^[31]. 假设 α_s 是第 s 类所对应的类内散度, 则 LLD 的类内散度 α 为

$$\alpha = \sum_{s=1}^c \alpha_s = \sum_{s=1}^c \sum_{i=1}^{c_s} \omega_i^s \|\mathbf{y}_i^s - \bar{\mathbf{y}}^s\| = \sum_{s=1}^c \sum_{i=1}^{c_s} \exp\left(-\frac{\|\mathbf{x}_i^s - \bar{\mathbf{x}}^s\|_{M^D}}{\sigma}\right) \|\mathbf{y}_i^s - \bar{\mathbf{y}}^s\| \quad (2)$$

其中, $\|\cdot\|_{M^D}$ 表示任意输入样本空间 M^D 距离度量, $\|\cdot\|$ 表示欧氏空间的 l_2 范数, σ 是自由变量.

如果将式 (1) 代入式 (2) 就可以得到 LLD 类内散度的另外一种比较直观的形式:

$$\alpha = \sum_{s=1}^c \alpha_s = \sum_{s=1}^c \sum_{i=1}^{c_s} \omega_i^s \|\mathbf{y}_i^s - \bar{\mathbf{y}}^s\| = U^T H_w U \quad (3)$$

其中, $H_w = \sum_{s=1}^c \sum_{i=1}^{c_s} \omega_i^s (\mathbf{x}_i^s - \bar{\mathbf{x}}^s)(\mathbf{x}_i^s - \bar{\mathbf{x}}^s)^T$ 为类内散度矩阵.

同样可以定义 LLD 准则的类间离散度 β 为^[31]:

$$\beta = \sum_{s=1}^c \omega^s \|\bar{\mathbf{y}}^s - \bar{\mathbf{y}}\| = \sum_{s=1}^c \exp\left(-\frac{\|\bar{\mathbf{x}}^s - \bar{\mathbf{x}}\|_{M^D}}{\sigma}\right) \|\bar{\mathbf{y}}^s - \bar{\mathbf{y}}\| = U^T H_b U \quad (4)$$

其中, $H_b = \sum_{s=1}^c \omega^s (\bar{\mathbf{x}}^s - \bar{\mathbf{x}})(\bar{\mathbf{x}}^s - \bar{\mathbf{x}})^T$ 为类间散度矩阵.

根据式 (3) 和 (4) 可以得到 LLD 判别准则^[31]:

$$\arg \max_{U^T U = I_d} J(U) = \arg \max_{U^T U = I_d} \frac{U^T H_b U}{U^T H_w U} \quad (5)$$

其中, $I_d = \text{diag}\{\underbrace{1, \dots, 1}_d\}$.

从 LLD 的类内散度、类间散度定义可以看出, 当自由变量 $\sigma = +\infty$ 时, 判别准则就退化为 LDA^[31]; 同时由于 LLD 准则在处理原始样本间距离时使用了距离度量 $\|\cdot\|_{M^D}$, 而该距离度量是不依赖具体原始样本空间类型的, 从而大大降低了原始

样本间距离度量对输入样本空间类型的依赖程度, 并且在一些非线性空间进行距离学习时, 该距离度量还可以自动地将原始样本进行非线性映射^[31], 实现 LLD 非线性学习能力, 因此 LLD 准则是一种具有较强泛化能力的非线性特征提取方法.

然而, LLD 方法也存在一些问题: 1) LLD 准则并没有提供根据原始样本特征去确定相应空间类型的方法. 这是因为在使用 LLD 方法处理实际问题时还是需要首先确定原始样本空间的空间类型, 并根据相应的空间类型去选择适当的距离度量, 这样才能保证较好的特征提取能力. 比如文献 [31] 在测试 LLD 算法时, 就首先确定原始样本空间属于非欧氏空间, 并选择适合该空间类型的距离度量: χ^2 距离, 因此取得了较好的非线性特征提取效果; 2) LLD 方法同 LDA 一样都面临小样本问题, 即在处理小样本高维数据集时容易导致矩阵发生奇异.

因此, 本文提出基于语境距离度量拉普拉斯最大间距判别准则 (Contextual-distance based Laplacian maximum margin criterion, CLMMC), 该方法由于使用最大间距准则, 从而在一定程度上避免了 LLD 方法具有的小样本问题; 同时 CLMMC 准则中使用语境距离度量替换 LLD 中的 $\|\cdot\|_{M^D}$ 度量, 由于语境距离度量更关注原始样本簇的本质结构而并不注重原始样本空间的类型^[34-35], 从而在一定程度上保证了 CLMMC 准则与原始样本空间类型的相对独立性, 于是较好地解决了在实际使用 LLD 准则过程中需要确定原始样本空间类型的问题; 通过引入新的语境距离度量计算形式并结合 QR 分解基本原理, 使得 CLMMC 准则在处理高维矢量数据时更具适应性和效率.

2 基于语境距离度量的拉普拉斯最大间距判别准则: CLMMC

2.1 CLMMC 准则

由于在 CLMMC 准则中引入了语境距离度量, 因此本文重新定义类内散度 α 、类间散度 β 如下:

$$\begin{aligned} \alpha &= \sum_{s=1}^c \alpha_s = \sum_{s=1}^c \sum_{i=1}^{c_s} \omega_i^s \|\mathbf{y}_i^s - \bar{\mathbf{y}}^s\| = \\ &= \sum_{s=1}^c \sum_{i=1}^{c_s} \exp\left(-\frac{p(\mathbf{x}_i^s \rightarrow \bar{\mathbf{x}}^s)^2}{\sigma_w}\right) \|\mathbf{y}_i^s - \bar{\mathbf{y}}^s\| = \\ &= U^T H_w U \quad (6) \\ \beta &= \sum_{s=1}^c \omega^s \|\bar{\mathbf{y}}^s - \bar{\mathbf{y}}\| = \sum_{s=1}^c \exp\left(-\frac{p(\bar{\mathbf{x}}^s \rightarrow \bar{\mathbf{x}})^2}{\sigma_b}\right) \|\bar{\mathbf{y}}^s - \bar{\mathbf{y}}\| = U^T H_b U \quad (7) \end{aligned}$$

其中, $H_w = \sum_{s=1}^c \sum_{i=1}^{c_s} \omega_i^s (\mathbf{x}_i^s - \bar{\mathbf{x}}^s)(\mathbf{x}_i^s - \bar{\mathbf{x}}^s)^T$, $H_b = \sum_{s=1}^c \omega^s (\bar{\mathbf{x}}^s - \bar{\mathbf{x}})(\bar{\mathbf{x}}^s - \bar{\mathbf{x}})^T$ 为半正定矩阵.

需要说明的是上述定义中虽然使用的符号与 LLD 准则中的符号相同, 但由于在 CLMMC 中使用了语境距离度量 $p(\mathbf{x}_i \rightarrow \mathbf{x}_j)$, 因此具体内容和直观含义是明显不同的. 至于式 (6) 和 (7) 中的语境度量的设定将在第 3 节详细讨论.

由此, 根据式 (6) 和 (7) 定义的类内散度、类间散度, 就可以得到 CLMMC 判别准则.

定义 2. 设 $\gamma \in [0, 1]$ 是一实数变量, 则 CLMMC 判别准则为

$$\arg \max_{U^T U = I_d} J(U) = \arg \max_{U^T U = I_d} \gamma \beta - (1-\gamma)\alpha = \arg \max_{U^T U = I_d} U^T (\gamma H_b - (1-\gamma)H_w) U \quad (8)$$

其中, $I_d = \text{diag}\{\underbrace{1, \dots, 1}_d\}$.

为了求解式 (8), 我们使用拉格朗日乘数法来构造式 (8) 相对应的拉格朗日函数:

$$L(\mathbf{v}, \lambda) = \mathbf{v}^T (\gamma H_b - (1-\gamma)H_w) \mathbf{v} - \lambda (\mathbf{v}^T \mathbf{v} - 1) \quad (9)$$

要想式 (8) 成立, 就必须有:

$$\frac{\partial L}{\partial \mathbf{v}} = 0 \quad (10)$$

即有下式成立:

$$(\gamma H_b - (1-\gamma)H_w) \mathbf{v} = \lambda \mathbf{v} \quad (11)$$

因此, CLMMC 准则的变换矩阵 U 就是通过选取 $\gamma H_b - (1-\gamma)H_w$ 矩阵 d 个最大特征值对应的正交特征单位向量 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d$ 组成的, 即 $U = [\mathbf{v}_1, \dots, \mathbf{v}_d]$.

根据定义 2 可知, CLMMC 准则中参数 γ 起到了权衡 H_b 与 H_w 的作用, 从而表明变换矩阵 U 实质上是关于参数 γ 的函数. 因此, 本文结合文献 [18] 的相关知识和方法从理论上讨论参数 γ 对 CLMMC 准则的影响. 为了便于讨论我们首先将式 (8) 表示成参数 γ 的函数形式:

$$\phi(\gamma) = \mathbf{v}(\gamma)^T (\gamma H_b - (1-\gamma)H_w) \mathbf{v}(\gamma) = \max_{\mathbf{v}^T \mathbf{v} = 1} \mathbf{v}^T (\gamma H_b - (1-\gamma)H_w) \mathbf{v} \quad (12)$$

定理 1. $\phi(\gamma)$ 是 γ 的单调增函数, 特别是当 H_w 非奇异时, $\phi(\gamma)$ 是严格单调增函数.

证明. 设 $\gamma_1, \gamma_2 \in [0, 1]$ 且 $\gamma_1 > \gamma_2$, \mathbf{v}_1 和 \mathbf{v}_2 分别为 $\gamma_1 H_b - (1-\gamma_1)H_w$ 、 $\gamma_2 H_b - (1-\gamma_2)H_w$ 最大

特征值对应的单位特征向量. 则:

$$\begin{aligned} \phi(\gamma_1) &= \mathbf{v}_1^T (\gamma_1 H_b - (1-\gamma_1)H_w) \mathbf{v}_1 \geq \\ & \mathbf{v}_2^T (\gamma_1 H_b - (1-\gamma_1)H_w) \mathbf{v}_2 = \\ & \gamma_1 \left[\mathbf{v}_2^T \left(H_b - \left(\frac{1}{\gamma_1} - 1 \right) H_w \right) \mathbf{v}_2 \right] = \\ & \gamma_1 \left[\mathbf{v}_2^T \left(H_b - \left(\frac{1}{\gamma_2} - 1 \right) H_w \right) \mathbf{v}_2 + \right. \\ & \left. \left(\frac{1}{\gamma_2} - \frac{1}{\gamma_1} \right) \mathbf{v}_2^T H_w \mathbf{v}_2 \right] = \\ & \frac{\gamma_1}{\gamma_2} \left[\mathbf{v}_2^T (\gamma_2 H_b - (1-\gamma_2)H_w) \mathbf{v}_2 + \right. \\ & \left. \left(1 - \frac{\gamma_2}{\gamma_1} \right) \mathbf{v}_2^T H_w \mathbf{v}_2 \right] = \\ & \frac{\gamma_1}{\gamma_2} \left[\phi(\gamma_2) + \left(1 - \frac{\gamma_2}{\gamma_1} \right) \mathbf{v}_2^T H_w \mathbf{v}_2 \right] \end{aligned}$$

$\because \gamma_1 > \gamma_2, H_w$ 是半正定矩阵

$$\therefore \left(1 - \frac{\gamma_2}{\gamma_1} \right) \mathbf{v}_2^T H_w \mathbf{v}_2 \geq 0 \quad (13)$$

$$\therefore \phi(\gamma_1) \geq \phi(\gamma_2) \quad (14)$$

根据式 (14) 得 $\phi(\gamma)$ 是 γ 的单调增函数.

如果 H_w 非奇异, 则 $(1-\gamma_2/\gamma_1)\mathbf{v}_2^T H_w \mathbf{v}_2 > 0$, 有 $\phi(\gamma_1) > \phi(\gamma_2)$ 严格成立, 所以 $\phi(\gamma)$ 是 γ 的严格单调增函数. \square

根据定理 1, 我们可以得到 $\phi(\gamma)$ 的取值范围, 假设类内散度矩阵的最小特征值为 λ' , 类间散度矩阵的最大特征值为 λ'' , 则 $\phi(\gamma) \in [-\lambda', \lambda'']$. 特别地, 对于 $\forall \gamma \in [0, 1]$ 有下式成立:

$$-\lambda' \leq \phi(\gamma) \leq \gamma \lambda'' - (1-\gamma)\lambda' \quad (15)$$

定理 2. $\phi(\gamma)$ 是关于 γ 的一个连续函数.

证明. 根据连续函数的定义: 假设任意给定 $\gamma_c \in [0, 1]$, 对任意 $\varepsilon > 0$, 存在一个 $\delta > 0$, 如果 $\forall \gamma$ 满足 $|\gamma - \gamma_c| < \delta$, 那么有:

$$|\phi(\gamma) - \phi(\gamma_c)| = |(\mathbf{v}(\gamma) - \mathbf{v}(\gamma_c))^T (H_b + H_w) \times (\mathbf{v}(\gamma) - \mathbf{v}(\gamma_c))| |\gamma - \gamma_c| \quad (16)$$

因为 $H_t = H_b + H_w$ 是半正定矩阵, 那么设矩阵 H_t 的最大特征值为 λ_t , 则由式 (16) 可得:

$$|\phi(\gamma) - \phi(\gamma_c)| \leq \lambda_t |\gamma - \gamma_c| < (\lambda_t + 1) |\gamma - \gamma_c| \quad (17)$$

根据式 (17), 取 $\delta = \varepsilon / (\lambda_t + 1) > 0$, 并代入式 (17), 则有 $|\phi(\gamma) - \phi(\gamma_c)| < \varepsilon$, 所以定理成立. \square

定理 3. 在 CLMMC 准则中, 当 H_w 非奇异时必定存在唯一的一个 $\gamma_p \in [0, 1]$, 使得 LLD 准则对

应的最优鉴别矢量即为 $\phi(\gamma_p)$ 所对应的单位特征向量.

证明. 根据 LLD 准则, 令 $c' = \arg \max_{\mathbf{v}^T \mathbf{v}=1} \mathbf{v}^T H_b \mathbf{v} / (\mathbf{v}^T H_w \mathbf{v}) = \lambda'' / \lambda' = \mathbf{v}^{*T} H_b \mathbf{v}^* / (\mathbf{v}^{*T} H_w \mathbf{v}^*)$, 因此对于任意单位向量 \mathbf{v} , 都有 $\mathbf{v}^T H_b \mathbf{v} / (\mathbf{v}^T H_w \mathbf{v}) \leq \lambda'' / \lambda'$, 因此, 取 $\gamma_p = \lambda' / (\lambda'' + \lambda')$, 则:

$$\gamma_p \mathbf{v}^{*T} H_b \mathbf{v}^* - (1 - \gamma_p) \mathbf{v}^{*T} H_w \mathbf{v}^* = \frac{\lambda'}{\lambda'' + \lambda'} \mathbf{v}^{*T} H_b \mathbf{v}^* - \left(1 - \frac{\lambda'}{\lambda'' + \lambda'}\right) \mathbf{v}^{*T} H_w \mathbf{v}^* = (18)$$

$$\frac{\lambda'}{\lambda'' + \lambda'} \left[\mathbf{v}^{*T} H_b \mathbf{v}^* - \left(\frac{\lambda''}{\lambda'}\right) \mathbf{v}^{*T} H_w \mathbf{v}^* \right] (19)$$

根据 LLD 准则, 因为 \mathbf{v}^* 是最大特征值 λ'' / λ' 所对应的单位特征向量, 因此对任意单位向量 \mathbf{v} 并结合式 (19) 都有:

$$\begin{aligned} \frac{\lambda'}{\lambda'' + \lambda'} [\mathbf{v}^{*T} H_b \mathbf{v}^* - \left(\frac{\lambda''}{\lambda'}\right) \mathbf{v}^{*T} H_w \mathbf{v}^*] &\geq \\ \frac{\lambda'}{\lambda'' + \lambda'} [\mathbf{v}^T H_b \mathbf{v} - \left(\frac{\lambda''}{\lambda'}\right) \mathbf{v}^T H_w \mathbf{v}] &= \\ \frac{\lambda'}{\lambda'' + \lambda'} \mathbf{v}^T H_b \mathbf{v} - \left(1 - \frac{\lambda'}{\lambda'' + \lambda'}\right) \mathbf{v}^T H_w \mathbf{v} &(20) \end{aligned}$$

根据式 (20), 显然式 (18) 等号左边即为 $\phi(\gamma_p)$, \mathbf{v}^* 即为 $\phi(\gamma_p)$ 对应的单位特征向量. 同时因为 H_w 非奇异, 则根据定理 1 和定理 2, γ_p 是唯一的. \square

需要说明的是, 在定理 3 中提到的 LLD 方法所对应的权值已经发生了变化, 即使用了式 (6) 和式 (7) 中定义的权值来代替, 是一种基于语境距离度量的 LLD, 从而我们将这种基于语境距离度量的 LLD 记为 CL-LLD.

2.2 CLMMC 准则的 QR 分解法

从式 (11) 可以看出, CLMMC 准则关键就是要求解矩阵 $\gamma H_b - (1 - \gamma) H_w$ 的 d 个最大特征值所对应的单位特征向量. 然而当原始样本空间 M^D 呈现高维状态时, 即维数 D 比较大, 求解上述矩阵的特征值具有较高的时间复杂度 ($O(D^3)$), 为此将 QR 分解方法引入到 CLMMC 准则中以降低求解特征值的时间复杂度.

为了便于 QR 分解, 本文使用文献 [31] 中的方法分别将式 (6) 和式 (7) 写成关于“迹运算”的等价形式, 即转换成具有拉普拉斯矩阵的形式. 如果令 $X = [\mathbf{x}_1^1, \dots, \mathbf{x}_{c_1}^1, \dots, \mathbf{x}_1^s, \dots, \mathbf{x}_{c_s}^s, \dots, \mathbf{x}_1^c, \dots, \mathbf{x}_{c_c}^c]$, $W_s = \text{diag}\{\omega_1^s, \dots, \omega_{c_s}^s\}$, $W_b = \text{diag}\{\omega^1, \dots, \omega^c\}$, $\mathbf{e}_{c_s} = \underbrace{[1, \dots, 1]}_{c_s}$, $\mathbf{e}_c = \underbrace{[1, \dots, 1]}_c$, 则:

$$\alpha = \text{tr}(U^T X L_w X^T U) (21)$$

其中, $L_w = \text{diag}\{L_1, \dots, L_s, \dots, L_c\}$, $L_s = W_s - 2W_s \mathbf{e}_{c_s} (\mathbf{e}_{c_s})^T / c_s + (\mathbf{e}_{c_s})^T W_s \mathbf{e}_{c_s} (\mathbf{e}_{c_s})^T / c_s^2$.

$$\beta = \text{tr}(U^T X B L_b B^T X^T U) (22)$$

其中, $B = \text{diag}\{\mathbf{e}_{c_1} / c_1, \dots, \mathbf{e}_{c_c} / c_c\}$, $L_b = W_b - 2W_b \mathbf{e}_c (\mathbf{e}_c)^T / c + (\mathbf{e}_c)^T W_b \mathbf{e}_c (\mathbf{e}_c)^T / c^2$.

由式 (21) 和式 (22), CLMMC 准则目标函数的拉普拉斯形式为

$$\arg \max_{U^T U = I_d} J(U) = \arg \max_{U^T U = I_d} \text{tr}(U^T X L X^T U) (23)$$

其中, $I_d = \text{diag}\{\underbrace{1, \dots, 1}_d\}$, $L = \gamma B L_b B^T - (1 - \gamma) L_w$.

根据文献 [29], 可以得到 $X = QR$, 其中 $Q \in M^{D \times r}$ 由一组正交列向量组成, $R \in M^{r \times r}$ 是上三角矩阵, $r = \text{rank}(X)$. 由文献 [29] 的定理 1 可知, 求解式 (23) 可以转化为求解下式:

$$\arg \max_{Z^T Z = I_d} J(Z) = \arg \max_{Z^T Z = I_d} Z^T R L R^T Z (24)$$

并且式 (23) 中的 U 与式 (24) 中的 Z 的关系为: $U = QZ$.

上述 QR 分解方法表明, 求解矩阵 $X L X^T$ 的 d 个最大特征值所对应的单位特征向量转换成求解矩阵 $R L R^T$ 的 d 个最大特征值所对应的单位特征向量, 这样做的好处就是使得原方法的时间复杂度降低为 $O(r^3)$, 特别相对于高维小样本数据更有优势.

3 CLMMC 准则的权值设定

在 CLMMC 准则中使用了语境距离度量 $p(\mathbf{x}_i \rightarrow \mathbf{x}_j)$, 从而一定程度上降低了对原始样本空间类型的依赖, 提高了该准则特征提取的泛化能力. 根据文献 [34–35], 我们知道语境距离度量是定义在相应的语境集 Ω 上, 该度量更关注于输入样本簇内在的几何结构, 而不是具体的原始样本空间的类型. 该距离度量是通过定义相应的结构化描述器函数 $f(\Omega)$ 来反映数据簇内在的几何特征, 即如果 \mathbf{x}_i 能融于语境集 Ω 所表明的数据内在的几何结构, 那么当把 \mathbf{x}_i 从 Ω 移走后并不会对原有数据内在结构产生太大影响, 而如果 \mathbf{x}_i 是孤立点或噪音样本, 那么当把 \mathbf{x}_i 从 Ω 移走后就会对原有数据内在几何结构产生较大影响. 通常在语境距离度量的概念中使用 $\delta f_i = \|f(\Omega) - f(\Omega \setminus \mathbf{x}_i)\|$ 来表示样本 \mathbf{x}_i 对 Ω 所表明的内在几何结构的贡献程度, 由此就可以定义语境距离度量^[34–35]:

$$p(\mathbf{x}_i \rightarrow \mathbf{x}_j) = |\delta f_i - \delta f_j| (25)$$

从式 (25) 可以看出语境距离度量实质表明的是语境集内部样本之间的不相似性, $p(\mathbf{x}_i \rightarrow \mathbf{x}_j)$ 越小越说明 \mathbf{x}_i 与 \mathbf{x}_j 属于同一个簇结构的可能性越大, 反之说明两个样本不属于同一簇。

在实际使用语境距离度量时, 簇结构描述器函数 $f(\Omega)$ 一般通过两种方法来定义: 一是质心法, 二是编码长度法 (Coding length). 所谓质心描述函数 $f(\Omega) = \frac{1}{|\Omega|} \sum_{\mathbf{x}_i \in \Omega} \mathbf{x}_i$ 即是通过语境集的质心来反映相应簇内在的几何结构, 然而通过研究发现质心法更适用于团状数据保持内在的几何结构, 而对于非团状数据表现不太明显^[34-35]; 编码长度法又称为编码率法, 该方法则基于信息论基本原理, 采用对原始信息重新编码所具有的信息量 (即编码长度) 来作为信息率失真函数 (Rate distortion gunction), 因此编码长度法更关注语境集中信息量的变化对簇内在几何结构的影响, 同时文献 [32, 34-35] 还分别从理论上和实例测试上来说明使用该簇结构描述器构造的语境距离度量更能保持全局的几何结构, 表现出更好的鲁棒性和适应性, 因此本文采用编码长度法来定义描述函数 $f(\Omega)$ 。

定义 3^[34-35]. 如果数据集 $\Omega = [\mathbf{x}_1, \dots, \mathbf{x}_k]$, $\bar{\mathbf{x}} = \Omega \mathbf{e}/k$, 那么令 $\bar{\Omega} = \Omega - \bar{\mathbf{x}} \mathbf{e}^T$, 则 Ω 的编码长度为

$$f(\Omega) = \frac{k+D}{2} \log \det \left(I + \frac{D}{\varepsilon^2 k} \bar{\Omega} \bar{\Omega}^T \right) + \frac{D}{2} \log \left(1 + \frac{\bar{\mathbf{x}}^T \bar{\mathbf{x}}}{\varepsilon^2} \right) \quad (26)$$

其中, ε 为可允许偏差, 通常令 $\varepsilon = \sqrt{10D/(K-1)}$ ^[32, 34-36], D 表示矢量维数。

3.1 权值的设定

根据以上分析并根据式 (25) 和式 (26), 可以较为容易地定义式 (6) 中的语境距离以及参数 σ :

$$p(\mathbf{x}_i^s \rightarrow \bar{\mathbf{x}}^s) = |\delta f_i^s - \delta \bar{f}^s| \quad (27)$$

其中, $\delta f_i^s = \|f(X_s) - f(X_s \setminus \mathbf{x}_i^s)\|$, $\delta \bar{f}^s$ 和 σ_w 分别为 $[\delta f_1^s, \dots, \delta f_{c_s}^s]$ 的均值和方差。

式 (27) 从形式上看与文献 [34] 中的语境距离相似, 不同的是在文献 [34] 中语境集是通过无监督 K -近邻方法得到的, 而本文的 CLMMC 作为一种有监督的方法则认为每类训练样本就是类内每个样本的语境集, 这一点与文献 [32] 相似. 然而在定义 (7) 中的语境距离 $p(\bar{\mathbf{x}}^s \rightarrow \bar{\mathbf{x}})$ 时却遇到了一定的困难。

从概念上讲, $p(\bar{\mathbf{x}}^s \rightarrow \bar{\mathbf{x}})$ 反映的是每类训练样本均值与总体均值之间的不相似性, 同时结合类间散度的直观意义, $p(\bar{\mathbf{x}}^s \rightarrow \bar{\mathbf{x}})$ 也反映了不同类样本

之间的不相似性, 因此为了保证类间最大, 即类间的不相似程度较高, 则要求 $p(\bar{\mathbf{x}}^s \rightarrow \bar{\mathbf{x}})$ 较大. 比较直观的定义方式就是使用文献 [32] 中定义方式 (即 $p(\bar{\mathbf{x}}^s \rightarrow \bar{\mathbf{x}}) = \|f(X) - f(X \setminus X_s)\|$), 该设定方法表明了每类训练样本的内在几何结构对保持样本全局结构的贡献, 然而当处理矢量模式的高维数据集 (比如基因数据、图像数据) 时, 发现该定义一定程度上容易导致 $p(\bar{\mathbf{x}}^s \rightarrow \bar{\mathbf{x}})$ 为无穷大量, 权值为无穷小量, 从而使得测试程序无法进行, 因此有必要讨论适合处理高维矢量模式和本文 CLMMC 方法的定义方式. 下面首先给出计算式 (7) 中的语境距离 $p(\bar{\mathbf{x}}^s \rightarrow \bar{\mathbf{x}})$ 和参数 σ_b 的算法, 然后从直观含义上解释算法的合理性。

算法 1. 计算式 (7) 中语境距离 $p(\bar{\mathbf{x}}^s \rightarrow \bar{\mathbf{x}})$ 和参数 σ_b 的算法。

步骤 1. 输入训练样本集 X 、类别数 c , 初始化 $s = 1$;

步骤 2. 计算第 s 类训练样本集 X_s 的均值 $\bar{\mathbf{x}}^s$;

步骤 3. 设第 k 类训练样本集 X_k ($k = 1, \dots, c, k \neq s$), 并令 $X'_k = X_k \cup \{\bar{\mathbf{x}}^s\}$, 分别计算 $\delta f_k^s = \|f(X'_k) - f(X_k)\|$ ($k = 1, \dots, c, k \neq s$);

步骤 4. 令 δf^s 为 $[\delta f_1^s, \dots, \delta f_c^s]$ 的均值;

步骤 5. 令 $s = s + 1$, 如果 $s \leq c$ 则转到步骤 2, 否则转到步骤 6;

步骤 6. 令 $\delta \bar{f}$ 和 σ_b 分别为 $[\delta f^1, \dots, \delta f^c]$ 的均值和方差;

步骤 7. 输出 $p(\bar{\mathbf{x}}^s \rightarrow \bar{\mathbf{x}})$ 、 σ_b 。

算法 1 中使用均值 $\bar{\mathbf{x}}^s$ 来代表第 s 类训练样本集 X_s , 这和一般处理方法相同. 然而在整个算法中最关键的是步骤 3 和步骤 4, 通过这两步算法计算得到了反映第 s 类与其他类别数据集之间的不相似性度量 δf^s . 在步骤 3 中定义新数据集 X'_k 并计算 δf_k^s 的目的在于表明第 s 类训练样本均值 $\bar{\mathbf{x}}^s$ 与第 k 类训练样本 X_k 之间的不相似性, 也就可以在一定程度上表明第 s 类训练样本集 X_s 与第 k 类训练样本 X_k 之间的不相似性, 从而可以看出步骤 4 计算得到的 δf^s 反映的是第 s 类训练样本集 X_s 与其他 $c-1$ 类训练样本之间的平均不相似性, 也反映了第 s 类训练样本集 X_s 内在几何结构与其他 $c-1$ 类训练样本几何结构之间的差异性, 从而体现了第 s 类训练样本集 X_s 内在几何结构对保持全局几何结构的贡献, 因此从这个意义上讲算法 1 处理原则是符合语境距离 $p(\bar{\mathbf{x}}^s \rightarrow \bar{\mathbf{x}})$ 直观含义的. 同时通过大量测试发现该算法确实一定程度上克服了文献 [32] 的定义方式在处理高维矢量数据时所遇到的困难. 诚然由于步骤 3 需要多次计算, 从而在一定程度上提高了算法 1 的时间复杂度。

算法 2. CLMMC 准则。

步骤 1. 输入样本集 X 、类别数 c , 使用式 (26) 和式 (27) 并调用算法 1 分别计算 $p(\mathbf{x}_i^s \rightarrow \bar{\mathbf{x}}^s)$ 、 σ_w 和 $p(\bar{\mathbf{x}}^s \rightarrow \bar{\mathbf{x}})$ 、 σ_b ;

步骤 2. 根据式 (21) 和式 (22) 分别计算 L_w 、 $BL_b B^T$;

步骤 3. 根据 QR 分解的基本原理, 得到 $X = QR$, 其中根据文献 [29], 可以得到 $X = QR$, 其中 $Q \in M^{D \times r}$ 由一组正交列向量组成, $R \in M^{r \times n}$ 是上三角矩阵, $r = \text{rank}(X)$;

步骤 4. 设定参数 γ , 并根据式 (24) 计算得到由 d 个最大特征值对应的单位特征向量组成的矩阵 Z ;

步骤 5. 使用公式 $U = QZ$ 得到变换矩阵 U ;

步骤 6. 根据式 (1) 输出输入样本集 X 对应的低维嵌入集 $Y = [\mathbf{y}_1, \dots, \mathbf{y}_n]$.

4 实验

本文通过测试一个人造数据集 2moons、四个 UCI 数据子集 (即: IRIS 数据集、Ionosphere 数据集、Segment 数据集、Glass 数据集)^[37]、两个高维的基因数据集 (Leukemia 数据集、Colon 数据集) (<http://www.molbio.princeton.edu/>) 以及两个人脸识别数据集 (ORL、Yale 数据集) (<http://www.cs.uiuc.edu/homes/dengcai2/>) 来说明本文提出的 CLMMC 方法在处理类内散度矩阵非奇异、类内散度矩阵奇异的问题以及在保持图像样本内在的几何结构方面都具有较好的特征提取能力. 由于人工数据集和 UCI 数据集维数相对较低, 在测试过程中使用 CLMMC 方法时不采用 QR 分解方法, 而在测试高维基因数据和人脸图像数据时使用 QR 分解方法. 在实验过程中都使用最近邻分类器.

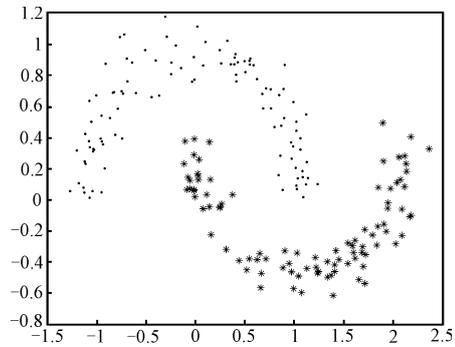
4.1 测试 2moons 数据集

人造 2moons 的数据集是一具有明显非线性流形结构 (见图 1, 其中图 1(a)、1(b) 分别是训练样本和测试样本). 从理论上讲欧氏空间是线性的黎曼流形空间, 而非线性黎曼流形空间是典型的非欧氏空间^[38], 从这一层面上讲 2moons 的数据集所蕴含的非线性流形结构表明该数据集的输入样本空间是非欧氏空间. 因此通过测试该数据集来表明本文 CLMMC 方法在处理输入样本空间是非欧氏空间时距离度量学习的有效性.

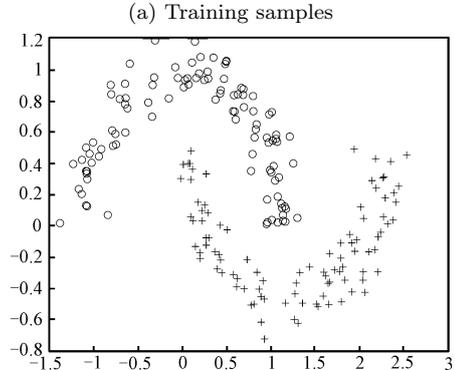
为了说明上述有效性, 我们对本文的 CLMMC 方法进行改造, 即使用欧氏空间的标准距离: Euclidean 范数 $\|\cdot\|$ 来替代本文 CLMMC 方法中语境距离, 我们把改造后的方法记为: Euclidean-CLMMC, 在此基础上同本文的 CLMMC 方法进

行比较. 测试过程中采用 10-折交叉验证, 并令 CLMMC 算法中参数 $\gamma = 2^{i-20}, i = 1, \dots, 20$, Euclidean-CLMMC 参数 $t = 2^i, i = -2, \dots, 7$, 测试结果见表 1.

从表 1 可以看出, 当输入样本空间是非欧氏空间时, 如果还使用经典的距离度量在一定程度上确实影响了特征提取的效果, 这就充分说明了本文的 CLMMC 方法由于引入了语境距离度量, 从而在一定程度上降低了距离度量学习对输入样本空间的依赖性, 提高了特征提取的效果.



(a) 训练数据



(b) 测试数据

图 1 2moons 数据集

Fig.1 2moons datasets

表 1 对 2moons 数据集的识别效果比较

Table 1 Recognition performance comparison on 2moons datasets

Algorithm	Training accuracy	Test accuracy
Euclidean-CLMMC	1	0.98
CLMMC	1	1

4.2 测试 UCI 数据集

UCI 数据集经常被一些特征提取方法通过测试来说明该方法的特征提取能力^[39-40]. 本文通过测试 4 个 UCI 数据子集 (见表 2) 来说明 CLMMC 方法在解决具有非奇异类内散度矩阵特点的实际问题时

表 2 Ionosphere、IRIS、segment、glass 数据集
Table 2 Ionosphere, IRIS, segment, and glass datasets

Datasets	The number of training samples	The number of test samples	The number of features	The number of subjects
Ionosphere	230	117	33	2
IRIS	100	50	4	3
Segment	1 540	770	18	7
Glass	140	72	9	6

表 3 对 Ionosphere、IRIS、segment、glass 数据集的识别效果比较
Table 3 Recognition performance comparison on ionosphere, IRIS, segment, and glass datasets

Datasets	Ionosphere		IRIS		Segment		Glass	
	Accuracy	Dim	Accuracy	Dim	Accuracy	Dim	Accuracy	Dim
LDA	0.78632	1	0.94	2	0.91948	5	0.77778	4
MMC	0.8974	5	0.94	4	0.95325	11	0.81944	7
LLD	0.92308	12	0.96	4	0.94805	15	0.81944	7
	($t = 512$)		($t = 0.0625$)		($t = 8$)		($t = 16$)	
CL-LLD	0.93162	5	0.96	4	0.95455	12	0.80556	8
CLMMC	0.94017	5	0.96	3	0.95844	10	0.83333	5
	($\gamma = 2^{-6}$)		($\gamma = 2^{-11}$)		($\gamma = 2^{-17}$)		($\gamma = 2^{-2}$)	

所表现出的特征提取能力,同时通过测试来验证本文定理 1、定理 3 所表明性质.

为了说明本文 CLMMC 方法具有较强的特征提取能力,我们将 CLMMC 方法同 LDA、MMC、LLD、CL-LLD 这 4 种特征提取方法进行比较.在测试过程中为了给不同算法选择合适的参数,本文使用 10-折交叉验证,从而令 CLMMC 算法中参数同第 4.1 节,LLD 算法中参数 $t = 2^i, i = -4, \dots, 9$.测试效果见表 3,并使用 Ionosphere 数据集的测试结果来表明特征提取对分类精度的影响(图 2),同时验证定理 1(图 3(a))、定理 3(图 3(b)).

根据表 3、图 2 和图 3 可以得到如下结论:

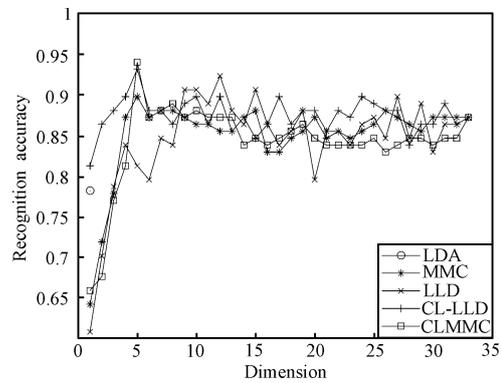
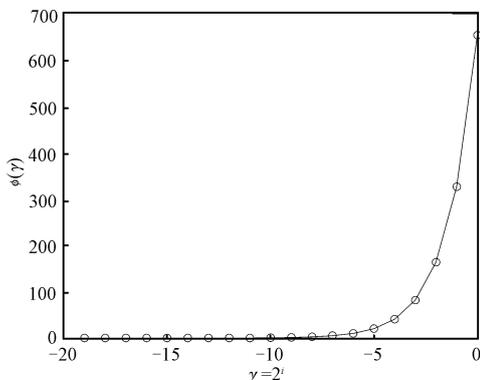
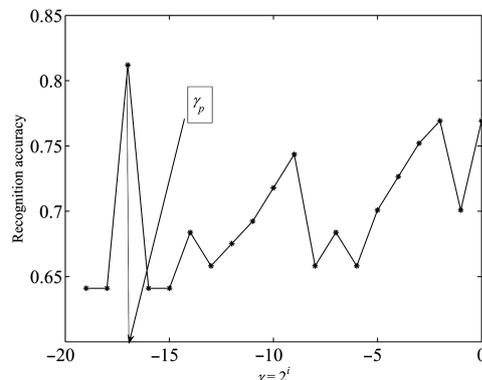


图 2 特征提取结果对分类的影响
Fig. 2 Recognition accuracy vs. dimension reduction



(a) 实验验证定理 1

(a) Experimental verification of Theorem 1



(b) 实验验证定理 3

(b) Experimental verification of Theorem 3

图 3 CLMMC 方法实验结果验证定理结论

Fig. 3 Experimental verification of the theorems by CLMMC method

1) 从表 3 可以看出, 本文的 CLMMC 方法同其他 4 种方法相比, 在处理具有非奇异类内散度矩阵特点的数据集时表现出较好的特征提取效果和分类精度, 这说明由于在 CLMMC 方法的权值中引入了语境距离度量, 从而使得该方法更关注输入样本空间本质的几何结构, 降低了对输入样本空间类型的依赖, 在一定程度上提高了该方法的泛化能力.

2) 图 2 说明的是上述 5 种特征提取方法在测试 Ionosphere 数据集时, 取得最好分类精度的情况下所表现的不同特征提取结果对分类精度的影响. 从图中可以看出本文的 CLMMC 方法同其他 4 种方法相似, 都具有随着特征提取维数的不断提高分类精度也有所变化的特点, 从这一层面上可以表明 CLMMC 符合特征提取方法的一般规律.

3) 图 3 表现的是本文 CLMMC 方法在使用 Ionosphere 数据集的测试结果来验证定理 1、定理 3 结论的效果. 由于 Ionosphere 数据集所对应的类内散度矩阵是非奇异的, 则根据定理 1 可知 $\phi(\gamma)$ 是 γ 的严格单调增函数, 这一结论完全可以根据图 3(a) 表现的实验结果得到验证; 而图 3(b) 则表明的是本文 CLMMC 方法在提取一个特征 (即 $d = 1$) 的情况下所对应的特征提取效果, 从图 3(b) 可以看出, 确实存在唯一的 γ_p 使得 CLMMC 与 CL-LLD 具有相同的特征提取效果, 从而有力地证明了定理 3 结论的正确性, 从这一层面上讲, 当类内散度矩阵非奇异并且特征提取为一维特征时, 本文的 CLMMC 方法完全可以替代 CL-LLD; 同时还可以从图 3(b) 看出本文 CLMMC 方法在提取相同维数特征后其特征提取效果是随着参数 γ 的变化而变化的.

4.3 测试基因数据集

基因数据作为典型的高维小样本数据经常被用来测试特征提取方法在处理小样本问题时的有效性^[41]. 本文使用 Leukemia、Colon 数据集 (见表 4) 来测试 CLMMC 方法在处理类内散度奇异问题的有效性.

在测试过程中, 本文的 CLMMC 方法中的参数 γ 选取范围和第 4.1 节相同. 然而通过测试发现, LLD 方法在处理基因数据时如果参数 t 选取太小将不能得到有效的权值, 因此在本测试过程中令 $t = 2^i, i = 8, \dots, 20$. 同时由于基因数据往往会导致类内散度发生奇异, 从而在使用 LDA、LLD、CL-LLD 方法测试之前首先运用 PCA 方法对相应的基因数据进行预处理, 即在每次使用上述方法之前本文只需选取 $n - 1$ (代表训练样本数) 个主成分. 即将 Leukemia、Colon 训练数据集的维数分别从 71 129 和 21 000 变换为 39 和 29, 尽管变换后维数大大降低, 但变换后的数据集包含了原数据集信息量的 99.5% 以上, 同时可以在一定程度上确保变换后的数据集对应的类内散度不会发生奇异性问题, 因此这样处理是合适的. 为了提高特征提取的有效性, 本测试过程使用 5-折交叉验证. 需要说明的是在使用 LDA 方法后只需提取 $c - 1$ (c 代表的是样本类别数) 维特征. 测试结果见表 5 和图 4.

根据表 5 和图 4 可以看出, 本文的 CLMMC 方法确实可以克服高维小样本数据的类内散度矩阵奇异性问题, 同时和 PCA+LDA、PCA+LLD、PCA+CL-LLD 的测试

表 4 Colon 和 Leukemia 数据集
Table 4 Colon and Leukemia datasets

Datasets	The number of training samples	The number of test samples	The number of features	The number of subjects	Rank
Colon	30	32	2 000	2	30
Leukemia	40	32	7 129	2	40

表 5 对 Colon 和 Leukemia 数据集识别效果比较
Table 5 Recognition performance comparison on Colon and Leukemia datasets

Datasets	Colon		Leukemia	
	Accuracy	Dim	Accuracy	Dim
PCA+LDA	0.71875	1	0.53125	1
PCA+LLD	0.90625 ($t = 512$)	9	0.84375 ($t = 4096$)	26
PCA+CL-LLD	0.90625	10	0.8750	15
CLMMC	0.9375 ($\gamma = 2^{-5}$)	15	0.9375 ($\gamma = 2^{-3}$)	9

效果相比具有较为明显的优势, 体现出 CLMMC 方法在处理高维的基因数据时具有较强的特征提取能力. 同时由于 CLMMC 方法引入了 QR 分解技术, 使得该方法在对高维数据进行特征提取时同其他最大间距判别准则相比具有较高的特征提取效率. 比如在上述测试过程中, 传统的最大间距判别准则需要分别求解 7129×7129 的散度矩阵 (Leukemia 数据)、 2000×2000 的散度矩阵 (Colon 数据) 特征值问题, 而本文的 CLMMC 方法只需分别求解 40×40 的散度矩阵 (Leukemia 训练数据集的秩为 40)、 30×30 的散度矩阵 (Colon 训练数据集的秩为 30) 特征值问题, 从而在一定程度上降低了求解高阶矩阵特征值问题的时间与空间复杂度.

4.4 测试人脸图像数据集

众所周知人脸图像数据集蕴含着内在的非线性流形结构^[11-17, 22-35], 即该原始样本空间属于非线性的黎曼流形空间. 为此通过测试 ORL (32×32)、Yale (32×32) 两个人脸图像数据集来反映本文提出的 CLMMC 方法在一定程度上能较好地保持输入样本空间内在的几何结构以及对原始样

本空间类型具有较低的依赖程度. ORL 人脸数据包含 40 个类别的人脸图像, 每一类别包含有 10 幅不同人脸表情的图像 (见图 5); Yale 人脸数据则包含有 15 个类别的人脸图像, 每一类别包含有 11 幅不同人脸表情的图像 (见图 6).

在本测试过程中为了有效地说明本文提出方法具有较好的特征提取能力, 我们使用 PCA+LDA、PCA+LPP、PCA+LLD 三种方法来对比测试本文提出的 PCA+CL-LLD 和 CLMMC 方法. 其中, PCA+LDA、PCA+LLD、PCA+CL-LLD 三种方法中的参数设置、PCA 方法的处理过程和第 4.3 节相同, 而在 PCA+LPP 测试过程中对热核参数 t 的设置则参考文献 [29] 的方法, 即 $t = (2^m)\sigma_0, m = -9, \dots, 10$, 其中, σ_0 代表训练样本的方差, PCA 数据预处理过程和其他几种方法相同. 同时为了提高测试结果的有效性, 在测试过程中分别对 ORL 和 Yale 数据集每类选取 3 个和 5 个样本作为训练样本来测试上述 5 种方法, 并分别使用 3-折交叉验证和 5-折交叉验证. 测试结果见表 6、图 7 和图 8.

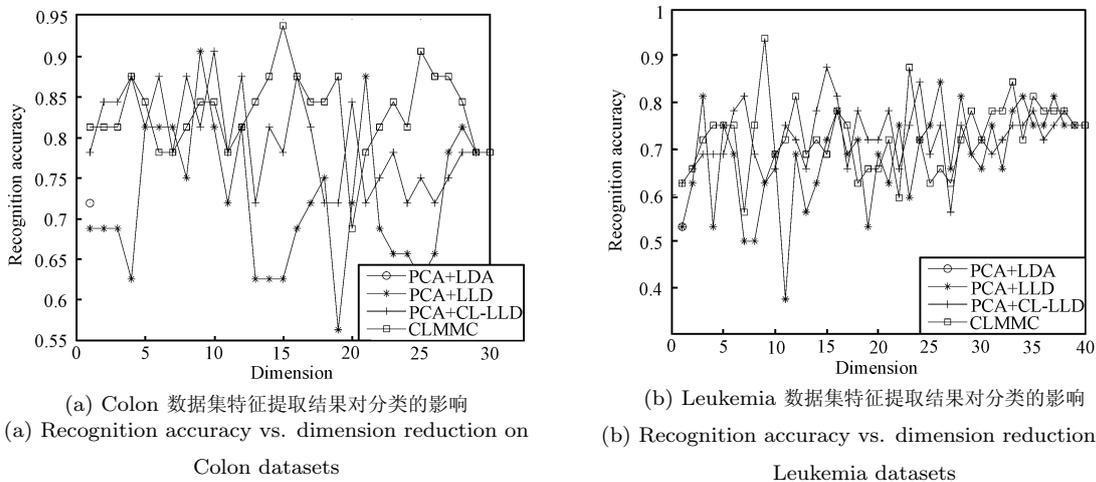


图 4 基因数据特征提取结果对分类的影响

Fig. 4 Recognition accuracy vs. dimension reduction on gene datasets

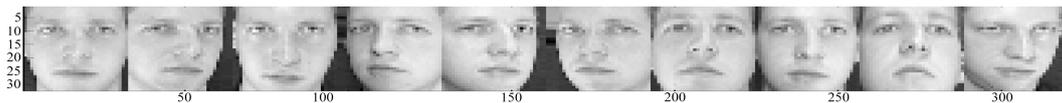


图 5 ORL 数据集中某一类所有图像

Fig. 5 All images in a certain class in ORL datasets

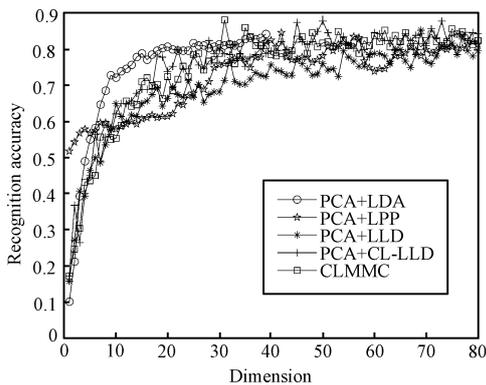


图 6 Yale 数据集中某一类所有图像

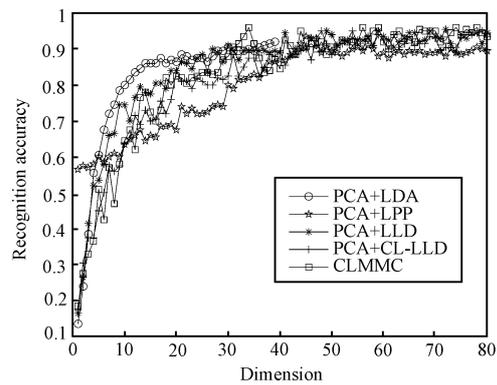
Fig. 6 All images in a certain class in Yale datasets

表 6 对 ORL 和 Yale 数据集识别效果比较
Table 6 Recognition performance comparison on ORL and Yale datasets

Datasets	ORL				Yale			
	3		5		3		5	
The number of train samples	Accuracy	Dim	Accuracy	Dim	Accuracy	Dim	Accuracy	Dim
PCA+LDA	0.84286	39	0.92	39	0.60833	14	0.76667	14
PCA+LPP	0.84659 ($m = 2^{-3}$)	42	0.9350 ($m = 2^4$)	45	0.61667 ($m = 2^{-5}$)	44	0.78889 ($m = 2^{-7}$)	74
PCA+LLD	0.85356 ($t = 2^{13}$)	69	0.950 ($t = 2^9$)	49	0.64167 ($t = 2^8$)	36	0.8 ($t = 2^{12}$)	35
PCA+CL-LLD	0.87501	45	0.950	48	0.63333	38	0.81111	50
CLMMC	0.88216 ($\gamma = 2^{-3}$)	31	0.960 ($\gamma = 2^5$)	34	0.65 ($\gamma = 2^{-3}$)	30	0.81111 ($\gamma = 2^{-8}$)	38



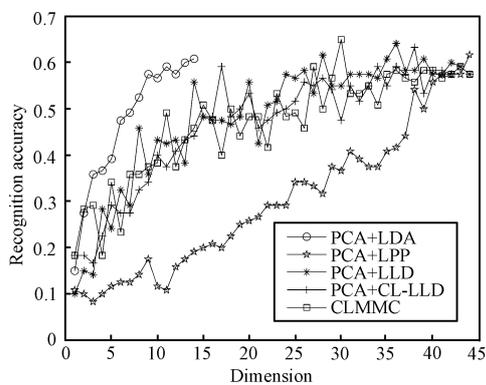
(a) 3 个训练样本
(a) 3 trains



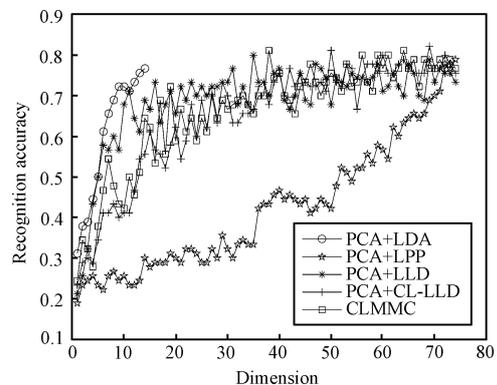
(b) 5 个训练样本
(b) 5 trains

图 7 ORL 数据特征提取结果对分类的影响

Fig. 7 Recognition accuracy vs. dimension reduction on ORL datasets



(a) 3 个训练样本
(a) 3 trains



(b) 5 个训练样本
(b) 5 trains

图 8 Yale 数据特征提取结果对分类的影响

Fig. 8 Recognition accuracy vs. dimension reduction on Yale datasets

根据以上测试结果可以看出, 本文提出的 PCA+CL-LLD 和 CLMMC 方法同其他方法相比具有更好的特征提取效果和分类精度, 这在一定程度上说明本文的方法由于引入了语境距离度量能更好地保持样本内在蕴含的几何结构, 降低了对原始样本空间类型的依赖程度。

5 总结

本文通过分析 LLD 方法存在的问题和不足, 通过引入语境距离度量并结合最大间距判别准则的基本原理提出一种基于语境距离度量的拉普拉斯最大间距判别准则: CLMMC. 该准则不但在一定程度上避免小样本问题, 而且由于语境距离度量更关注原始样本簇内在的本质结构而不是原始样本空间的类型, 从而降低了该准则对特定样本空间的依赖程度. 同时通过引入计算语境距离度量的新算法并结合 QR 分解的基本原理提高了 CLMMC 在处理高维矢量模式数据时的适应性和效率. 并从理论上讨论 CLMMC 准则具有的基本性质以及与 LLD 准则的内在联系. 最后通过测试证明 CLMMC 准则具有一定的有效性. 然而, CLMMC 准则也存在着一定的不足, 比如如何从理论上分析并结合运用实际去寻找更适合高维矢量数据的语境距离度量的计算形式将是我们以后工作的方向。

References

- Bian Zhao-Qi, Zhang Xue-Gong. *Pattern Recognition (Second Edition)*. Beijing: Tsinghua University Press, 2001. 87–90
(边肇祺, 张学工. 模式识别 (第二版). 北京: 清华大学出版社, 2001. 87–90)
- Jolliffe I T. *Principal Component Analysis*. New York: Springer-Verlag, 1986
- Tang F, Crabb R, Tao H. Representing images using nonorthogonal Harr-like bases. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, **29**(12): 2120–2134
- Skocaj D, Leonarids A, Bischof H. Weighted and robust learning of subspace representations. *Pattern Recognition*, 2007, **40**(5): 1556–1569
- Park M S, Choi J Y. Theoretical analysis on feature extraction capability of class-augmented PCA. *Pattern Recognition*, 2009, **42**(11): 2353–2362
- Fisher R A. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 1936, **7**: 179–188
- Hsieh P F, Wang D S, Hsu C W. A linear feature extraction for multiclass classification problems based on class mean and covariance discriminant information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **28**(2): 223–235
- Kyperountas M, Tefas A, Pitas I. Weighted piecewise LDA for solving the small sample size problem in face verification. *IEEE Transactions on Neural Network*, 2007, **18**(2): 506–519
- Turk M, Pentland A. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 1991, **3**(1): 71–86
- Cui Guo-Qin, Gao Wen. Face recognition based on two-layer generate virtual data for SVM. *Chinese Journal of Computers*, 2005, **28**(3): 368–376
(崔国勤, 高文. 基于双层虚拟视图和支持向量的人脸识别方法. 计算机学报, 2005, **28**(3): 368–376)
- Fu Y, Yuan J S, Li Z, Huang T S, Wu Y. Query-driven locally adaptive Fisher faces and expert-model for face recognition. In: *Proceedings of the International Conference on Image Processing*. San Antonio, USA: IEEE, 2007. 141–144
- Ye Jian-Hua, Liu Zheng-Guang. Multi-modal face recognition based on local binary pattern and Fisherfaces. *Computer Engineering*, 2009, **35**(11): 193–195
(叶剑华, 刘正光. 基于 LBP 和 Fisherfaces 的多模态人脸识别. 计算机工程, 2009, **35**(11): 193–195)
- Zheng W S, Lai J H, Yuen P C, Li S T. Perturbation LDA: learning the difference between the class empirical mean and its expectation. *Pattern Recognition*, 2009, **42**(5): 764–779
- Fibregas J, Faundez-Zanuy M. Biometric dispersion matcher versus LDA. *Pattern Recognition*, 2009, **42**(9): 1816–1823
- Wang X G, Tang X O. A unified framework for subspace face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(9): 1222–1228
- Xiong H L, Swamy M N S, Ahmad O M. Two-dimensional FLD for face recognition. *Pattern Recognition*, 2005, **38**(7): 1121–1124
- Peng J, Zhang P, Riedel N. Discriminant learning analysis. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2008, **38**(6): 1614–1625
- Song Feng-Xi, Zhang Da-Peng, Yang Jing-Yu, Gao Xiu-Mei. Adaptive classification algorithm based on maximum scatter difference discriminant criterion. *Acta Automatica Sinica*, 2006, **32**(4): 541–549
(宋枫溪, 张大鹏, 杨静宇, 高秀梅. 基于最大散度差鉴别准则的自适应分类算法. 自动化学报, 2006, **32**(4): 541–549)
- Li H F, Jiang T, Zhang K S. Efficient and robust feature extraction by maximum margin criterion. *IEEE Transactions on Neural Networks*, 2006, **17**(1): 157–165
- Zheng W M, Zou C R, Zhao L. Weighted maximum margin discriminant analysis with kernels. *Neurocomputing*, 2005, **67**: 357–362
- Liu J, Chen S C, Tan X Y, Zhang D Q. Comments on “efficient and robust feature extraction by maximum margin criterion”. *IEEE Transactions on Neural Networks*, 2007, **18**(6): 1862–1864
- Schraudolph N N, Gunter S, Vishwanathan S V N. Fast iterative kernel PCA. In: *Proceedings of the Conference on Neural Information Processing Systems*. British Columbia, Canada: MIT Press, 2006. 1225–1232
- Li J, Li X L, Tao D C. KPCA for semantic object extraction in images. *Pattern Recognition*, 2008, **41**(10): 3244–3250
- Chen B, Liu H W, Bao Z. A kernel optimization method based on the localized kernel Fisher criterion. *Pattern Recognition*, 2008, **41**(3): 1098–1109
- Tenenbaum J B, de Silva V, Langford J C. A global geometric framework for nonlinear dimensionality reduction. *Science*, 2000, **290**(5500): 2319–2323
- Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 2000, **290**(5500): 2323–2326
- Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 2003, **15**(6): 1373–1396

- 28 He X F, Niyogi P. Locality preserving projections. In: Proceedings of the 16th Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2003. 153–160
- 29 Wang H X, Chen S B, Hu Z L, Zheng W M. Locality-preserved maximum information projection. *IEEE Transactions on Neural Networks*, 2008, **19**(4): 571–585
- 30 Yang W K, Wang J G, Ren M W, Yang J Y, Zhang L, Liu G H. Feature extraction based on Laplacian bidirectional maximum margin criterion. *Pattern Recognition*, 2009, **42**(11): 2327–2334
- 31 Zhao D L, Lin Z C, Xiao R, Tang X O. Linear Laplacian discrimination for feature extraction. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Minnesota, USA: IEEE, 2007. 1–7
- 32 Zhang W, Lin Z C, Tang X O. Tensor linear Laplacian discrimination (TLLD) for feature extraction. *Pattern Recognition*, 2009, **42**(9): 1941–1948
- 33 Zheng W S, Lai J H, Li S Z. 1D-LDA vs. 2D-LDA: when is vector-based linear discriminant analysis better than matrix-based? *Pattern Recognition*, 2008, **41**(7): 2156–2172
- 34 Zhao D L, Lin Z C, Xiao R, Tang X O. Laplacian PCA and its applications. In: Proceedings of the 11th International Conference on Computer Vision. Rio de Janeiro, Brazil: IEEE, 2007. 1–8
- 35 Zhao D L, Lin Z C, Xiao R, Tang X O. Contextual distance for data perception. In: Proceedings of the 11th International Conference on Computer Vision. Rio de Janeiro, Brazil: IEEE, 2007. 1–8
- 36 Ma Y, Derksen H, Hong W, Wright J. Segmentation of multivariate mixed data via lossy data coding and compression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, **29**(9): 1546–1562
- 37 Blake C L, Merz C J. UCI repository of machine learning databases [Online], available: <http://www.ics.uci.edu/mllearn/MLRepository.html>, Mar 17, 2009
- 38 Wang Jue, Zhou Zhi-Hua, Zhou Ao-Ying. *Machine Learning and Its Application*. Beijing: Tsinghua Press, 2006. 135–164 (王珏, 周志华, 周傲英. 机器学习及其应用. 北京: 清华大学出版社, 2006. 135–164)
- 39 Chen S C, Zhu Y L, Zhang D Q, Yang J Y. Feature extraction approaches based on matrix pattern: MatPCA and MatFLDA. *Pattern Recognition Letters*, 2005, **26**(8): 1157–1167
- 40 Wang Z, Chen S C. New least squares support vector machines based on matrix patterns. *Neural Processing Letters*, 2007, **26**(1): 41–56
- 41 Gao Jun, Wang Shi-Tong. Matrix pattern based minimum within-class scatter support vector machines. *Acta Electronica Sinica*, 2009, **37**(5): 1051–1057 (皋军, 王士同. 基于矩阵模式的最小类内散度支持向量机. 电子学报, 2009, **37**(5): 1051–1057)

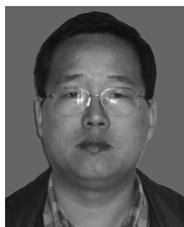


皋 军 盐城工学院信息工程学院副教授, 江南大学物联网学院博士研究生. 主要研究方向为人工智能, 模式识别, 数据挖掘和模糊系统. 本文通信作者.

E-mail: gjxllin@yahoo.cn

(**GAO Jun** Associate professor at the School of Information Engineering, Yancheng Institute of Technology,

Ph. D. candidate at the School of Internet of Things, Jiangnan University. His research interest covers artificial intelligence, pattern recognition, data mining, and fuzzy system. Corresponding author of this paper.)

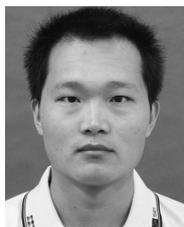


王士同 江南大学物联网学院教授. 主要研究方向为人工智能, 模式识别, 数据挖掘, 神经网络, 模糊系统, 医学图像处理 and 生物信息学.

E-mail: wxotong@yahoo.com.cn

(**WANG Shi-Tong** Professor at the School of Internet of Things, Jiangnan University. His research interest covers

artificial intelligence, pattern recognition, data mining, neural networks, fuzzy system, medical image processing, and bioinformation.)



王晓明 西华大学数学与计算机学院讲师. 主要研究方向为人工智能和模式识别. E-mail:

wangxiaomingtian@yahoo.com.cn

(**WANG Xiao-Ming** Lecturer at the School of Mathematics and Computer Engineering, Xihua University. His research interest covers artificial intelligence and pattern recognition.)