

具有模糊聚类功能的双向二维无监督特征提取方法

皋军^{1,2,3} 孙长银¹ 王士同⁴

摘要 依据最大间距判别准则 (Maximum margin criterion, MMC) 的基本原理, 并结合模糊技术和张量理论, 提出一种矩阵模式的模糊最大间距判别准则 (Matrix model fuzzy maximum margin criterion, MFMMC), 并在此基础上形成具有模糊聚类功能的双向二维无监督特征提取方法 (Two-directional two-dimensional unsupervised feature extraction method with fuzzy clustering ability, (2D)²UFFCA). 该方法不但能直接实现矩阵模式数据的模糊聚类, 而且还可以对矩阵模式数据进行双向二维特征提取, 实现特征降维. 同时我们还从几何的直观含义出发, 合理地设定矩阵模式的模糊最大间距判别准则中的调节参数 γ , 并从理论上证明其合理性. 为了提高特征提取的效率, 还提出一种能有效计算矩阵模式数据的投影变换矩阵的方法. 实验结果表明该方法具有上述优势.

关键词 张量模式, 双向二维特征提取, 矩阵模式的模糊最大间距判别准则, 模糊聚类

引用格式 皋军, 孙长银, 王士同. 具有模糊聚类功能的双向二维无监督特征提取方法. 自动化学报, 2012, 38(4): 549–562

DOI 10.3724/SP.J.1004.2012.00549

(2D)²UFFCA: Two-directional Two-dimensional Unsupervised Feature Extraction Method with Fuzzy Clustering Ability

GAO Jun^{1,2,3} SUN Chang-Yin¹ WANG Shi-Tong⁴

Abstract In this paper, based on the principles of the maximum margin criterion (MMC) and by introducing the fuzzy method and the tensor theory into it, a novel matrix model fuzzy maximum margin criterion (MFMMC) is proposed. Also, on the basis of it, a two-directional two-dimensional unsupervised feature extraction method with fuzzy clustering ability ((2D)²UFFCA) is constructed. This method can directly realize fuzzy clustering of matrix model data. And it can also achieve the two-directional two-dimensional feature extraction of them, that is, the realization of dimension reduction. At the same time, the adjusting parameter γ in the matrix model fuzzy maximum margin criterion is defined reasonably from the respect of geometry intuition, which is proved theoretically. In order to improve the efficiency of feature extraction, an effective method which can find out the projection matrices of matrix model data is presented. The results of tests show the above advantages of the method.

Key words Tensor model, two-directional two-dimensional feature extraction, matrix model fuzzy maximum margin criterion (MFMMC), fuzzy clustering

Citation Gao Jun, Sun Chang-Yin, Wang Shi-Tong. (2D)²UFFCA: two-directional two-dimensional unsupervised feature. *Acta Automatica Sinica*, 2012, 38(4): 549–562

收稿日期 2011-04-25 录用日期 2011-09-14
Manuscript received April 25, 2011; accepted September 14, 2011

国家自然科学基金 (90820002, 60903100, 61005008), 江苏省自然科学基金 (BK2011417), 江苏省新型环保重点实验室开放课题 (AE201068), 江苏省计算机信息处理重点实验室开放课题 (KJS1126) 资助

Supported by National Natural Science Foundation of China (90820002, 60903100, 61005008), Natural Science Foundation of Jiangsu Province (BK2011417), Open Project of Key Laboratory for Advanced Technology in Environmental Project of Jiangsu Province (AE201068), Open Project of Key Laboratory for Computer Information Processing Technology of Jiangsu Province (KJS1126)

本文责任编辑 刘成林

Recommended by Associate Editor LIU Cheng-Lin

1. 东南大学自动化学院 南京 210096 2. 盐城工学院信息工程学院 盐城 224001 3. 苏州大学江苏省计算机信息处理重点实验室 苏州 215006 4. 江南大学数字媒体学院 无锡 214122

1. School of Automation, Southeast University, Nanjing 210096
2. School of Information Engineering, Yancheng Institute of Technology, Yancheng 224001 3. Key Laboratory for Com-

puter Information Processing Technology, Soochow University, Suzhou 215006 4. School of Digital Media, Jiangnan University, Wuxi 214122

特征提取 (Feature extraction) 作为一种有效的特征降维方法在模式识别、图像处理和计算机视觉中得到了成功的运用^[1-4]. 特征提取的目的是通过相应的特征变换映射将高维的原始样本空间转换为低维的表示空间^[5], 而低维的表示空间所保留的判别信息的多少往往用来作为评判特征提取方法效率的标准. 主成分分析 (Principal component analysis, PCA)^[6-7] 和线性判别分析 (Linear discrimination analysis, LDA)^[8-9] 作为两种较为经典的特征提取方法被广泛地研究. PCA 是一种无监督的特征提取方法, 该方法主要通过所构造的协方差矩阵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

puter Information Processing Technology, Soochow University, Suzhou 215006 4. School of Digital Media, Jiangnan University, Wuxi 214122

的低维投影. LDA 方法则在充分使用训练样本类标号的前提下, 依据最小化类内散度、最大化类间散度的原则得到高维原始空间的低维嵌入, 实现有监督的特征提取, 从而使 LDA 方法比 PCA 方法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问题数据时类内散度矩阵容易发生奇异, 即所谓的小样本问题 (Small sample size problem, SSS). 为此研究者从不同的角度提出了许多基于 LDA 的改进方法来克服这一问题, 比如 PCA + LDA^[10]、Fisher 脸方法^[3-4]、DLA (Discriminant learning analysis)^[11]等. 特别是一些基于最大间距原则的方法: 最大散度差判别分析 (Maximum scatter difference discriminant criterion, MSD)^[12]、最大间距判别分析 (Maximum margin criterion, MMC)^[13-14]等, 该类方法通过将 LDA 方法的最大化类间散度与类内散度的比值形式转变成最大化类间散度与类内散度的差的形式, 从而较为容易地解决了 LDA 方法面临的小样本问题, 而且与其他方法相比具有相对较低的时间复杂度和更明确的直观含义^[15].

目前, 智能识别领域越来越多的数据呈现张量模式^[16-19], 如果还使用传统的子空间学习方法处理这些具有高维特征的张量模式数据, 即将高维的张量模式数据转化为矢量模式加以处理, 在一定程度上会导致所谓的维数灾难 (Curse of dimensionality)^[20], 同时会破坏原始数据内在的空间结构和相关性, 因此近来基于张量模式数据的子空间学习方法得到广泛研究, 特别是基于矩阵模式的张量子空间学习方法被研究得更为深入和详尽. 这是因为基于矩阵模式数据的张量子空间学习方法的数学理论完备、模型简单, 而且所得结果可以平滑地推广到高序张量 (High-order tensors) 模式数据上. 如具有单向特征提取功能的二维主成分分析 (Two-dimensional PCA, 2DPCA)^[21]、二维线性判别分析 (Two-dimensional LDA, 2DLDA)^[22-24] 和二维最大散度差判别准则 (Two-dimensional MSD, 2DMSD)^[25], 以及具有双向特征提取功能的双向二维主成分分析 (Two-directional Two-dimensional PCA, (2D)²PCA)^[26] 和双向二维线性判别分析 (Two-directional Two-dimensional LDA, (2D)²LDA^[27])¹ 等方法都得到了广泛研究.

值得一提的是, 近来文献 [28-31] 分别从不同角度将模糊理论引入特征提取方法中. 文献

[28-29] 分别基于 MMC 和 LDA 提出两种无监督的特征提取方法: FMSDC 和 FLDC (LDA-based fuzzy clustering), 上述方法不但能实现模糊聚类, 还可以实现特征提取; 文献 [30-31] 将模糊理论分别引入 LDA 和 2DLDA, 提出 CFLDA (Complete fuzzy LDA) 和 F2DLDA (Fuzzy 2DLDA), 提高了相应监督特征提取方法的泛化能力.

然而, 研究发现, 虽然 (2D)²PCA, (2D)²LDA 和 2DMSD + PCA 方法能取得较好的特征提取效果, 但并不能直接进行模式分类; 而 FMSDC 和 FLDC 方法也只能处理相对低维的矢量模式数据; CFLDA 和 F2DLDA 方法尽管通过引入了模糊概念, 使得有监督特征提取结果更符合实际, 但由于在求解隶属度时都使用了先验方法, 导致上述方法也不能直接进行模式分类. 因此, 本文在继承上述方法优点的基础上, 结合 MMC 方法提出一种矩阵模式的模糊最大间距判别准则 (Matrix model fuzzy maximum margin criterion, MFMMC), 并依据该准则形成具有模糊聚类功能的双向二维的无监督特征提取方法 (Two-directional two-dimensional unsupervised feature extraction method with fuzzy clustering ability, (2D)²UFFCA), 贡献如下: 1) 首次提出矩阵模式的模糊最大散度差判别准则; 2) 不但可以实现矩阵模式数据的直接聚类, 而且还可以实现双向二维的无监督特征提取; 3) 根据几何的直观含义确定 MFMMC 方法中调节参数 γ , 并从理论上证明其合理性; 4) 为了提高该方法特征提取效率, 提出一种能有效计算矩阵模式数据的正交投影变换矩阵的方法.

本文内容组织如下: 第 1 节简要介绍 MMC 和 (2D)²LDA 的相关知识; 第 2 节详细讨论 (2D)²UFFCA 方法; 第 3 节进行实验研究; 第 4 节总结全文并提出待解决的问题.

1 相关知识

为了便于讨论, 本节简要介绍最大间距判别分析方法 (MMC) 和双向二维的线性判别分析方法 ((2D)²LDA).

1.1 最大间距判别分析方法: MMC

定义 1^[8]. 假设有 n 个样本组成的样本集 $D = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$, 它们分别属于 C 个不同的类, 其中大小为 n_k 的样本子集 D_k 属于第 k 类, 给定分类决策平面的法向量 $\boldsymbol{\omega}$, 则类内散度、类间散度分别为 $\boldsymbol{\omega}^T S_W \boldsymbol{\omega}$, $\boldsymbol{\omega}^T S_B \boldsymbol{\omega}$. 其中类内散度矩阵 S_W 、第 k 类均值 \mathbf{u}_k 、类间散度矩阵 S_B 和样本总体均值 \mathbf{u} 分别

¹尽管该文献也将方法记为 2DLDA, 但本质上该方法是双向二维的特征提取方法, 为了区别于其他单向二维特征提取方法, 本文将该方法记为 (2D)²LDA.

为

$$S_W = \sum_{k=1}^C \sum_{\mathbf{x} \in D_k} (\mathbf{x} - \mathbf{u}_k)(\mathbf{x} - \mathbf{u}_k)^T \quad (1)$$

$$\mathbf{u}_k = \frac{1}{n_k} \sum_{\mathbf{x} \in D_k} \mathbf{x}, k = 1, 2, \dots, C \quad (2)$$

$$S_B = \sum_{k=1}^C n_k (\mathbf{u} - \mathbf{u}_k)(\mathbf{u} - \mathbf{u}_k)^T \quad (3)$$

$$\mathbf{u} = \frac{1}{n} \sum_{\mathbf{x} \in D} \mathbf{x} \quad (4)$$

定义 2^[13]. 根据定义 1, 最大间距判别分析方法目标函数为

$$\arg \max_{\omega^T \omega = 1} J(\omega) = \arg \max_{\omega^T \omega = 1} (\omega^T S_B \omega - \gamma \omega^T S_W \omega) \quad (5)$$

定义 2 表明 MMC 方法确实可以解决小样本问题, 且明显满足式 (5) 的最优鉴别矢量 ω^* 为 $S_B - \gamma S_W$ 最大特征值对应的单位特征向量. 定义 2 还表明, 调节参数 γ 对求解最优鉴别矢量 ω^* 有至关重要的作用, 文献 [12, 14] 虽然对该参数做了一定的理论分析, 但并没有给出设定该参数的具体方法.

1.2 双向二维线性判别分析: (2D)²LDA

假设存在矩阵模式 (Matrix model) 数据集 $\tilde{D} = \{X_1, \dots, X_n\}, \forall X_i \in \mathbf{R}^{m_1 \times m_2}$, 不同于 2DLDA 方法, (2D)²LDA 方法对于任意的一个矩阵模式数据 X_i 同时存在列方向上的投影矩阵 $U \in \mathbf{R}^{m_1 \times l_1}$ 和行方向上的投影矩阵 $V \in \mathbf{R}^{m_2 \times l_2}$, 使得 X_i 映射为 $Y_i = U^T X_i V$, 其中 $l_1 < m_1, l_2 < m_2$, 由此可得数据集 \tilde{D} 的嵌入集 $\tilde{S} = \{Y_1, \dots, Y_n\}$. 同时令数据集 $\tilde{D}(\tilde{S})$ 可分为 C 个不同的类, 其中大小为 n_k 的样本子集 $\tilde{D}_k(\tilde{S}_k)$ 属于第 k 类, $\bar{X}_k(\bar{Y}_k)$ 为第 k 类的均值, $\bar{X}(\bar{Y})$ 为 $\tilde{D}(\tilde{S})$ 的总体均值, 则 (2D)²LDA 方法的类内散度 α 和类间散度 β 分别为

$$\begin{aligned} \alpha &= \sum_{k=1}^C \sum_{i=1}^{n_k} \|Y_i - \bar{Y}_k\|_F^2 = \\ &= \sum_{k=1}^C \sum_{i=1}^{n_k} \text{tr}((Y_i - \bar{Y}_k)(Y_i - \bar{Y}_k)^T) = \\ &= \sum_{k=1}^C \sum_{i=1}^{n_k} \text{tr}(U^T (X_i - \bar{X}_k) V V^T (X_i - \bar{X}_k)^T U) \end{aligned} \quad (6)$$

$$\begin{aligned} \beta &= \sum_{k=1}^C \|\bar{Y}_k - \bar{Y}\|_F^2 = \\ &= \sum_{k=1}^C \text{tr}((\bar{Y}_k - \bar{Y})(\bar{Y}_k - \bar{Y})^T) = \\ &= \sum_{k=1}^C \text{tr}(U^T (\bar{X}_k - \bar{X}) V V^T (\bar{X}_k - \bar{X})^T U) \end{aligned} \quad (7)$$

其中, $\text{tr}(\cdot)$ 是矩阵迹运算.

定义 3^[27]. 根据式 (6) 和式 (7) 定义的类内散度 α 和类间散度 β , (2D)²LDA 判别分析方法目标函数为

$$\arg \max_{U^T U = I_{l_1}, V^T V = I_{l_2}} J(U, V) = \arg \max_{U^T U = I_{l_1}, V^T V = I_{l_2}} \frac{\beta}{\alpha} \quad (8)$$

为了有效地求解式 (8) 中两个投影矩阵 U 和 V , 文献 [27] 使用了迭代方式优化方法. 本文为了在一定程度上提高算法效率, 提出一种新的方法求解类似问题.

2 具有模糊聚类功能的双向二维无监督特征提取方法: (2D)²UFFCA

目前, 大多数基于矩阵模式的特征提取方法仅能实现样本的特征降维功能, 而不能直接完成模式的分类, 因此设计一种既能实现矩阵模式数据的特征提取又能解决该模式数据的直接分类问题的方法具有一定的研究价值和现实意义. 本文基于此提出了 (2D)²UFFCA 方法.

2.1 矩阵模式的模糊最大间距判别准则: MFMMC

定义 4. 假设存在可以分为 C 类的矩阵模式数据集 $\tilde{D} = \{X_1, \dots, X_n\} \subseteq \mathbf{R}^{m_1 \times m_2}$, 对 $\forall X_i$ 同时存在列方向和行方向上的两个投影矩阵 $U \in \mathbf{R}^{m_1 \times l_1}$ 和 $V \in \mathbf{R}^{m_2 \times l_2}$, 使 X_i 映射为 $Y_i = U^T X_i V$, 其中 $l_1 < m_1, l_2 < m_2$, 由此可得数据集 \tilde{D} 的嵌入集 $\tilde{S} = \{Y_1, \dots, Y_n\}$. 则 MFMMC 方法的类内散度 α^{MFMMC} 和类间散度 β^{MFMMC} 可以分别表示为

$$\begin{aligned} \alpha^{\text{MFMMC}} &= \sum_{k=1}^C \sum_{i=1}^n \mu_{ik}^m \|Y_i - \bar{Y}_k\|_F^2 = \\ &= \sum_{k=1}^C \sum_{i=1}^n \mu_{ik}^m \text{tr}((Y_i - \bar{Y}_k)(Y_i - \bar{Y}_k)^T) = \end{aligned}$$

$$\begin{aligned} & \text{tr}(U^T (\sum_{k=1}^C \sum_{i=1}^n \mu_{ik}^m (X_i - \bar{X}_k) V V^T \times \\ & (X_i - \bar{X}_k)^T) U) = \\ & \text{tr}(V^T (\sum_{k=1}^C \sum_{i=1}^n \mu_{ik}^m (X_i - \bar{X}_k)^T U U^T \times \\ & (X_i - \bar{X}_k)) V) \end{aligned} \quad (9)$$

$$\begin{aligned} \beta^{\text{MFMMC}} &= \sum_{k=1}^C \sum_{i=1}^n \mu_{ik}^m \|\bar{Y}_k - \bar{Y}\|_F^2 = \\ & \sum_{k=1}^C \sum_{i=1}^n \mu_{ik}^m \text{tr}((\bar{Y}_k - \bar{Y})(\bar{Y}_k - \bar{Y})^T) = \\ & \text{tr}(U^T (\sum_{k=1}^C \sum_{i=1}^n \mu_{ik}^m (\bar{X}_k - \bar{X}) V V^T \times \\ & (\bar{X}_k - \bar{X})^T) U) = \\ & \text{tr}(V^T (\sum_{k=1}^C \sum_{i=1}^n \mu_{ik}^m (\bar{X}_k - \bar{X})^T U U^T \times \\ & (\bar{X}_k - \bar{X})) V) \end{aligned} \quad (10)$$

其中, μ_{ik} 是模糊隶属度, 表示第 i 个样本隶属于第 k 类可能性大小, 且满足 $\sum_{k=1}^C \mu_{ik} = 1$; m 为模糊常数, 一般设定 $m > 1$; $\bar{X}_k(\bar{Y}_k)$ 表示 $\tilde{D}(\tilde{S})$ 第 k 类的模糊聚类中心, 而 $\bar{X}(\bar{Y})$ 表示 $\tilde{D}(\tilde{S})$ 总体均值, 即 $\bar{X} = \frac{1}{n} \sum_{X_i \in \tilde{D}} X_i$ ($\bar{Y} = \frac{1}{n} \sum_{Y_i \in \tilde{S}} Y_i$).

定义 5. 根据定义 4 中式 (9) 和式 (10) 可得矩阵模式的模糊最大间距判别分析的目标函数为

$$\begin{aligned} & \arg \max_{U^T U = I_{l_1}, V^T V = I_{l_2}} J(U, V) = \\ & \arg \max_{U^T U = I_{l_1}, V^T V = I_{l_2}} \beta^{\text{MFMMC}} - \gamma \alpha^{\text{MFMMC}} \end{aligned} \quad (11)$$

定义 4 和定义 5 表明, MFMMC 方法与 (2D)²LDA 中的类内散度和类间散度存在着明显的差异: 1) MFMMC 方法中的类内散度和类间散度均基于模糊理论; 2) MFMMC 方法中的 $\bar{X}_k(\bar{Y}_k)$ 表示 $\tilde{D}(\tilde{S})$ 第 k 类的模糊聚类中心, 而不是 $\tilde{D}(\tilde{S})$ 第 k 类的均值, 且上述两个分量是基于无监督方法得到的, 而不是基于有监督的方法.

2.2 实现矩阵模式数据的双向特征提取

为了一定程度上有效求解式 (11), 可以采用类

似于文献 [27] 中提出的迭代方法进行求解. 然而通过分析可知, 上述迭代方法所得的迭代序列一般并不能保证收敛, 即所得的两个投影矩阵的解只是局部最优解的近似解^[32]. 尽管文献 [32] 提出了一种新颖的迭代方法并证明了该方法产生的迭代序列是收敛的, 可以得到局部最优解, 从理论上较好地解决了通过迭代优化方法实现张量模式数据特征提取的一个基本问题. 然而, 由于该求解过程考虑了投影矩阵 U 和 V 的非线性关系, 因此就必须使用非线性优化方法才能有效求解^[33], 这就导致在实际计算过程中会不断引入一些舍入误差, 从而在一定程度上不能保证迭代序列局部收敛, 得到的往往还是局部最优解的近似解^[33].

本文根据矩阵理论和文献 [33] 的基本思想提出一种能有效求解式 (11) 的近似方法. 该方法通过假设 U 和 V 是正交矩阵, 并在此基础上改写 α^{MFMMC} 和 β^{MFMMC} 中的模糊散度矩阵, 最后通过直接求解矩阵特征值对应的单位特征矢量得到式 (11) 解的近似值, 这一过程不需要进行任何的迭代优化, 因此一定程度上降低了方法的时间和空间复杂度, 提高了求解投影矩阵的效率.

定理 1. 假设 U 和 V 是正交矩阵, 问题 (11) 可以通过如下的特征值分解方法得到:

假设 U 和 V 是正交矩阵, 则 α^{MFMMC} 和 β^{MFMMC} 中的模糊散度矩阵可以分别改写为

$$S_w^{V'} = \sum_{k=1}^C \sum_{i=1}^n \mu_{ik}^m ((X_i - \bar{X}_k)(X_i - \bar{X}_k)^T) \quad (12)$$

$$S_b^{V'} = \sum_{k=1}^C \sum_{i=1}^n \mu_{ik}^m ((\bar{X}_k - \bar{X})(\bar{X}_k - \bar{X})^T) \quad (13)$$

$$S_w^{U'} = \sum_{k=1}^C \sum_{i=1}^n \mu_{ik}^m ((X_i - \bar{X}_k)^T (X_i - \bar{X}_k)) \quad (14)$$

$$S_b^{U'} = \sum_{k=1}^C \sum_{i=1}^n \mu_{ik}^m ((\bar{X}_k - \bar{X})^T (\bar{X}_k - \bar{X})) \quad (15)$$

如果假设 \mathbf{v}_i 表示 U 或 V 的第 i 个列向量, 则上述问题可以转变为

$$\lambda_1 = \arg \max_{\mathbf{v}^T \mathbf{v} = 1} \mathbf{v}^T (S_b' - \gamma S_w') \mathbf{v} \quad (16)$$

$$\lambda_k = \arg \max_{\mathbf{v}_k^T \mathbf{v}_k = 1, \mathbf{v}_k^T \mathbf{v}_i = 0 (i=1, \dots, k-1)} \mathbf{v}^T (S_b' - \gamma S_w') \mathbf{v} \quad (17)$$

从式 (16) 看出, \mathbf{v}_1 是 $S_b' - \gamma S_w'$ 的最大特征值对应的特征向量. 而 \mathbf{v}_k 则通过式 (17) 并结合拉格

朗日乘数法求解:

$$L_k = \mathbf{v}_k^T (S'_b - \gamma S'_w) \mathbf{v}_k - \lambda (\mathbf{v}_k^T \mathbf{v}_k - 1) - \sum_{q=1}^{k-1} \lambda_q \mathbf{v}_k^T \mathbf{v}_q \quad (18)$$

根据局部最优解的必要条件 $\frac{\partial L_k}{\partial \mathbf{v}_k} = 0$, 则得:

$$2(S'_b - \gamma S'_w) \mathbf{v}_k - 2\lambda \mathbf{v}_k - \sum_{q=1}^{k-1} \lambda_q \mathbf{v}_q = 0 \quad (19)$$

在式 (19) 的两边同时左乘 \mathbf{v}_k^T , 则 $\lambda = \lambda_k = \mathbf{v}_k^T (S'_b - \gamma S'_w) \mathbf{v}_k$. 同时令: $\tilde{\lambda}_{k-1} = [\lambda_1, \dots, \lambda_{k-1}]^T$, $Q_{k-1} = [\mathbf{v}_1, \dots, \mathbf{v}_{k-1}]$, 并用 $\mathbf{v}_j^T (j = 1, \dots, k-1)$ 左乘式 (19), 则

$$\tilde{\lambda}_{k-1} = 2Q_{k-1}^T (S'_b - \gamma S'_w) \mathbf{v}_k \quad (20)$$

将式 (20) 代入式 (19), 得:

$$(I - Q_{k-1} Q_{k-1}^T) (S'_b - \gamma S'_w) \mathbf{v}_k = \lambda_k \mathbf{v}_k \quad (21)$$

根据式 (21) 可知 \mathbf{v}_k 为 $(I - Q_{k-1} Q_{k-1}^T) (S'_b - \gamma S'_w)$ 最大特征值对应的单位特征向量.

根据定理 1 可得求解满足式 (11) 的近似解 U 和 V 的算法.

算法 1. 求解满足式 (11) 的 U 和 V 近似值的算法

步骤 1. 根据定理 1 的假设前提重新构造式 (12) ~ (15): $S_w^{V'}$, $S_b^{V'}$, $S_w^{U'}$ 和 $S_b^{U'}$;

步骤 2. 计算 $U = [\mathbf{u}_1, \dots, \mathbf{u}_{l_1}]$, $V = [\mathbf{v}_1, \dots, \mathbf{v}_{l_2}]$:

1) 设定 $l_1(l_2)$ 的初值;

2) 计算 $S_w^{V'} - \gamma S_b^{V'}$, $S_w^{U'} - \gamma S_b^{U'}$ 最大特征值对应的单位特征向量 $\mathbf{u}_1, \mathbf{v}_1$;

3) 计算 $(I - Q_{k-1}^U Q_{k-1}^{U'}) (S_b^{V'} - \gamma S_w^{V'})$, $(I - Q_{k-1}^V Q_{k-1}^{V'}) (S_b^{U'} - \gamma S_w^{U'})$ 最大特征值对应的单位特征向量 $\mathbf{u}_k (k = 2, \dots, l_1)$, $\mathbf{v}_k (k = 2, \dots, l_2)$; 其中 $Q_{k-1}^U = [\mathbf{u}_1, \dots, \mathbf{u}_{k-1}]$, $Q_{k-1}^V = [\mathbf{v}_1, \dots, \mathbf{v}_{k-1}]$.

对于算法 1 需要说明两点: 1) 算法 1 中假设 U 和 V 是正交矩阵是为了简化上述四种模糊散度矩阵, 如此处理虽然一定程度上降低了算法求解式 (11) 的精度, 但这样做是有一定依据的. 通过传统迭代优化方法^[27] 求解式 (11) 时, 在求解某个投影矩阵时也是假定另一个投影矩阵为常量, 因此从这一层面上讲, 本文方法即是使用单位矩阵 I 替换矩阵 UU^T , VV^T ; 2) 通过实际运行算法 1 后得到的解应该为定义 4 中的形式, 即 $U \in \mathbf{R}^{m_1 \times l_1}$ 和 $V \in \mathbf{R}^{m_2 \times l_2}$, 而不是假设前提中的完整的正交矩阵.

2.3 实现矩阵模式数据的模糊聚类

传统的模糊聚类方法一般依据模糊隶属度实现

模式数据的无监督分类, 本文也依据这一原理并结合 Lagrange 乘数法来求解满足式 (11) 的模糊隶属度.

根据 Lagrange 乘数法, 求解式 (11) 对应的 Lagrange 公式为

$$L = J(U, V) - \eta_1 (U^T U - I_{l_1}) - \eta_2 (V^T V - I_{l_2}) + \sum_{i=1}^n \lambda_i \left(\sum_{k=1}^C \mu_{ik} - 1 \right) \quad (22)$$

将式 (22) 对 μ_{ik} 求偏导数并令其为 0, 得:

$$\mu_{ik} = \left(\frac{\lambda_i}{m(\gamma \|Y_i - \bar{Y}_k\|_F^2) - \|\bar{Y}_k - \bar{Y}\|_F^2} \right)^{\frac{1}{m-1}} \quad (23)$$

又因为 $\sum_{j=1}^C \mu_{ij} = 1$, 所以有

$$\lambda_i^{\frac{1}{m-1}} = \frac{1}{\sum_{j=1}^C (\gamma \|Y_i - \bar{Y}_j\|_F^2 - \|\bar{Y}_j - \bar{Y}\|_F^2)^{\frac{1}{m-1}}} \quad (24)$$

将式 (24) 代入式 (23), 可得式 (25) (见下页).

同理可得 \tilde{S} 第 k 类的模糊聚类中心:

$$\bar{Y}_k = \frac{\sum_{i=1}^n \mu_{ik}^m \left(Y_i - \frac{1}{\gamma} \bar{Y} \right)}{\sum_{i=1}^n \mu_{ik} \left(1 - \frac{1}{\gamma} \right)} \quad (26)$$

一般要求模糊隶属度满足 $\mu_{ik} \in [0, 1]$, 且由于模糊常数 m 的选择都要求式 (25) 的分子内部表达式大于 0, 然而当某样本 X_i 经过双向二维特征提取后, 到第 k 类聚类中心的距离小于或等于第 k 类聚类中心到总体均值的距离的 $1/\sqrt{\gamma}$ 倍 (即满足式 (24)) 时, 将使得式 (25) 分子内部表达式为小于或等于 0 的数, 从而导致无法使用式 (25) 求解样本 X_i 相对于第 k 类的模糊隶属度, 因此有必要通过其他方法来解决这一问题.

$$\text{tr}(U^T (X_i - \bar{X}_k) V V^T (X_i - \bar{X}_k)^T U) \leq \frac{1}{\gamma} \text{tr}(U^T (\bar{X}_k - \bar{X}) V V^T (\bar{X}_k - \bar{X})^T U) \quad (27)$$

通过研究满足式 (27) 的样本所具有的直观几何含义后发现, 当某个样本 X_i 能满足式 (27), 那么可以认定该样本严格隶属于第 k 类, 即将该样本硬划分到第 k 类, 也即规定 $\mu_{ik} = 1, \mu_{ij} = 0 (k \neq j)$. 下面通过一个具体的例子 (见图 1)^[28] 说明上述规定的合理性.

$$\mu_{ik} = \frac{(\|Y_i - \bar{Y}_k\|_F^2 - \frac{1}{\gamma}\|\bar{Y}_k - \bar{Y}\|_F^2)^{\frac{1}{m-1}}}{\sum_{j=1}^C (\|Y_i - \bar{Y}_j\|_F^2 - \frac{1}{\gamma}\|\bar{Y}_j - \bar{Y}\|_F^2)^{\frac{1}{m-1}}} = \frac{(\text{tr}(U^T(X_i - \bar{X}_k)VV^T(X_i - \bar{X}_k)^T U) - \frac{1}{\gamma}\text{tr}(U^T(\bar{X}_k - \bar{X})VV^T(\bar{X}_k - \bar{X})^T U))^{\frac{1}{m-1}}}{\sum_{j=1}^C (\text{tr}(U^T(X_i - \bar{X}_j)VV^T(X_i - \bar{X}_j)^T U) - \frac{1}{\gamma}\text{tr}(U^T(\bar{X}_j - \bar{X})VV^T(\bar{X}_j - \bar{X})^T U))^{\frac{1}{m-1}}} \quad (25)$$

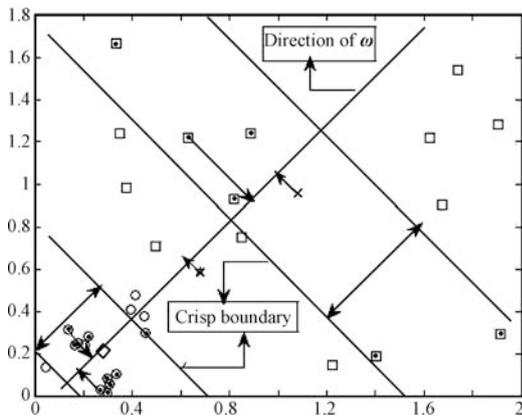


图 1 硬划分示意图

Fig. 1 Sketch map of the crisp section

其中，“○”和“□”表示两类样本，“×”、“◇”分别表示两类样本的聚类中心，“☆”表示总体样本均值， $\omega = U = V \in \mathbf{R}^{2 \times 1}$ ， $\gamma = 1$ ，“Crisp boundary”表示硬划分边界，“↔”连接的两个硬划分编辑的区域为硬划分区域。

对图 1 中加“●”的样本进行理论分析后，发现这些样本确实能满足式 (27)，因此不能使用式 (25) 计算这类样本所对应的模糊隶属度，但依据这些样本在图 1 中所具有的直观几何含义，可以将这一类型的样本进行硬划分处理，因此从这一层面上讲，上述硬划分的规定是合理的，类似的方法在文献 [34] 中也得到了成功的运用。

式 (25) 和式 (27) 也反映了参数 γ 的重要性，这是因为该参数一定程度上决定了被硬划分样本的数目，同时也决定了会不会出现某一个样本被硬划分多次的情况，因此有必要探讨该参数的设定问题。

2.3.1 调节参数 γ 的设定

式 (25) 和式 (27) 表明，参数 γ 一定程度上决定着聚类的效果，特别是在判断一个样本是否采用硬划分时尤为重要。从直观意义上看，为了避免一个

样本被多次硬划分到不同的类，一般最好要求该样本被双向投影后到某类聚类中心的距离应该小于或等于不同类别中心之间最小距离的 $1/\sqrt{N}$ ($N \geq 4$) 倍。因此，根据这一直观的含义并通过如下定理可以在一定程度上解决参数 γ 设定的问题，同时也可以避免出现同一样本被硬划分多次的情况。

定理 2. 设某样本 X_i ，第 k 类聚类中心 \bar{X}_k 和总体均值 \bar{X} 分别经过双向二维投影后，如果令 $\gamma = \max\{\gamma_1, \dots, \gamma_C\}$ ，其中 $\gamma_k = \frac{N \max_j \text{tr}(U^T(\bar{X}_j - \bar{X})VV^T(\bar{X}_j - \bar{X})^T U)}{\min_{k \neq k^*} \text{tr}(U^T(\bar{X}_{k^*} - \bar{X}_k)VV^T(\bar{X}_{k^*} - \bar{X}_k)^T U)}$ ($N \geq 4$)，则式 (27) 可以写为

$$\text{tr}(U^T(X_i - \bar{X}_k)VV^T(X_i - \bar{X}_k)^T U) \leq \frac{1}{N} \min_{k \neq k^*} \text{tr}(U^T(\bar{X}_{k^*} - \bar{X}_k)VV^T(\bar{X}_{k^*} - \bar{X}_k)^T U) \quad (28)$$

证明. 根据式 (27)，有：

$$\text{tr}(U^T(X_i - \bar{X}_k)VV^T(X_i - \bar{X}_k)^T U) \leq \frac{1}{\gamma_k} \text{tr}(U^T(\bar{X}_k - \bar{X})VV^T(\bar{X}_k - \bar{X})^T U) \quad (29)$$

根据定理的假设前提，将 γ_k 代入式 (29) 并运用 $\max_j \text{tr}(U^T(\bar{X}_j - \bar{X})VV^T(\bar{X}_j - \bar{X})^T U) \geq \text{tr}(U^T(\bar{X}_k - \bar{X})VV^T(\bar{X}_k - \bar{X})^T U)$ 的条件可得式 (28)。□

根据定理 2 可知，调节参数 γ 的设定方法具有较为明显的几何直观含义，从而一定程度上有利于提高聚类的稳定性。同时式 (28) 也表明了参数 N 对模糊聚类效果的影响，即 N 越大，该样本被硬划分的可能性越小，反之该样本被硬划分的可能性越大，从而也说明在聚类过程中， N 决定了被硬划分样本的数目。因此从 N 的极限形式来看，当 $N \rightarrow \infty$ 时，该聚类过程可以进化为绝对的模糊聚类过程；当 $N \rightarrow 0$ 时，该聚类过程退化为经典的硬划分聚类。

2.4 确定数据集 \tilde{D} 的模糊聚类中心

为了有效地实现第 2.2 节和第 2.3 节的功能，还需解决一个关键的问题：基于数据集 \tilde{D} 的第 k 类聚

类中心 \bar{X}_k 如何进行有效的计算. 该矩阵不但会影响算法 1 计算式 (11) 的近似解, 在一定程度上也会影响最终模糊聚类的效果, 因此有必要讨论该设定问题.

然而, 由于式 (11) 中同时存在双向特征提取矩阵 U 和 V , 从而导致通过求解满足式 (11) 的必要条件直接求解 \bar{X}_k 尤为困难. 而在第 2.2 节中假设 U 和 V 是正交矩阵, 这为在一定程度上有效求解 \bar{X}_k 提供了思路.

定理 3. 假设 U 和 V 是正交矩阵, 如果式 (11) 有解, 则:

$$\bar{X}_k = \frac{\sum_{i=1}^n \mu_{ik}^m (X_i - \frac{1}{\gamma} \bar{X})}{\sum_{i=1}^n \mu_{ik}^m (1 - \frac{1}{\gamma})} \quad (30)$$

证明. 根据定理假设 V 是正交矩阵并结合式 (11) 对应的 Lagrange 公式 (22), 则满足式 (11) 必须有: $\frac{\partial L}{\partial \bar{X}_k} = 0$, 则有:

$$2U^T \sum_{i=1}^n \mu_{ik}^m (\bar{X}_k - \bar{X}) U = -2\gamma U^T \sum_{i=1}^n \mu_{ik}^m (X_i - \bar{X}_k) U \quad (31)$$

又因为根据定理假设前提 U 是正交矩阵, 所以有:

$$\sum_{i=1}^n \mu_{ik}^m (\bar{X}_k - \bar{X}) = -\gamma \sum_{i=1}^n \mu_{ik}^m (X_i - \bar{X}_k) \quad (32)$$

则有:

$$\sum_{i=1}^n \mu_{ik}^m (\gamma - 1) \bar{X}_k = \sum_{i=1}^n \mu_{ik}^m (\gamma X_i - \bar{X}) \quad (33)$$

所以式 (30) 成立. \square

从定理 3 可知, 在运行算法 1 时, 如果取满足式 (30) 的 \bar{X}_k , 那么所得投影矩阵的近似值是近似程度最好的而且是最稳定的. 同时通过观察式 (26) 与式 (30) 发现, 当对式 (30) 中 \bar{X}_k 的两边分别同时作用 U^T 和 V 时, 所得到的变换矩阵正是式 (26) 中的 Y_k , 即是数据集 \tilde{D} 的第 k 类聚类中心 \bar{X}_k 通过提取后所得 Y_k 即是数据集 \tilde{S} 的第 k 类聚类中心, 这一方法在许多有监督特征提取方法^[22, 25-27] 中被广泛使用, 以便在实现特征提取的同时尽可能地保持原始输入样本内在的全局结构.

通过以上分析, 可得具有模糊聚类功能的双向二维无监督特征提取方法: (2D)²UFFCA.

算法 2. (2D)²UFFCA

步骤 1. 给定误差控制量 ε , 最大迭代次数 $maxIter$, 令 p 的初值为 0 并随机产生模糊隶属度矩阵 $\Lambda^p = (\mu_{ik}^p)_{n \times C}$, 设定参数 γ, m, N, l_1, l_2 初始值.

步骤 2. 根据式 (30) 计算初始聚类中心 X_k^p ($k = 1, \dots, C$).

步骤 3. 根据算法 1 计算 $U^p \in \mathbf{R}^{m_1 \times l_1}$ 和 $V^p \in \mathbf{R}^{m_2 \times l_2}$.

步骤 4. 根据定理 2 设定参数 γ .

步骤 5. 计算模糊隶属度矩阵 $\Lambda^{p+1} = (\mu_{ik}^{p+1})_{n \times C}$: 对于 $\forall X_i \in \tilde{D}$, 如果该样本满足式 (28), 则令 $\mu_{ik}^{p+1} = 1, \mu_{ij}^{p+1} = 0$ ($k \neq j$). 否则使用式 (25) 求解 μ_{ik}^{p+1} .

步骤 6. 根据式 (30) 计算聚类中心 \bar{X}_k^{p+1} ($k = 1, \dots, C$).

步骤 7. 如果 $|J(U, V)^{p+1} - J(U, V)^p| \leq \varepsilon$ 或 $p \geq maxIter$ 时, 输出模糊隶属度矩阵 Λ^{p+1} , 聚类中心 \bar{X}_k^{p+1} ($k = 1, \dots, C$), 根据算法 1 计算投影矩阵 $U^{p+1} \in \mathbf{R}^{m_1 \times l_1}$ 和 $V^{p+1} \in \mathbf{R}^{m_2 \times l_2}$. 否则令 $p = p + 1$, 转至步骤 3.

3 实验

通过测试相应的数据集来说明 (2D)²UFFCA 方法的基本聚类能力、处理大数据集的聚类能力和处理矩阵模式数据的无监督特征提取能力. 仿真实验运行的计算机系统为 Vista 操作系统、Inter Core2 T5500 1.66 G CPU 处理器、1.5 G 内存, 使用 Matlab 7.0 仿真软件包.

3.1 测试基本的聚类能力

UCI 数据集^[35] 和基因数据集 Yeast galactose data^[36] 常被用来测试算法的聚类效果^[28, 37]. 在本测试过程中, 通过测试 UCI 数据的 5 个数据子集: Iris 数据集、Glass 数据集、Lymphography 数据集、Zoo 数据集以及 Yeast galactose data 数据集说明 (2D)²UFFCA 方法的基本聚类能力. 为了反映本文方法的有效性, 首先需要将上述 5 个矢量表示的数据集转换成相应的矩阵模式, 如表 1 所示, 同时在测试过程中令参数 $\varepsilon = 1E - 5$, $maxIter = 100$, $\gamma = 2$, $m = 2$, $n = 4$, 并将测试精度 (Rand index^[38]) 和收敛性与 FCM 方法所得结果进行比较, 如表 2 和图 2 所示.

为了说明本文方法参数 N 对模糊划分的影响, 分别令 $N = 4, 400, 4000, 40000$ 来测试 Iris 数据集, 并用该数据集的 150 个样本隶属于第一类的模糊隶属度来加以说明, 如图 3 所示. 另一方面, 通过分别测试 Yeast galactose data 数据集不同的矩阵模式以及不同特征提取效果对精度的影响来说明本

文方法的聚类效果与具体的矩阵模式的相关性,如图 4 所示.

通过对以上测试的分析,在一定程度上可以得到如下基本结论:

1) 表 2 中两种方法的聚类精度的比较结果以及图 2 所表示的两种方法在处理 Iris 数据集的收敛性对比说明 (2D)²UFFCA 算法具有基本的聚类能力. 在处理传统矢量模式数据特别是高维基因数据时,只要选择合适的矩阵模式同时选取适当的列(行)投影矩阵 $U \in \mathbf{R}^{m_1 \times l_1}$ ($V \in \mathbf{R}^{m_2 \times l_2}$) 就可以取得较为合理的聚类效果,从而一定程度上说明本文方法在处理传统模式数据时不会因为使用矩阵模式而降低方法的聚类有效性.

图 2(a) 是本文方法取 $l_1 = 2, l_2 = 1$ 时测试 Iris 数据的收敛性表现,从图中看出,本文方法只需迭代 12 次就可以满足相应的收敛条件,从而一定程度上可以说明两个问题: a) 本文方法在实验过程中满足 $|J(U, V)^{p+1} - J(U, V)^p| \leq \varepsilon$ 时退出,这表明在使用矩阵模式处理相应数据时一定程度上可以保证迭代收敛; b) 本文方法具有较好的收敛速度.

2) 图 3 是本文方法选取不同参数 N 并结合 $l_1 = 2, l_2 = 1$ 的情况下通过测试 Iris 数据集得到的,表明参数 N 对模糊划分程度的影响. 当选取 $N = 4, 400, 4000$ 时,在 150 个样本中分别有 135, 24, 2 个样本属于硬划分(分别对应图 3(a)~3(c)),

表 1 Iris, Glass, Lymphography, Zoo, Yeast galactose data 数据集

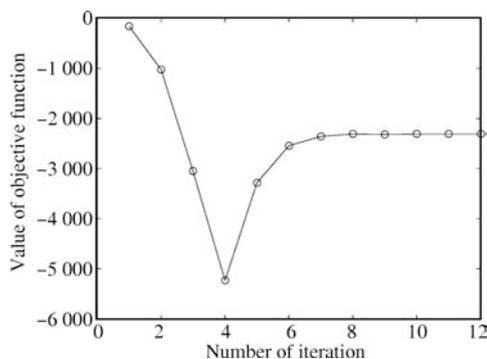
Table 1 Iris, Glass, Lymphography, Zoo, and Yeast galactose data datasets

Datasets	Number of samples	Number of features	Number of subjects	Matrix model
Iris	150	4	3	2×2
Glass	214	9	6	3×3
Lymphography	148	18	4	6×3
Zoo	101	16	5	4×4
Yeast galactose data	205	80	4	10×8 8×10

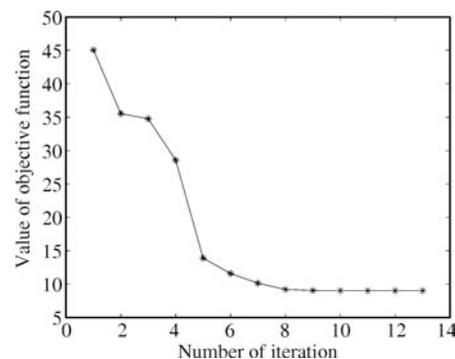
表 2 对 Iris, Glass, Lymphography, Zoo, Yeast galactose data 数据集的识别精度比较

Table 2 Comparison of recognition performances on Iris, Glass, Lymphography, Zoo, and Yeast galactose data datasets

Algorithm	Datasets				
	Iris	Glass	Lymphography	Zoo	Yeast galactose data
FCM	0.893	0.7117	0.5820	0.7992	0.85567
(2D) ² UFFCA	0.92671 ($l_1 = 2, l_2 = 1$)	0.73454 ($l_1 = 3, l_2 = 2$)	0.59469 ($l_1 = 4, l_2 = 1$)	0.8996 ($l_1 = 2, l_2 = 4$)	0.9847 (10×8) ($l_1 = 3, l_2 = 4$) 0.97972 (8×10) ($l_1 = 2, l_2 = 4$)



(a) (2D)²UFFCA 算法
(a) (2D)²UFFCA method



(b) FCM 算法
(b) FCM method

图 2 (2D)²UFFCA 和 FCM 方法收敛性比较

Fig. 2 Comparison of iterative convergence on (2D)²UFFCA and FCM methods

而当 $N = 40\ 000$ 时则没有一个样本被硬划分 (对应图 3(d)), 这充分说明本文方法中的参数 N 确实起到了控制模糊划分的作用, 从而验证定理 2 中参数 N 所具有的性质. 通过测试发现, 随着参数 N 增大, 本文方法的聚类精度也有所提高, 如当 $N = 4$ 时精度为 0.926 71, 而当 $N = 400, 4\ 000, 40\ 000$ 时精度可达 0.957 49, 这说明本文方法具有作为模糊聚类方法的基本特点, 即模糊程度越高一般越能反映样本的客观分类.

3) 图 4 是本文方法在测试基因数据集 Yeast galactose data 时分别选取 10×8 和 8×10 两种矩阵模式对应的聚类精度示意图. 从图中可以看出, 当将数据集 Yeast galactose data 转换成不同的矩阵模式时, 本文方法的测试精度有一定的差异性, 这说明其聚类效果一定程度上依赖于相应的矩阵模式. 同时还可以看出, 相对同一种矩阵模式, 不同列 (行) 方向的特征提取效果也在一定程度上影响本文方法的有效性.

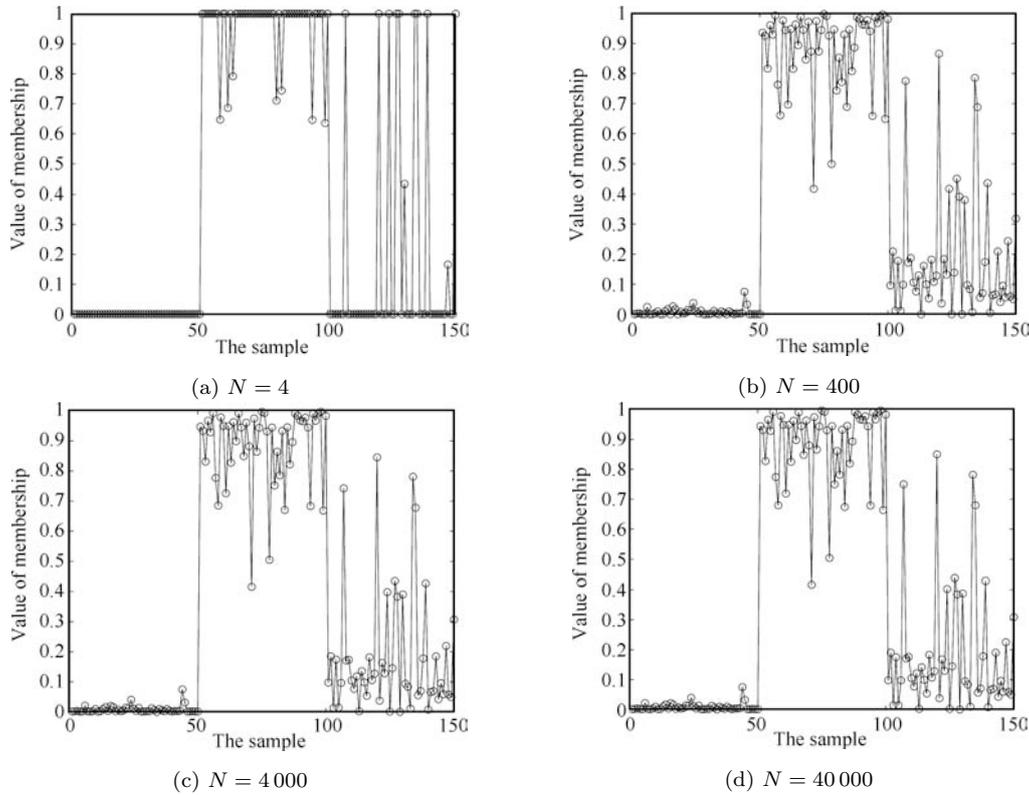


图 3 参数 N 对 $(2D)^2$ UFFCA 算法模糊划分的影响
Fig. 3 Influence of parameter N on fuzzy partition of $(2D)^2$ UFFCA

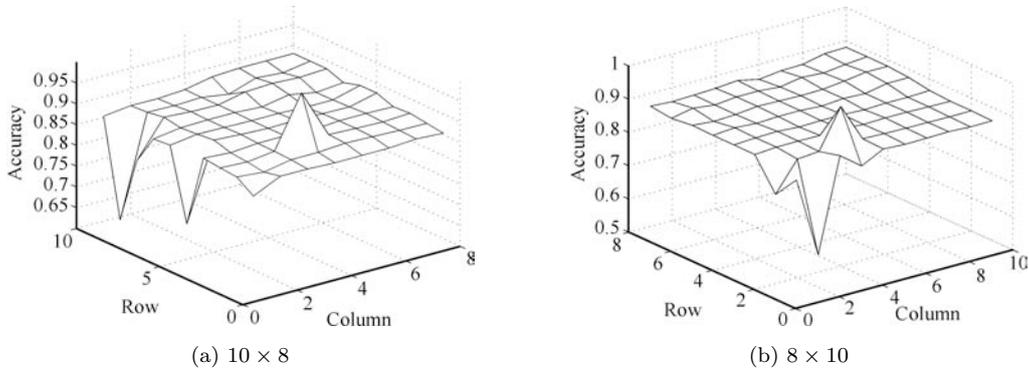


图 4 不同矩阵模式对 $(2D)^2$ UFFCA 方法的聚类效果的影响
Fig. 4 Influences of matrix models on clustering result of $(2D)^2$ UFFCA

3.2 测试大数据集的聚类效果

为了有效地说明本文方法具有较强的模糊聚类能力, 选取 1 个 UCI 数据子集^[36]— Shuttle 数据集, 以及手写体数字数据集 USPS、MNIST (<http://www.cs.uiuc.edu/homes/dengcai2/>) 的 2 个数据子集共 3 个大容量数据集来测试本文方法, 如表 3 和图 5 所示. USPS、MNIST 作为两种具有矩阵模式的图像数据, 分别有 9 298、4 000 个样本, 为了提高测试效率, 在本测试过程中在上述两种数据集中分别随机选取 2 500、2 160 个样本作为测试样本.

为了测试本文方法在处理大数据集时的聚类效果, 使用 FCM、FCS^[34] 方法与 (2D)²UFFCA 方法进行测试比较. 从散度的角度上讲, 上述三种方法具有一定的可比较性^[34]. 在测试过程中令 FCS 方法中 $\eta = 2$, $\beta = 0.5$, 本文方法中 $\gamma = 2$, 其他两种方法中共有的参数则令为 $\varepsilon = 1E-5$, $maxIter = 50$, $m = 2$. 测试精度比较见表 4.

通过对上述实验结果的分析, 可以得到如下结论:

1) 本文方法具有较强的聚类能力. 特别是在处理矢量模式的大容量 Shuttle 数据集时, 尽管将该数据转换成矩阵模式加以处理, 但并没有降低对该数据集的识别能力, 这在一定程度上说明本文方法具有处理大容量矢量模式数据的模糊聚类能力.

2) 在处理两种手写体数据集时, 本文方法与 FCM、FCS 方法相比具有更好的聚类精度. 如在测试 USPS 数据时, 本文方法可以有效聚类 2 124 个样本, 分别比 FCS 和 FCM 方法多识别 401 个和 739 个样本; 在测试 MNIST 数据时本文方法可以有效识别 1 580 个样本. 这是因为手写体图像数据

是典型的矩阵模式数据, 而本文方法在测试上述两种数据集时, 并没有将矩阵模式转换为矢量模式进行处理, 而是对矩阵模式数据直接进行模糊聚类, 较好地保持了原图像数据间的空间结构信息, 从而提高了模糊聚类能力; 同时由于本文方法在实现模糊聚类的过程中充分地使用了适合图像数据的特征提取技术, 也即本文方法能根据聚类过程中所得到的双向二维特征变换矩阵来实现原图像数据的特征降维, 使得变换后的数据显得更为简单, 从而提高模糊聚类的效果. 尽管 FCM、FCS 方法也考虑了类内散度矩阵和类间散度矩阵对模糊聚类的影响, 但显然不具备本文方法通过使用特征提取技术来实现类间最大、类内最小的这一技术.

3.3 测试特征提取能力

为了有效说明 (2D)²UFFCA 方法的无监督的特征提取能力, 在本测试过程中使用两个经典的人脸图像数据集: ORL (32×32)、Yale (32×32) (<http://www.cs.uiuc.edu/homes/>). ORL 数据集包含 40 个类别的样本, 每一类别由 10 种不同的人脸表情组成, 如图 6 所示; Yale 数据集包含 15 个类别的人脸图像, 每一类由 11 种不同表情的人脸图像组成, 如图 7 所示.

在测试过程中, 使用 4 种无监督的特征提取方法: PCA, 2DPCA, (2D)²PCA, FLDC 方法与本文的 (2D)²UFFCA 方法进行测试比较. 在此过程中分别选取上述两种人脸图像数据集中每类的前 2 个、3 个、4 个和 5 个样本作为训练样本, 剩余样本作为测试样本, 为了更好地表现文中算法 1 所具有的特征值分解的效率和 (2D)²UFFCA 方法总体特征提取的效果, 分别从测试精度 (最近邻分类器 1-NN)、特征提取维数和一次特征值分解所消耗的 CPU 时间

表 3 Shuttle, USPS, MNIST 数据集

Table 3 Shuttle, USPS, and MNIST datasets

Datasets	Number of samples	Number of features	Number of subjects	Matrix model
Shuttle	14 500	9	7	3×3
USPS	2 500	256	10	16×16
MNIST	2 160	784	10	28×28

表 4 Shuttle, USPS 和 MNIST 数据集的识别精度比较

Table 4 Comparison of recognition performances on Shuttle, USPS, and MNIST datasets

Algorithm	Datasets		
	Shuttle	USPS	MNIST
FCM	0.4267	0.5540	0.5821
FCS	0.46295	0.6890	0.5830
(2D) ² UFFCA	0.5666 ($l_1 = 3, l_2 = 1$)	0.84953 ($l_1 = 2, l_2 = 10$)	0.73552 ($l_1 = 19, l_2 = 3$)

来加以说明. 在此测试过程中, 令 FLDC 方法和本文方法中的调节参数分别为 $\gamma = 0.1$ 和 $\gamma = 2$, 两者具有的共同参数设定为 $\varepsilon = 1E - 5, maxIter = 30, m = 2$, 且假设在测试 PCA 和 FLDC 方法时将测试图像投影到 d 维子空间, 测试 2DPCA 方法时将测试图像投影到 $32 \times d$ 的张量子空间, 而在测试 $(2D)^2PCA$ 和本文方法时则将测试图像投影到 d^2 (即 $l_1 = l_2 = d$) 的张量子空间. 测试结果见表 5、表 6 和图 8.

根据测试两类人脸识别数据所得的结果, 可以得到如下结论:

1) 从测试精度和特征提取的维数上来看, 表 5、

表 6 反映出 $(2D)^2UFFCA$ 方法具有较强的特征提取能力, 说明本文方法由于引进了模糊理论而使得特征提取效果更符合人脸图像所处的现实外部环境, 如光照的强弱特点等^[39], 提高了特征提取后的分类精度. 尽管 FLDC 方法也是通过引进模糊理论实现特征提取, 但从实验结果来看该方法的特征提取效果同本文方法相比还略显不足. 这是由于 FLDC 方法只能处理矢量模式数据, 而没有充分考虑人脸图像数据之间具有的空间结构信息, 特别是 FLDC 方法将求解模糊隶属度和实现特征提取作为两个独立的过程来实现, 而本文的方法则将求解隶属度与实现特征提取进行了有机的统一, 从而在一定程度上

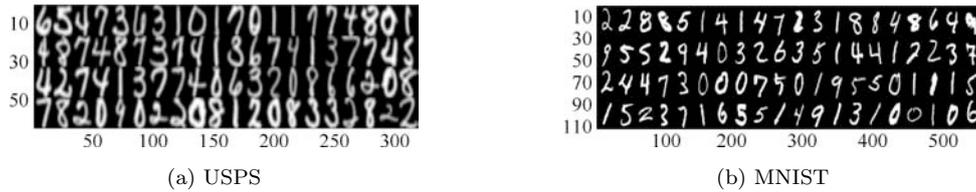


图 5 手写体图像数据 0~9

Fig. 5 Examples of handwritten digits from 0 to 9



图 6 ORL 数据集中某一类所有图像

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



图 7 Yale 数据集中某一类所有图像

Fig. 7 All images in a certain class in Yale datasets

表 5 对 ORL 数据集识别效果比较

Table 5 Comparison of recognition performances on ORL datasets

Number of training samples	2		3		4		5	
	Accuracy (Dimension)	Time (s)	Accuracy (Dimension)	Time (s)	Accuracy (Dimension)	Time (s)	Accuracy (Dimension)	Time (s)
PCA	0.6938 (73)	16.7701	0.71786 (121)	15.9121	0.80417 (156)	14.8201	0.84 (201)	14.2117
FLDC	0.6938 (214)	59.264	0.7250 (276)	56.16	0.82083 (403)	63.753	0.855 (446)	64.975
2DPCA	0.71875 (13×32)	0.014	0.76429 (18×32)	0.014	0.8625 (14×32)	0.015	0.89 (7×32)	0.014
$(2D)^2PCA$	0.72188 (10^2)	0.017	0.76786 (19^2)	0.016	0.8667 (18^2)	0.016	0.905 (18^2)	0.018
$(2D)^2UFFCA$	0.72813 (20^2)	0.019	0.7750 (18^2)	0.024	0.8708 (17^2)	0.029	0.905 (17^2)	0.034

表 6 对 Yale 数据集识别效果比较
Table 6 Comparison of recognition performances on Yale datasets

Number of training samples	2		3		4		5	
	Accuracy (Dimension)	Time (s)	Accuracy (Dimension)	Time (s)	Accuracy (Dimension)	Time (s)	Accuracy (Dimension)	Time (s)
PCA	0.4222 (29)	18.1741	0.45 (44)	17.4409	0.52381 (57)	17.1757	0.5778 (74)	16.6297
FLDC	0.4444 (189)	59.906	0.45 (153)	57.843	0.54286 (340)	62.285	0.6 (249)	59.748
2DPCA	0.5778 (7×32)	0.015	0.63333 (7×32)	0.016	0.67619 (6×32)	0.014	0.71111 (7×32)	0.015
(2D) ² PCA	0.5778 (7^2)	0.016	0.63333 (8^2)	0.015	0.68571 (8^2)	0.015	0.71111 (8^2)	0.017
(2D) ² UFFCA	0.5778 (8^2)	0.051	0.64167 (7^2)	0.062	0.69524 (8^2)	0.073	0.73333 (6^2)	0.079

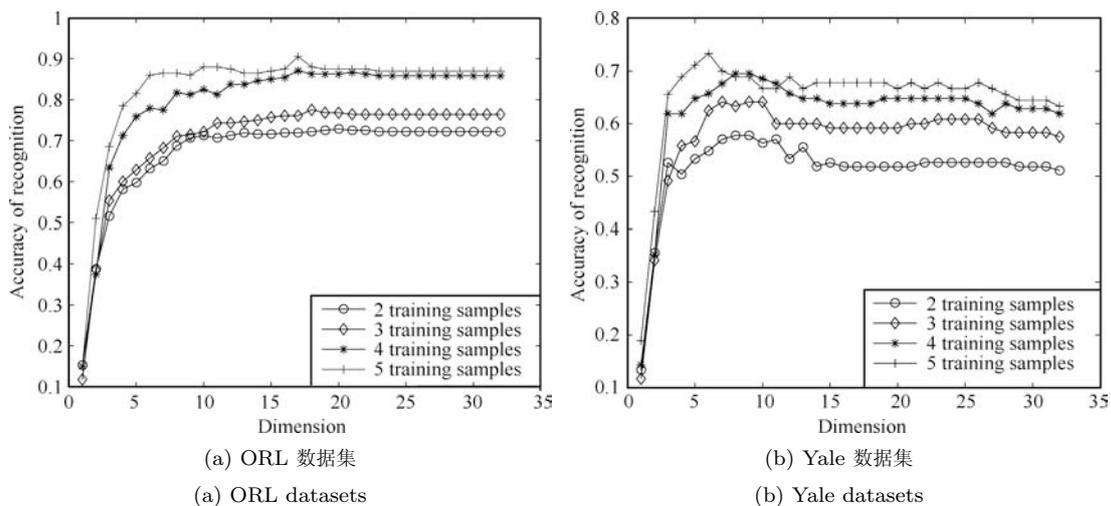


图 8 不同训练样本数对 (2D)²UFFCA 方法特征提取效果的影响

Fig. 8 Performances of (2D)²UFFCA method with different numbers of training samples

提高了图像样本模糊隶属度的区分度, 再次进行特征提取时会反映不同样本对特征提取效果的贡献程度.

2) 从进行特征值分解所消耗的 CPU 时间来看, 本文使用算法 1 来求解满足式 (11) 的 U 和 V 近似值具有一定的有效性, 同时在实验过程中也尝试使用文献 [32] 中的迭代方法来求解, 所得 CPU 时间是算法 1 的 1000 倍左右, 从而说明算法 1 是可行的.

3) 从图 8 可以看出, 本文方法随着训练样本数量的增加, 测试精度不断提高, 从而在一定程度上反映特征提取效果与训练样本数目之间成某种正比例

的关系, 这符合作为一种无监督特征提取方法所应具备的基本条件.

4 总结

本文运用模糊技术和张量理论, 提出一种矩阵模式的模糊最大间距判别准则, 并在此基础上形成具有模糊聚类功能的双向二维特征提取方法 (2D)²UFFCA. 该方法不但能直接实现矩阵模式数据的无监督聚类, 而且还可以对矩阵模式数据进行双向二维特征提取, 实现特征降维. 同时从理论上较好地解决了矩阵模式的模糊最大间距判别准则中的调节参数 γ 的问题, 并提出一种能有效求解计算矩

阵模式数据的正交投影变换矩阵的方法. 最后通过测试 UCI 数据集、手写体数据集和人脸图像数据集表明本文方法的有效性. 但从另一方面来说, 文献 [40] 中的方法在实现特征提取后又使用了具有明确直观含义的特征选择方法来进一步降低特征的冗余度, 从而提高了算法特征降维的效果. 因此, 研究如何引进合适的特征选择技术将是我们的下一个研究目标. 此外, 由于算法 1 计算得到的投影矩阵 U 和 V 会直接影响到本文方法的聚类效果和特征提取能力, 因此如何改进算法 1 以提高方法的稳定性, 也将是我们面临的难题.

References

- 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 IEEE 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)
- Bian Zhao-Qi, Zhang Xue-Gong. *Pattern Recognition (Second Edition)*. Beijing: Tsinghua University Press, 2000. 87–90
(边肇祺, 张学工. 模式识别 (第二版). 北京: 清华大学出版社, 2000. 87–90)
- Jolliffe I T. *Principal Component Analysis*. New York: Springer, 1986
- 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**(2): 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
- Wang X, Tang X. A unified framework for subspace face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(9): 1223–1228
- 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**(2): 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, Jiang T, Zhang K. Efficient and robust feature extraction by maximum margin criterion. *IEEE Transactions on Neural Networks*, 2006, **17**(1): 157–165
- Zheng W M, Zou C, 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
- Yan S, Xu D, Yang Q, Zhang L, Tang X, Zhang H. Multilinear discriminant analysis for face recognition. *IEEE Transactions on Image Processing*, 2007, **16**(1): 212–220
- Lei Z, Chu R, He R, Liao S, Li S Z. Face recognition by discriminant analysis with Gabor tensor representation. In: Proceedings of the International Conference of Biometrics. Seoul, Korea: Springer, 2007. 87–95
- Kim Y D, Choi S. Color face tensor factorization and slicing for illumination-robust recognition. In: Proceedings of the International Conference of Biometrics. Seoul, Korea: Springer, 2007. 19–28
- Lu H P, Plataniotis K N, Venetsanopoulos A N. Uncorrelated multilinear discriminant analysis with regularization and aggregation for tensor object recognition. *IEEE Transactions on Neural Networks*, 2009, **20**(1): 103–123
- Ren C X, Dai D Q. Incremental learning of bidirectional principal components for face recognition. *Pattern Recognition*, 2010, **43**(1): 318–330
- Yang J, Zhang D, Frangi A F, Yang J Y. Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(1): 131–137
- Li M, Yuan B. 2D-LDA: a statistical linear discriminant analysis for image matrix. *Pattern Recognition Letters*, 2005, **26**(5): 527–532
- Xiong H L, Swanmy M N S, Ahmad M O. Two-dimensional FLD for face recognition. *Pattern Recognition*, 2005, **38**(7): 1121–1124
- Jing X Y, Wong H S, Zhang D. Face recognition based on 2D Fisherface approach. *Pattern Recognition*, 2006, **39**(4): 707–710
- Wang J G, Yang W K, Lin Y S, Yang J Y. Two-directional maximum scatter difference discriminant analysis for face recognition. *Neurocomputing*, 2008, **72**(1–3): 352–358
- Zhang D Q, Zhou Z H. (2D)²PCA: two-directional two-dimensional PCA for efficient face representation and recognition. *Neurocomputing*, 2005, **69**(1–3): 224–231
- Ye J P, Janardan R, Li Q. Two-dimensional linear discriminant analysis [Online], available: <http://books.nips.cc/nips17.html>, November 29, 2011

- 28 Gao Jun, Wang Shi-Tong. Fuzzy maximum scatter difference discriminant criterion based clustering algorithm. *Journal of Software*, 2009, **20**(11): 2939–2950
(皋军, 王士同. 基于模糊最大散度差判别准则的聚类方法. 软件学报, 2009, **20**(11): 2939–2950)
- 29 Li C H, Kuo B C, Lin C T. LDA-based clustering algorithm and its application to an unsupervised feature extraction. *IEEE Transactions on Fuzzy Systems*, 2011, **19**(1): 152–163
- 30 Yang W K, Yan H, Wang J G, Yang J Y. Face recognition using complete fuzzy LDA. In: Proceedings of the 19th International Conference on Pattern Recognition. Tampa, USA: IEEE, 2008. 1–4
- 31 Yang W K, Yan X Y, Zhang L, Sun C Y. Feature extraction based on fuzzy 2DLDA. *Neurocomputing*, 2010, **73**(10–12): 1556–1661
- 32 Wang H, Yan S C, Huang T S, Tang X O. A convergent solution to tensor subspace learning. In: Proceedings of the 20th International Joint Conference on Artificial Intelligence. Hyderabad, India: AAAI, 2007. 629–634
- 33 Wang F, Wang X. Neighborhood discriminant tensor mapping. *Neurocomputing*, 2009, **72**(7–9): 2035–2039
- 34 Wu K L, Yu K, Yang M S. A novel fuzzy clustering algorithm based on a fuzzy scatter matrix with optimality tests. *Pattern Recognition Letters*, 2005, **26**(5): 639–652
- 35 Blake C L, Merz C J. UCI machine learning repository [Online], available: <http://www.ics.uci.edu/~mllearn/MLRepository.html>, November 29, 2011
- 36 Medvedovic M, Yeung K Y, Bumgarner R E. Bayesian mixture model based clustering of replicated microarray data [Online], available: <http://bioinformatics.oxfordjournals.org/content/20/8/12-22.full.pdf>, November 29, 2011
- 37 Chung F L, Wang S T, Deng Z H, Shu C, Hu D. Clustering analysis of gene expression data based on semi-supervised clustering algorithm. *Soft Computing*, 2006, **10**(11): 981–993
- 38 Jain A K, Dubes R C. *Algorithms for Clustering Data*. New Jersey: Prentice Hall, 1988
- 39 Kwak K C, Pedrycz W. Face recognition using a fuzzy Fisherface classifier. *Pattern Recognition*, 2005, **38**(10): 1717–1732
- 40 Yang J, Zhang D, Yong X, Yang J Y. Two-dimensional discriminant transform for face recognition. *Pattern Recognition*, 2005, **38**(7): 1125–1129



皋 军 东南大学自动化学院博士后, 盐城工学院信息工程学院副教授. 主要研究方向为人工智能, 模式识别, 数据挖掘, 模糊系统. 本文通信作者.

E-mail: gjxllin@yahoo.cn

(**GAO Jun** Postdoctoral at the School of Automation, Southeast University and associate professor at the School of Information Engineering, Yancheng Institute of Technology. His research interest covers artificial intelligence, pattern recognition, data mining, and fuzzy system. Corresponding author of this paper.)



孙长银 东南大学自动化学院教授. 主要研究方向为人工智能, 神经网络, 智能控制理论与方法, 模式识别.

E-mail: cysun@seu.edu.cn

(**SUN Chang-Yin** Professor at the School of Automation, Southeast University. His research interest covers artificial intelligence, neural networks, theory and design of intelligent control systems, and pattern recognition.)



王士同 江南大学数字媒体学院教授. 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: wxwangst@yahoo.com.cn

(**WANG Shi-Tong** Professor at the School of Digital Media, Jiangnan University. His research interest covers artificial intelligence, pattern recognition, data mining, neural networks, fuzzy system, medical image processing, and bioinformatics.)