一种新的核线性鉴别分析算法及其在人脸识别上的应用*)

郑宇杰[1] 杨静宇[1] 吴小俊[2] 王卫东[1] 张丽丽[1]

(南京理工大学计算机系 南京 210094)[1] (江苏科技大学电子信息学院 镇江 212003)[2]

摘 要 基于核策略的核 Fisher 鉴别分析(KFD)算法已成为非线性特征抽取的最有效方法之一。但是先前的基于核 Fisher 鉴别分析算法的特征抽取过程都是基于2值分类问题而言的。如何从重叠(离群)样本中抽取有效的分类特征没有得到有效的解决。本文在结合模糊集理论的基础上,利用模糊隶属度函数的概念,在特征提取过程中融入了样本的分布信息,提出了一种新的核 Fisher 鉴别分析方法——模糊核鉴别分析算法。在 ORL 人脸数据库上的实验结果验证了该算法的有效性。

关键词 核策略,核 Fisher 鉴别分析,模糊核 Fisher 鉴别分析,特征提取,人脸识别

A New Kernel Discriminant Analysis Algorithm and its Application to Face Recognition

ZHENG Yu-Jie[1] YANG Jing-Yu[1] WU Xiao-Jun[2] WANG Wei-Dong[1] ZHANG Li-Li[1]

(Department of Computer Science, Nanjing University of Science & Technology, Nanjing 210094)[1]

(School of Electronics and Information, Jiangsu University of Science & Technology, Zhenjiang 212003)[2]

Abstract Kernel Fisher discriminative analysis(KFD)algorithm based on kernel trick has been one of the effective nonlinear feature extraction methods. But all previous nonlinear feature extraction methods based on KFD algorithm which procedures are based on solving binary classification problem. How to extract effective discriminative information from overlapping(outlier)samples is still open. In this paper, a new KFD algorithm named fuzzy kernel discriminative analysis(FKFD)is proposed. In the proposed algorithm, fuzzy K-nearest neighbor(FKNN)algorithm is incorporated into the process of KFD algorithm and the corresponding fuzzy membership degrees are gained. Therefore, distribution information of samples is embedded in the proposed algorithm through fuzzy membership degrees. Experimental results on ORL face database demonstrate the effectiveness of the proposed algorithm.

Keywords Kernel trick, Kernel Fisher discriminative analysis, Fuzzy kernel Fisher discriminative analysis, Feature extraction, Face recognition

1 引言

基于核函数的学习算法已经成为目前模式识别和机器学习研究领域的一个研究热点,特别是随着统计学习理论研究的成熟,涌现了大量的基于核函数的学习算法,如支持向量机[1](SVMs)、核主成分分析[2]、核独立成分分析[3]以及核线性鉴别分析[4~6]等等[7]。其中核线性鉴别分析由于可以抽取得到原始样本的有效非线性分类信息,被广泛应用于模式识别的各个领域。如 Mika 等提出了针对两类鉴别分析的方法;Baudat 等进一步提出了多类的核线性鉴别分析方法。最近,Yang 提出了一个有效的核线性鉴别分析的框架[8,9]:KPCA+LDA,该框架揭示了核线性鉴别分析算法的本质,并且给出了一个简单的特征提取步骤。但是,先前的关于核线性鉴别分析算法的特征提取过程都是基于2值分类问题而言的,也就是说它是用来解决是否属于还是不属于某一类的特征提取问题。当数据样本存在重叠点或者离群点的情况下,这种简单的处理2值问题的算法显然会遇到困难。如何克服这个缺点以及从这些重叠(离群)样本中抽取得到有效的分类特征是本文要关注的问题。

基于对此类问题的关注,本文引入了模糊集[10]的思想,提出了一种新的基于模糊集理论的模糊核线性鉴别分析算法。该算法充分利用了样本的分布信息,有助于更有效地得

*)国家自然科学基金资助(编号:60472060)、南京理工大学科研发展基金。郑宇杰 博士研究生。

心法和二分法的思想,设计出了求解多维0-1背包问题的一种改进的遗传算法。它通过计算 KMP 问题(1)的位有效度,然后按位有效度重新排序,使其近似具有单调性。这样,使用含中值杂交算子等组成部分的改进遗传算法求解,能更快地收敛到近似最优解。实验结果表明,其对于求解多维0-1背包问题不但可行,而且非常有效。

参考文献

1 Lorie J, Savage L. Three problem in capital rationing. Journal of Business, 1955, 28: 229~239
2 Manne A, Markowitz H. On the solution of discrete programming problem. Econometrica, 1957, 25: 84~110
3 Gavish B, Pirkul H. Efficient algorithms for solving multiconstraint 0-1 knapsack problems to optimality. Mathematical Programming, 1985, 31: 78~105
4 Balas E, Martin C. Pivot and Complement - a Heuristic for 0-1 Programming. Manegement Science, 1980, 26(1): 86~96
5 Freville A. The multidimensional 0-1 knapsack problem: An overview. European Journal of Operational Research, 2004, 155: 1~21
6 Holland J. Adaptation in Natural and Artificial Systems. MI: Ann Arbor, MI: University of Michigan Press, 1975; Cambridge, MA: MIT Press, 1992
7 Goldberg D. Genetic Algorithms in Search, Optimization, and Machine Learning. Reading, MA, Addison-Wesley, 1989
8 Sakawa M, Kato K, Shibano T. An interactive fuzzy satisficing method for multiobjective multidimensional 0-1 knapsack problems through genetic algorithms. In: Proceedings of 1996 IEEE International Conference on Evolutionary Computation, Piscataway, NJ, USA, 1996
9 Sakawa M, Kato K. Genetic algorithms with double strings for 0-1 programming problems. European Journal of Operational Research, 2003, 144: 581~597
10 Bian R, Chen Z, Yuan Z. Improved Crossover Strategy of Genetic Algorithms and Analysis of Its Performance. In: Proceedings of the 3th world congress on Intelligent Control and Automation, Hefei, P. R. China, 2000

到样本的分类信息，特别是有助于得到重叠（离群）样本的分类信息。在ORL人脸数据库上的实验结果验证了该算法的有效性。

2 核线性鉴别分析

2.1 基础知识

非线性映射的思想是将原始空间线性不可分的样本经过一个非线性映射 Φ 函数投影到高维空间，转变成高维空间的线性可分情况。

$$\Phi: R^n \to H \quad x \mapsto \Phi(x) \tag{1}$$

式(1)表示将一个在低维空间 R^n 的原始样本通过非线性映射 Φ 映射到高维空间 H。在此高维空间 H 中利用 Fisher 鉴别准则得到最优鉴别矢量集，此时就相当于得到了原始空间的非线性最优鉴别矢量集。高维空间 H 中的 Fisher 鉴别准则如下：

$$J^{\Phi}(\varphi)=\frac{\varphi^T S_b^{\Phi}\varphi}{\varphi^T S_t^{\Phi}\varphi},\varphi\neq 0 \tag{2}$$

其中的

$$S_b^{\Phi}=\frac{1}{M}\sum_{i=1}^{c} l_i(m_i^{\Phi}-m_0^{\Phi})(m_i^{\Phi}-m_0^{\Phi})^T \tag{3}$$

$$S_t^{\Phi}=\frac{1}{M}\sum_{i=1}^{M}(\Phi(x_i)-m_0^{\Phi})(\Phi(x_i)-m_0^{\Phi})^T \tag{4}$$

分别代表了高维空间 H 中的类间散布矩阵和总体散布矩阵；$x_1,x_2,\cdots,x_M$ 表示 R^n 空间的 M 个训练样本；l_i 代表第 i 类样本的样本数目，并且满足 $\sum_{i=1}^{c} l_i=M$；m_i^{Φ} 是第 i 类样本在高维空间的平均值，m_0^{Φ} 代表高维空间所有样本的平均值。

此时我们通过在高维空间 H 的 Fisher 鉴别准则函数可以得到原始空间的非线性鉴别矢量集。然而在具体非线性算法中，非线性映射函数 Φ 一般而言都比较复杂，而核策略则避免了这个非线性映射的显式运算过程，用内积函数来实现具体的操作。

2.2 核线性鉴别分析

由以上的分析可知，在高维空间 H(Hilbert空间)中利用 Fisher 鉴别准则求最优鉴别矢量集可以看成是求广义特征方程 $S_b^{\Phi}\varphi=\lambda S_t^{\Phi}\varphi$ 的特征向量集。由于其任意一个特征向量都可以表示成特征空间向量的线性组合，因此有

$$\varphi=\sum_{j=1}^{M} a_j\Phi(x_j)=Q\alpha \tag{5}$$

其中 $Q=[\Phi(x_1),\cdots,\Phi(x_M)]$，$\alpha=(a_1,\cdots,a_M)^T$。

将式(5)代入式(2)，那么我们可以得到如下的 Fisher 鉴别准则：

$$J^K(\alpha)=\frac{\alpha^T(KWK)\alpha}{\alpha^T(KK)\alpha} \tag{6}$$

其中的矩阵

$$K=\widetilde{K}-I_M\widetilde{K}-\widetilde{K}I_M+I_M\widetilde{K}I_M \tag{7}$$

上式中 $I_M=(1/M)_{M*M}$ 为单位矩阵；$\widetilde{K}=Q^TQ$ 是一个 $M*M$ 的矩阵，其中的元素为

$$\widetilde{K}_{ij}=\Phi(x_i)^T\Phi(x_j)=(\Phi(x_i)\cdot\Phi(x_j))=k(x_i,y_j) \tag{8}$$

其中 $k(x,y)$ 就是给定的非线性映射 Φ 所对应的核函数。

式(6)中的 $W=diag(W_1,\cdots,W_c)$，其中的 W_j 是所有元素为 $1/l_j$ 的 l_j*l_j 维的矩阵。

下面考虑对中心化的矩阵 K 进行 QR 分解。假设 $\gamma_1,\gamma_2,\cdots,\gamma_m$ 是矩阵 K 的前 m 个最大特征值 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_m$ 所对应的特征向量。

令 $P=(\gamma_1,\gamma_2,\cdots,\gamma_m)$，$\Lambda=diag(\gamma_1,\gamma_2,\cdots,\gamma_m)$。此时我们可以得到 $K=P\Lambda P^T$。显而易见，$P^TP=I$，其中的 I 为一个单位矩阵。

将 $K=P\Lambda P^T$ 代入式(6)，那么可得到

$$J^K(\alpha)=\frac{(\Lambda^{1/2}P^T\alpha)^T(\Lambda^{1/2}P^TWP\Lambda^{1/2})(\Lambda^{1/2}P^T\alpha)}{(\Lambda^{1/2}P^T\alpha)^T\Lambda(\Lambda^{1/2}P^T\alpha)} \tag{9}$$

设

$$\beta=\Lambda^{1/2}P^T\alpha \tag{10}$$

式(9)就变成如下的形式

$$J(\beta)=\frac{\beta^T S_b\beta}{\beta^T S_t\beta} \tag{11}$$

其中

$$S_b=\Lambda^{1/2}P^TWP\Lambda^{1/2},S_t=\Lambda \tag{12}$$

因此最优特征向量可以通过求取矩阵 $S_t^{-1}S_b$ 前 $d(d\leqslant c-1)$ 个最大特征值所对应的特征向量 $\beta_1,\beta_2,\cdots,\beta_d$ 得到。

显然，我们可以从式(10)得出，对于一个给定的 β，至少有一个 α 满足 $\alpha=P\Lambda^{-1/2}\beta$。当我们通过式(12)得到 $\beta_1,\beta_2,\cdots,\beta_d$ 之后，我们能得到一组对应的满足式(6)的 $\alpha_j=P\Lambda^{-1/2}\beta_j(j=1,2,\cdots,d)$。

由此，满足式(2)的最优特征向量集为：

$$\varphi_j=Q\alpha_j=QP\Lambda^{-1/2}\beta_j,j=1,2,\cdots,d \tag{13}$$

以上的推导给出了求广义特征方程 $S_b^{\Phi}\varphi=\lambda S_t^{\Phi}\varphi$ 的特征值和特征矢量。然而，按照此过程的 KFD 算法比较复杂，Yang 提出了一个简单的框架：KPCA+LDA。

在 Yang 的算法中，式(13)可看成分2步骤进行。

步骤1：

$$y=\left(\frac{\gamma_1}{\sqrt{\lambda_1}},\cdots,\frac{\gamma_m}{\sqrt{\lambda_m}}\right)^T(\Phi(x_1),\cdots,\Phi(x_M))^T\Phi(x)$$
$$=\left(\frac{\gamma_1}{\sqrt{\lambda_1}},\cdots,\frac{\gamma_m}{\sqrt{\lambda_m}}\right)^T[k(x_1,x),\cdots,k(x_M,x)] \tag{14}$$

此过程能看成是采用 KPCA 算法将样本从 Hilbert 空间 H 转化到欧氏空间 R^m。

步骤2：

$$\varphi=B^Ty \tag{15}$$

其中 $B=(\beta_1,\cdots,\beta_d)$。此过程可看作是 R^m 空间的 LDA 算法。

由此，我们可以看到 KFD 算法被大大简化。以下的算法中我们将基于 KPCA+LDA 的框架讨论。

3 模糊核线性鉴别分析

现有的核线性鉴别分析方法都是处理2值问题的。当有重叠（离群）数据存在的时候，如何有效地抽取得到这些样本的特征信息呢？很明显，模糊集理论是一个很好的选择。利用模糊隶属度函数表示数据的分布，此时的2值问题就变成了一个可以利用隶属度函数将样本划分到多类的问题。因此，有效地利用样本的隶属度函数，样本的分布信息就能被融合到特征抽取过程之中，有助于得到更多的分类信息，特别是有助于得到重叠（离群）样本的特征信息。

3.1 模糊K近邻算法(FKNN)

在本文的算法中，模糊隶属度函数通过 FKNN[11,12] 算法得到。假设有 c 个已知的模式 $w_1,w_2,\cdots,w_c$，$X=\{x_i\}$，$i=1,2,\cdots,M$ 为训练样本集。X 中的每一个样本均属于一个已知的类别 w_j，也就是说 $x_i\in w_j$，其中 $i=1,2,\cdots,M$，$j=1,2,\cdots,c$。通过 FKNN 得到相应隶属度函数的计算步骤如下：

步骤1：计算训练集中的任意2个样本之间的欧氏距离。

步骤2：将得到的距离矩阵中的本身距离置为无限大（对角线位置）。

步骤3：将每一列的距离值按照从小到大次序排列，每列分别得到对应的一组最小距离的样本所属的类别（在 k 近邻

的情况下，那么返回的是一组 k 个数值的向量)。

步骤 4：根据下列公式计算隶属度值：

$$\mu_{ij}=\begin{cases}0.51+0.49(n_{ij}/k), & i\text{ 所属的类别与 } j \text{ 所属的类别相同}\\ 0.49(n_{ij}/k), & i\text{ 所属的类别与 } j \text{ 所属的类别不同}\end{cases} \tag{16}$$

上式中的 n_{ij} 表示应该属于第 j 类的样本现属于第 i 类的个数。

得到隶属度函数之后，那么每一类的中心点定义如下：

$$\omega_i=\frac{\sum_{j=1}^{M}u_{ij}x_j}{\sum_{j=1}^{M}u_{ij}} \tag{17}$$

因此，我们能得到相应的类中心矩阵 ω 和模糊隶属度矩阵 U：

$$U=\lfloor \mu_{ij}\rfloor \quad i=1,2,\cdots,c;j=1,2,\cdots,M \tag{18}$$

$$\omega=[\omega_i] \quad i=1,2,\cdots,c \tag{19}$$

3.2 模糊核线性鉴别分析

模糊核线性鉴别分析的关键是如何将训练样本的分布信息融入到相应的散布矩阵之中，特别是考虑如何从重叠(离群)的样本得到有效的分类信息，关键的问题是如何将相应的样本分布信息集成到特征抽取的过程之中。本文利用模糊隶属度函数和得到的类中心点重新定义了相应的散布矩阵，有效地利用了样本的分布信息。

在模糊类内散布矩阵的定义之中，那些距离样本中心点越近的样本对于分类所做的贡献越大。因此，根据隶属度函数，我们可以得到在模糊核线性鉴别分析中的类内散布矩阵定义如下：

$$FS_w=\sum_{i=1}^{c}\Big(\sum_{x_j\in w_i}u_{ij}{}^{p}(x_j-\omega_i)(x_j-\omega_i)^T\Big) \tag{20}$$

其中的 p 是用户定义的常量，控制模糊隶属度函数的影响。

与类内散布矩阵定义相反，在模糊类间散布矩阵的定义中，每类的类中心离所有样本的中心越远，那么其对分类所做的贡献越大。因此，我们得到的模糊类间散布矩阵的定义如下：

$$FS_b=\sum_{i=1}^{c}\Big(\big(1-\sum_{x_j\in w_i}u_{ij}{}^{p}/\sum_{j=1}^{M}u_{ij}{}^{p}\big)(\omega_i-\bar{x})(\omega_i-\bar{x})^T\Big) \tag{21}$$

同样地，式(21)中的 p 含义同式(20)。

相应地，总体散布矩阵可以定义成如下的形式：

$$FS_t=FS_b+FS_w \tag{22}$$

由此，我们得到了所有散布矩阵的定义，并且每一样本对于分类的影响也被考虑进相应的模糊散布矩阵的定义之中。

基于以上的定义，可以得到关于模糊核线性鉴别分析的完整算法描述：

步骤 1：首先采用 KPCA 算法将原始图像降维到 m 维空间。

步骤 2：在经过 KPCA 降维之后的 m 维空间中，利用 FKNN 算法得到模糊隶属度矩阵 U 和类中心矩阵 ω。

步骤 3：根据 ω 和 U 分别计算出在 m 维空间的模糊类内散布矩阵 FS_w、模糊类间散布矩阵 FS_b 和模糊总体散布矩阵 FS_t。然后将 FS_b 和 FS_t 代替 Fisher 鉴别准则中的 S_b 和 S_t，计算得到 m 维空间的最优鉴别矢量集。

步骤 4：得到所有样本的最优特征数据，选择分类器进行分类。

4 实验结果

我们的实验数据采用 ORL[13] 人脸数据库。ORL 人脸数据库包括从 1992 年 4 月到 1994 年 4 月剑桥大学实验室拍摄的一系列人脸图像。具体为 40 个人，每个人由不同表情或不同视点的 10 幅图像构成，倾斜角度不超过 20 度。人脸库中的部分人脸图像如图 1 所示。

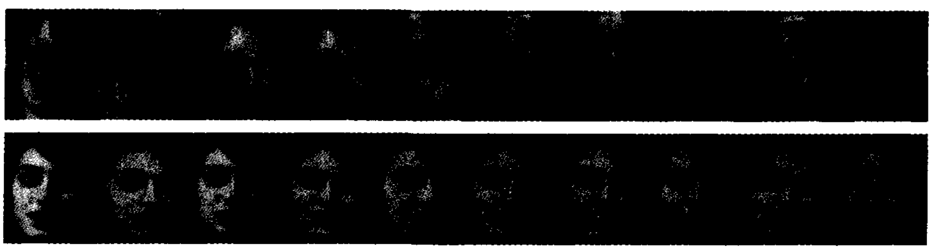

图 1　ORL 人脸数据库部分人脸图像

表 1　ORL 人脸数据库识别率(均值和方差)

# Training sample / class(ϑ)	4	5	6
KFD (Polynomial)	90.17 ±1.43	92.20 ±0.86	94.25 ±1.35
FKFD (Polynomial)	91.38 ±1.44	93.35 ±0.88	95.13 ±1.25
KFD (Sigmoid)	88.71 ±1.23	92.10 ±1.85	94.19 ±1.91
FKFD (Sigmoid)	89.12 ±1.19	93.00 ±1.68	94.94 ±1.04

实验中的训练数据集和测试数据集均随机生成，分别从每类中取 $\vartheta=4,5,6$，构成训练样本集。每次实验中的训练样本集均随机产生，数据库中训练样本集之外的数据构成测试样本集。最终抽取得到 39 维(i.e. $c-1$)特征；模糊散布矩阵中的参数 $p=2$。其中的核函数采用 2 种典型的核函数形式：

1)多项式核函数

$$k(x,y)=(x\cdot y'+1)^2 \tag{23}$$

2)Sigmoid 核函数

$$k(x,y)=\tanh\lfloor q(x\cdot y')+\Theta\rfloor \tag{24}$$

其中的参数 $q=0.01,\Theta=-4$。

最后采用最近邻分类器进行分类。我们的实验中，在每个不同训练样本数目下均做 10 次不同的实验。表 1 显示了在不同的实验方法中 10 次不同结果的平均识别率和方差比较。

图 2 和图 3 分别显示了在不同的核函数(Polynomial 和 Sigmoid)下，每类随机抽取 6 个训练样本的 10 次不同的识别率比较。

从表 1 可以得出，FKFD 的平均识别率优于传统的 KFD 特征提取方法。图 2 和图 3 的 10 次不同样本情况下的识别率比较图说明了 FKFD 方法有较好的鲁棒性，特别是在采用 Sigmoid 核函数的情况下。从以上的实验结果中可以得出，模糊核线性鉴别分析算法利用模糊隶属度函数融入了样本的

分布信息，通过在散布矩阵的定义之中引入相应的隶属度函数以考虑每一样本对鉴别矢量所做的贡献，消除重叠(离群)样本对于分类产生的影响。

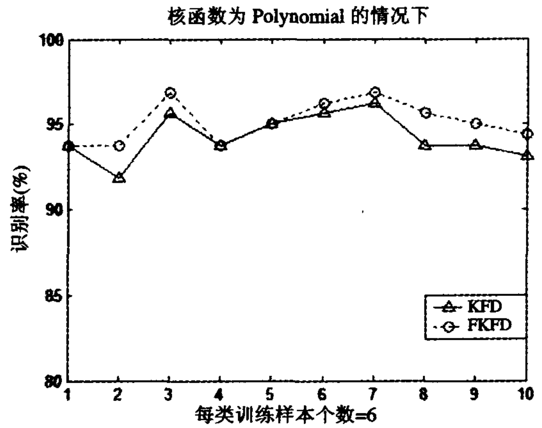

图 2　KFD 和 FKFD 特征提取方法识别率比较(采用 Polynomial 核函数)

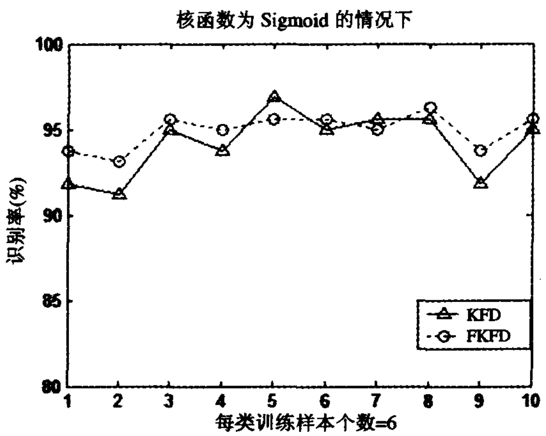

图 3　KFD 和 FKFD 特征提取方法识别率比较(采用 Sigmoid 核函数)

结束语　本文提出了一种新的基于模糊集理论的模糊核线性鉴别分析。随着模糊集理论的引入，样本的分布信息通过相应的隶属度函数被融入特征提取过程之中，有助于从重叠(离群)样本之中抽取得到有效的分类信息；并且基于 KPCA+LDA 的核线性鉴别分析方法给出了具体的算法。如何进一步对样本的分布情况进行研究，有效地利用样本的分布信息，特别是重叠(离群)样本的信息是我们今后要进一步关注的问题。

参 考 文 献

1　Vapnik V. The Nature of Statistical Learning Theory [M]. Springer, N Y, 1995

2　Schölkopf B, Smola A, Muller K R. Nonlinear component analysis as a kernel eigenvalue problem [J]. Neural Computation, 1998, 10(5): 1299～1319

3　Bach F R, Kernel M I. Independent Component Analysis [J]. Journal of Machine Learning Research, 2002, 3: 1～48

4　Mika S, Rätsch G, Weston J, et al. Fisher discriminant analysis with kernels [A]. In: IEEE International Workshop on Neural Networks for Signal Processing IX, Madison (USA), August, 1999. 41～48

5　Mika S, Rätsch G, Schölkopf B, et al. Invariant feature extraction and classification in kernel spaces [A]. Advances in Neural Information Processing Systems 12. Cambridge: MIT Press, 1999

6　Baudat G, Anouar F. Generalized discriminant analysis using a kernel approach [J]. Neural Computation, 2000, 12(10): 2385～2404

7　Müller K R, Mika S, Rätsch G, et al. An introduction to kernel-based learning algorithms [J]. IEEE Transactions on Neural Networks, 2001, 12(2): 181～201

8　Yang J, Jin Z, Yang J Y, et al. Essence of kernel Fisher discriminant: KPCA plus LDA [J]. Pattern Recognition, 2004, 37: 2097～2100

9　Yang J, Frangi A F, Yang J Y, et al. KPCA plus LDA: A Complete Kernel Fisher Discriminant Framework for Feature extraction and Recognition [J]. IEEE Trans Pattern Analysis and Machine Intelligence, 2005, 27(2): 230～244

10　Zadeh L A. Fuzzy sets [J]. Info Control, 1965, 8: 338～353

11　Keller J M, Gray M R, Givens J R. A fuzzy k-nearest neighbor algorithm [J]. IEEE Trans Syst Man Cybernet, 1985, 15(4): 580～585

12　Kwak K C, Pedrycz W. Face recognition using a fuzzy fisherface classifier [J]. Pattern Recognition, 2005, 38: 1717～1732

13　ORL face database. http://www.uk.research.att.com/facedatabase.html

(上接第 185 页)

3　Bykowski A, Rigotti C. A condensed representation to find frequent patterns. In: PODS, ACM, Santa Barbara, USA, pages May 2001. 267～273

4　Kryszkiewicz M, Gajek M. Concise representation of frequent patterns based on generalized disjunction-free generators. In: PAKDD, Springer, Taipei, Taiwan, pages May 2002. 159～171

5　Zaki M J, Hsiao C J. CHARM: An Efficient Algorithm for Closed Itemset Mining. In: Proc of 2002 SIAM Intl Conf Data Mining (SDM '02), Apr 2002. 457～473

6　Per J, Han H, Mao R. Closet: An efficient algorithm for mining frequent closed itemsets. SIGMOD Intl Workshop on Data Mining and Knowledge Discovery, May 2000

7　Baralis E, Chiusano S, Garza P. On support thresholds in associative classification. In: Proc of the 2004 ACM Symposium on Applied Computing, Nicosia, Cyprus, 2004. 225～232

8　Pudi V, Haritsa J. Generalized Closed Itemsets for Association Rule Mining. In: Proc of Intl Conf on Data Engineering (ICDE' 03), Mar. 2003. 1057～1073

9　Merz C J, Murphy P. UCI repository of machine learning database. http://www.cs.uci.edu/～mlearn/MLRepository.html

10　Quinlan J. C4.5: program for classification learning. Morgan Kaufmann. 1992

11　Liu B, Hsu W, Ma Y. Integrating classification and association rule mining. KDD'98, New York, NY, Aug. 1998

12　Yin X, Han J. CPAR: Classification based on predictive association rules. In: SIAM International Conference on Data Mining (SDM'03), San Francisco, CA, May 2003. 957～965