

非线性映射与特征提取: KMSE 模型与核主分量分析技术*

徐 勇¹ 宋枫溪² 李维杰¹

(哈尔滨工业大学深圳研究生院 深圳 518055)¹ (炮兵学院二系 合肥 230031)²

摘 要 分析表明, KMSE 模型准则中正则项的使用相当于引入了一个与核矩阵特征值直接相关的项以度量模型的泛化性能。根据矩阵特征值知识, 可知核主分量分析实际上为 KMSE 模型应用过程中的一个中间步骤。此时, KMSE 的作用表现为将样本在特征空间中的主分量映射为指示其类别的计算输出值。KMSE 模型可看作是在特征空间的主分量分析基础上进一步实施特征变换的过程。本文全面阐述了 KMSE 模型与 KFDA, LS-SVM, 核主分量分析以及 Bayesian 判别函数间的理论关系。此外, 通过分类实验测试了 KMSE、核主分量分析与本文方法的性能。

关键词 KMSE, 核主分量分析, 特征提取, 自动化技术, 特征空间

Nonlinear Mapping and Feature Extraction: KMSE Model and Kernel Principal Component Analysis Technique

XU Yong¹ SONG Feng-xi² LI Wei-jie¹

(Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518055, China)¹

(New Star Research Institute of Applied Technology, Hefei 230031, China)²

Abstract Analysis shows that the regularization term in the KMSE model is a term that aims to make the model generalize well and that is directly associated with the eigenvalue of the kernel matrix. According to matrix theory, we can know that kernel principal component analysis is an interim process of implementing KMSE. In fact, the KMSE model will further convert the obtained principal components into the computation outputs to indicate their class labels. The KMSE model can be considered to be a model that implements feature transform based on the principal component analysis in the feature space. This paper also presents clearly the theoretical relationships between KMSE and KFDA, LS-SVM, KPCA AND Bayesian decision function. In addition, the performance of KMSE, KPCA and the method developed in this paper are tested through classification experiments.

Keywords KMSE, Kernel principal component analysis, Feature extraction, Automation technique, Feature space

1 引言

作为自动化技术中的一个重要方法, 核方法在模式自动分类领域得到了大量应用^[1-9]。概括地讲, 作为一类非线性方法, 核方法的实现相当于如下两个过程的顺序执行: 将原空间(输入空间)非线性映射到一高维新空间(特征空间)的过程与在特征空间中实施线性特征抽取或建立线性预测模型的过程。

核最小平方误差方法(KMSE)^[11]是核方法中使用得较多的方法之一, 在人脸识别、机器学习等领域的应用都见诸文献^[4, 8-12]。最小平方误差方法^[13, 14]有如下理论性质: 当样本数趋于无穷时, 该方法以最小均方误差逼近贝叶斯判别函数^[13, 15]。KMSE 模型还可在形式上等效于 LS-SVM, 即最小平方支持向量机^[9]与核鉴别分析^[16]。

为提高 KMSE 模型的泛化性能, 常将正则项引入其目标函数即准则函数中。本文的分析表明, 引入正则项的目标函数包含的两项内容分别表示模型需在训练集上取得较小错误率, 且需使 $\alpha^T K \alpha$ (K 为核矩阵 $(K)_{ij} = k(x_i, x_j)$, α 为特征空间中的判别矢量) 对应较小值。本文根据矩阵特征值知识, 分析了 $\alpha^T K \alpha$ 的可能取值范围。深入的理论分析还表明, 核主分量分析实际上为 KMSE 模型应用过程中的一个中间步骤, 此时 KMSE 的作用表现为将样本在特征空间中的主分量

映射为指示其类别的计算输出值。实际上, KMSE 模型也可看作是在核主分量分析基础上实施特征提取的过程。此外, 本文总结了 KMSE 与核主分量分析、KFDA(核 Fisher 鉴别分析)以及 LS-SVM(最小平方支持向量机)间的关系。KMSE、核主分量分析与本文方法的对比实验分析支持理论分析得出的结论。

本文后续内容安排如下: 第 2 节为 KMSE 模型介绍; 第 3 节给出从准则函数引出 KMSE 模型的过程, 阐释了准则函数正则项之值与核矩阵特征值间关系, 形式化地给出了核主分量分析作为中间步骤的 KMSE 模型及其实现方案; 第 4 节为实验及结果分析, 最后是全文总结。

2 从 MSE 到 KMSE 模型

我们在本文中只考虑分别以 1 与 -1 作为类别标示的两类问题(也称 1 与 -1 为两类样本的理想输出值)。假设输入空间为 R^d , 非线性映射 ϕ 将输入空间映射到一高维特征空间, 并在特征空间中考虑分类问题。令训练样本数为 l , 其中对应类别标示 1 的训练样本共 l_1 个: $\phi(x_1), \phi(x_2), \dots, \phi(x_{l_1})$, 对应类别标示 -1 的训练样本共 l_2 个($l_1 + l_2 = l$): $\phi(x_{l_1+1}), \phi(x_{l_1+2}), \dots, \phi(x_l)$ 。最小平方误差方法基于如下方程组^[16]:

$$\Phi W = B \quad (1)$$

* 本文受国家自然科学基金重大国际(地区)合作项目(60620160097)及面上项目(60602038), 广东省自然科学基金(06300862)资助。

其中

$$W = \begin{bmatrix} w \\ w_0 \end{bmatrix}, B = [1 \cdots 1 \cdots -1 \cdots -1]^T, \\ \Phi = \begin{bmatrix} \phi(x_1) & \cdots & \phi(x_l) \\ 1 & & 1 \end{bmatrix}^T \quad (2)$$

矢量 B 中第 i 个分量 B_i 对应第 i 个样本 x_i 的类别标示。我们将 W 称为鉴别方向, 将 w 称为判别矢量, 将 w_0 称为阈值项。式(1)称为 MSE(最小平方误差)模型对应的方程组。下面我们将从 MSE 模型出发导出 KMSE 模型。

依据再生核理论, (1)式中判别矢量 w 可改写为

$$w = \sum_{i=1}^l \alpha_i^0 \phi(x_i) \quad (3)$$

引入核函数 $k(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$, 则方程组(1)可变形为

$$GA = B \quad (4)$$

$$A = [\alpha_1^0 \quad \alpha_2^0 \quad \cdots \quad \alpha_l^0 \quad w_0]^T \quad (5)$$

其中 $(G)_{i,i+1} = 1, (G)_{i,j} = k(x_i, x_j), i, j = 1, 2, \dots, l$ 。

此处 B 同(2)式, 方程组(4)的解可由其法方程组(6)求出:

$$(G^T G)A = G^T B \quad (6)$$

式(6)的解即为 KMSE 模型之解。基于方程组(4)可计算样本的计算输出值并判别其分类。具体分类方案如下: 根据 $y(x) = w_0 + \sum_{i=1}^l \alpha_i^0 k(x_i, x)$ 得出样本 x 的计算输出值后, 若 $y(x) \geq 0$, 则将 x 分类到第一类(即类别标示为 1 的类)中; 否则, 认为 x 属于第二类。

3 准则函数与最小平方误差模型

3.1 正则项与核主分量分析

原样本空间的最小平方误差准则为下式取得极小值:

$$O = \frac{1}{2} (y - XW - bu)^T (y - XW - bu) \quad (7)$$

其中 X 由原空间中所有训练样本组成且 $X = [x_1 \quad x_2 \quad \cdots \quad x_l]^T$, y 为由所有样本的类别标示(1 与 -1)组成的矢量, u 为元素全为 1 的矢量, bu 指代用 XW 表示 y 时的偏置, 常数 b 称为偏置系数。对应式(7)的数学模型称为 MSE(最小平方误差)模型, W 称为相应的判别矢量。而在由函数 ϕ 映射而得的特征空间中, MSE 模型对应的准则为如下函数取得极小值:

$$O_\phi = \frac{1}{2} (y - \Phi W_\phi - \beta u)^T (y - \Phi W_\phi - \beta u) \quad (8)$$

$$\Phi = [\phi(x_1) \quad \cdots \quad \phi(x_l)]^T \quad (9)$$

由再生核理论, 特征空间中的判别矢量 W_ϕ 也可如式(3)一样表达为特征空间中所有样本的线性组合形式: $W_\phi = \sum_{i=1}^l \alpha_i \phi(x_i)$, 此时 O_ϕ 可变形为

$$O_\phi = \frac{1}{2} (y - K^T \alpha - \beta u)^T (y - K^T \alpha - \beta u) \quad (10)$$

此处 β 为标量, 而矢量 $\alpha = [\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_l]^T$, 矩阵 K 依据核函数 $k(x_i, x_j)$ 定义为 $(K)_{ij} = \phi(x_i) \cdot \phi(x_j) = k(x_i, x_j), i, j = 1, 2, \dots, l$ 。易知式(10)在如下方程组成立条件下取得极值:

$$\begin{bmatrix} KK^T & Ku \\ (Ku)^T & l \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} Ky \\ u^T y \end{bmatrix} \quad (11)$$

由于 $G = [K \quad u], K^T = K$ (K 为对称矩阵), 容易看出式(11)与式(6)形式上完全相同, 可将方程组(11)和(6)之一作为 KMSE 模型对应的线性方程组。

需注意的是, 方程组(11)为“欠定”方程组, 无唯一解。为使获得有较好泛化性能的解, 可将特征空间中的目标函数修改为

$$O_\phi = \frac{1}{2} \mu W_\phi^T W_\phi + O_\phi \quad (12)$$

此处 μ 为一小正数, $\frac{1}{2} \mu W_\phi^T W_\phi$ 称为正则项。对应式(12)极小值的 KMSE 模型称为正则 KMSE 模型。由矩阵 K 和矢量 W_ϕ 的定义可知^[9]

$$W_\phi^T W_\phi = \alpha^T K \alpha \quad (13)$$

因此, 式(12)可改写为

$$O_\phi = \frac{1}{2} \mu \alpha^T K \alpha + O_\phi \quad (14)$$

显然, 正则 KMSE 模型对应的方程组应使得式(14)取得极小值。式(14)右端包含两项, 因此取得极小值要求这两项之值均较小。我们将二次型 $\alpha^T K \alpha$ 称作正则因子, 将 μ 称作正则参数。在参数 μ 固定的条件下, 第一项 $\frac{1}{2} \mu \alpha^T K \alpha$ 取得较小值的条件是 $\alpha^T K \alpha$ 之值较小。下面将对正则因子 $\alpha^T K \alpha$ 的取值范围进行分析。

引理 1^[17] 若 q_1, q_2, \dots, q_l 为矩阵 K 的 l 个单位特征向量, 相应特征值分别为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_l$, 则有 $q_i^T K q_i = \lambda_i$ 。

由代数理论还可知, q_1, q_2, \dots, q_l 为 l 维空间中的一组基, 因此 l 维空间中任一向量 p 均可表示为 $p = \sum_{i=1}^l \gamma_i q_i$, 可以证明, 如下定理成立。

定理 1 l 维空间中的任意单位向量 p_1 对应的 $p_1^T K p_1$ 之值均在区间 $[\lambda_l \quad \lambda_1]$ 范围内。

证明: 假设 p 为 l 维空间中的任一单位向量, 则可令 $p = \sum_{i=1}^l \gamma_i q_i$ 。由 $q_i^T K q_i = \lambda_i$ 可知

$$p^T K p = \sum_{i=1}^l (\gamma_i q_i)^T K (\gamma_i q_i) = \sum_{i=1}^l \gamma_i^2 q_i^T K q_i \\ = \sum_{i=1}^l \gamma_i^2 \lambda_i \quad (15)$$

由于 $\|p\|^2 = (\sum_{i=1}^l \gamma_i q_i)^T \sum_{i=1}^l \gamma_i q_i$, 据 q_i 的正交性(即 q_i 为单位正交向量: $q_i^T q_i = 1, q_i^T q_j = 0, i \neq j$)可知

$$\|p\|^2 = \sum_{i=1}^l \gamma_i^2 \quad (16)$$

因此, 若将 p 单位化为 $p_1 = p / \|p\|$, 则有

$$p_1^T K p_1 = \frac{\sum_{i=1}^l \gamma_i^2 \lambda_i}{\sum_{i=1}^l \gamma_i^2} \quad (17)$$

显然, $\lambda_l \leq p_1^T K p_1 \leq \lambda_1$, 且 $p_1 = q_1$ 时取得极大值 λ_1 , 而 $p_1 = q_l$ 时取得极小值 λ_l 。定理 1 得证。据定理 1, 我们可得出如下推论。

推论 1 式(14)的正则因子 $\alpha^T K \alpha$ 的取值范围为矩阵 K 的最小与最大特征值之间。

从模型应用的角度分析, 式(14)右端第二项较小意味着在训练集上能取得较小错误率, 第一项较小则意味着模型将有较好泛化性能。因此, 正则 KMSE 模型实质上要求模型同时具有较小训练误差与较好泛化性能。如若矩阵 K 的一较大特征值对应的特征向量 α 可抽取出对数据有较好描述能力的特征并对应较小的训练误差, 但较大的正则因子 $\alpha^T K \alpha$ 意味其并不对应最好的泛化性能。另一方面, 若将矩阵 K 的较小特征值对应的特征向量作为 KMSE 模型中的 α , 虽然会对应较小的正则因子, 但很难抽取出对样本描述能力强的特征从而往往会对应较大训练误差。从这个意义上讲, 正则 KMSE 模型的解强调较小训练误差与较好泛化性能之间的一个平衡; 而 K 的特征向量不适合直接作为 KMSE 模型中的 α , 原因是其很难满足准则(14)取得极小值且模型同时对

应较好泛化性能与较低错误率的要求。

不过,式(14)的正则项之值与矩阵 K 的特征值直接相关的事实却指出:可借助 K 的特征值技术求解 KMSE 模型。实际上,由上面分析可知,使式(14)取得极小值的 l 维向量 α 必位于以矩阵 K 的特征矢量为基的空间之中。因此,KMSE 模型必可在此基础上求解。我们在下面给出具体的求解步骤。

式(14)关于 α, β 取得极值的条件为式(14)中的 O_i^r 对 α, β 的偏导均为零。我们据此得知,下式成立即是包含正则项的准则(14)取得极值的条件。

$$\begin{bmatrix} K + \mu_2 I & u \\ u^T K^T & l \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} y \\ u^T y \end{bmatrix} \quad (18)$$

此处 μ_2 指代式(14)中的正则参数 μ 。据定理 1,若令 $Q = [q_1 \ q_2 \ \dots \ q_l]$,则可设 $\alpha = \sum_{i=1}^l \theta_i q_i = Q\theta$,其中 $\theta = [\theta_1 \ \theta_2 \ \dots \ \theta_l]^T$ 。

因此,式(18)可改写为

$$\begin{bmatrix} KQ + \mu_2 Q & u \\ u^T K^T Q & l \end{bmatrix} \begin{bmatrix} \theta \\ \beta \end{bmatrix} = \begin{bmatrix} y \\ u^T y \end{bmatrix} \quad (19)$$

由于 K 为对称矩阵,根据矩阵特征值知识,我们有 $K^T Q = KQ = \Lambda Q$, Λ 为对角阵,且其对角元素依次为 $\lambda_1 \lambda_2 \dots \lambda_l$ 。

求出 (19) 之解 $\Theta = [\theta \ \beta]^T$ 后,对样本 x 的分类可依据由(19)式计算出的其对应输出值 $y(x)$ 进行。具体步骤如下:首先计算 x 与所有训练样本间的核函数 $k(x, x_i), i=1, 2, \dots, l$ 。然后,计算

$$v_k(x) = Q[k(x, x_1) + \mu_2 k(x, x_2) + \mu_2 \dots k(x, x_l) + \mu_2]^T \quad (20)$$

之值。最后,计算

$$y(x) = \theta^T v_k(x) + \beta^T u \quad (21)$$

并依据如下原则分类:若 $y(x) \geq 0$,则 x 属于第一类,否则属于第二类。

式(20)实际上就是在特征空间中提取样本主分量的过程。提取出主分量后,式(21)的变换将主分量映射为样本的计算输出值。这表明,核主分量分析实际上为 KMSE 模型应用的一个中间步骤。

3.2 KMSE 模型及其它

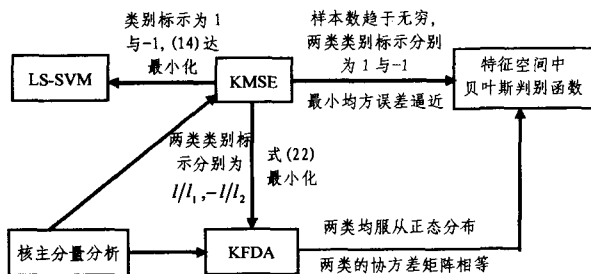


图1 KMSE与其它模型间关系示意图

除了本文的分析指出核主分量分析可作为实现 KMSE 模型的中间步骤之外,根据已有的研究还可知 KFDA 本质上等效于建立在核主分量分析基础之上的线性鉴别分析方法^[3]。此外,依据文献[15]对 MSE 模型的论述可知:在两类样本的类别标示分别为 1 与 -1 且样本数趋近无穷的条件下,KMSE 具有逼近特征空间的贝叶斯判别函数的特性。而在两类均服从正态分布且协方差矩阵相等的条件下,特征空间的贝叶斯判别函数又等效于 KFDA。此外,还可说明:在类

别标示为 $l/l_1, -l/l_2$ 的条件下以式(22)的最小化作为准则的 KMSE 模型等效于 KFDA,在类别标示为 1 与 -1 的条件下以式(14)的最小化作为准则的 KMSE 模型等效于 LS-SVM^[9]。由此,可由图 1 作为如上模型间理论关系的示意图。

$$O_i^r = \frac{1}{2} \mu \alpha^T \alpha + O_i \quad (22)$$

4 实验及结果分析

我们在几个基准数据集上进行实验。其中每个数据集均包含 100 个训练子集与测试子集。模型在第一个训练子集上进行训练,然后在所有测试子集上做分类测试。选用高斯核 $k(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / \sigma)$ 作为核函数,并将 σ 取值为第一个训练子集数据的协方差矩阵二范数的平方。表 1 列出了 KMSE、核主分量分析以及本文方法的平均分类错误率与标准偏差。数据项 $a(b)$ 中的 a 表示平均分类错误率,而 b 表示分类错误率标准差。其中,核主分量分析的结果在所提取特征 50 维,并采用最小距离分类器(即认为样本属于类均值距离其最近的类别)条件下得出。本文方法的结果在 3.1 节所示分类原则下取得。KMSE 亦按照若样本的计算输出值大于零则被分类到类别标示为 1 的类别,否则被分类为类别标示为 -1 的类别的原则分类。实验显示,本文方法取得与原 KMSE 模型相近的分类错误率,且低于核主分量分析方法所取得的分类错误率。本文方法取得优于核主分量分析的实验结果的事实说明将核主分量变换为计算输出有助于取得更好的分类性能。

表 1 三种方法的平均分配错误率和标准偏差

	KMSE 结果	核主分量分析结果	本文方法结果
german	26.04%(2.39%)	48.7%(2.54%)	25.42%(2.34%)
Titanic	22.7%(0.30)	32.3%(0.30%)	22.72%(0.27%)
flare-solar	31.7%(1.90%)	59.8%(1.73%)	30.7%(1.70%)
heart	7.2%(2.82%)	47.9%(3.69%)	9.3%(3.33%)

结束语 本文清楚地揭示了准则中包含正则项的 KMSE 模型与核主分量分析间的关系。理论分析显示,本文给出的正则 KMSE 模型的正则因子必在核矩阵最小特征值与最大特征值范围之间取值。实际上,核主分量分析可视为 KMSE 模型应用过程中的一个中间步骤。KMSE 模型中的矩阵变换可被分解为如下两个过程:一是主分量分析过程即提取样本在特征空间中的主分量的过程,二是在主分量基础上继续进行特征提取以使用提取后的特征逼近其理想输出值(即类别标示)的过程。除了与核主分量分析有密切关联外,KMSE 模型还与 KFDA,LS-SVM 以及特征空间的贝叶斯判别函数直接相关的特点显示了 KMSE 模型理论上的典型性和重要性。对比实验还显示:KMSE 模型实现过程中将核主分量变换为计算输出有助于取得更好的分类性能。

参考文献

- [1] Ruiz A, López-de-Teruel P E. Nonlinear Kernel-based Statistical Pattern Analysis. IEEE Transactions on Neural Networks, 2001,12(1):16-32
- [2] Muller K-R, Mika S, Ratsch G, et al. An introduction to kernel-based learning algorithms. IEEE Trans. on Neural Network, 2001,12(1):181-201

(下转第 232 页)

$(a_i, b_m), (a_j, b_n), (a_k, b_h) \in F_p$, 其中 i, j, m, n, k, h 为任意的自然数, $(a_1, \beta_1), (a_2, \beta_2), (\alpha, \beta) \in L$, 定义

$$r_2^5((a_k, b_h) \otimes (a_i, b_m), (a_k, b_h) \otimes (a_j, b_n)) = (a_i \otimes (a_j \vee a_k), b_m \otimes (b_n \vee b_h)),$$

$$t_2^5((\alpha_1, \beta_1), (\alpha_2, \beta_2)) = (\alpha_1 \vee \alpha_2, \beta_1 \vee \beta_2),$$

则对于任意 $T \in \mathfrak{S}_H$, T 关于 (r_2^5, t_2^5) 的 (α, β) - i 型闭, $(\alpha, \beta) \in L, i = I, II$.

证明: 根据 $T(r_2^5((a_k, b_h) \otimes (a_i, b_m), (a_k, b_h) \otimes (a_j, b_n))) = T(a_i \otimes (a_j \vee a_k), b_m \otimes (b_n \vee b_h)) = T((a_i, b_m) \otimes ((a_j, b_n) \vee (a_k, b_h))) = T(((a_i, b_m) \otimes (a_j, b_n)) \vee ((a_i, b_m) \otimes (a_k, b_h))) = T((a_i, b_m) \otimes (a_j, b_n)) \vee T((a_i, b_m) \otimes (a_k, b_h)) = t_2^5(T((a_i, b_m) \otimes (a_j, b_n)), T((a_i, b_m) \otimes (a_k, b_h))) \geq (\alpha, \beta) \otimes t_2^5(T((a_i, b_m) \otimes (a_j, b_n)), T((a_i, b_m) \otimes (a_k, b_h)))$

并且有 $t_2^5(T((a_i, b_m) \otimes (a_j, b_n)), T((a_i, b_m) \otimes (a_k, b_h))) \geq t_2^5((\alpha, \beta) \otimes T((a_i, b_m) \otimes (a_j, b_n)), (\alpha, \beta) \otimes T((a_i, b_m) \otimes (a_k, b_h)))$

因此对于任意 $T \in \mathfrak{S}_H$, T 关于 (r_2^5, t_2^5) 的 (α, β) - i 型闭, $(\alpha, \beta) \in L, i = I, II$.

在实际应用中, 往往需要合理的推理规则。由上述所得到的推理规则可以看出这些推理规则具备两个特点: 推理规则中含有自然语言; 其次它们均有语义部分和语法部分。因此这些规则是可靠有效的。

参考文献

[1] Borns D W, Mack J M. An algebraic introduction on mathematical logic [M]. Berlin: Springer, 1975
 [2] Goguen J A. The logic of inexact concepts [J]. Synthese, 1969, 19: 325-373
 [3] Pavelka J. On fuzzy logic I, II, III [J]//Zeitschr. F, ed. Math. Logic and Grundlagend Math., 1979, 25: 45-52, 119-134, 447-464
 [4] Novak V. First-order fuzzy logic [J]. Studia Logica, 1982, 46 (1): 87-109
 [5] Bolc L, Borowik P. Many-valued Logics [M]. Berlin: Springer-Verlag, 1992

[6] Ben-Eliyahu R, Dechter R. Default reasoning using classical logic [J]. Artificial Intelligence, 1996, 84: 113-150
 [7] 徐扬. 格蕴涵代数[J]. 西南交通大学学报, 1993, 28(1): 20-27
 [8] 徐扬, 秦克云. 格值命题逻辑(I). 西南交通大学学报, 1993, 1(2): 123-128
 [9] 秦克云, 徐扬. 格值命题逻辑(II)[J]. 西南交通大学学报, 1994, 2(1): 22-27
 [10] 徐扬, 秦克云. 关于格值逻辑系统的研究[C]//西南交通大学百年校庆论文集. 成都: 西南交通大学出版社, 1996: 72-80
 [11] 徐扬, 秦克云, 宋振明. 格值一阶逻辑的 FM 算法[J]. 科学通报, 1997(10): 1052-1055
 [12] Xu Y, Qin K Y, Liu J, et al. L-valued propositional logic $L_{\vee \wedge}$ [J]. Information Science, 1999, 114: 205-235
 [13] Xu Y, Liu J, Song Z, et al. On semantics of L-valued first-order logic [J]. International Journal Gen. Systems, 2000, 29: 53-79
 [14] Xu Y, Song Z, Qin K, et al. Syntax of L-valued first-order logic $L_{\vee \wedge}$ [J]. Mult. -valued logic, 2001, 7: 213-257
 [15] Xu Y, Ruan D, Liu J. Approximate reasoning based on lattice-valued propositional logic $L_{\vee \wedge}$ [J]//Ruan D. and Kerre E E, eds. Fuzzy If-Then Rules in Computational Intelligence: Theory and Applications. Kluwer Academic Publishers, 2000: 81-105
 [16] 秦克云, 徐扬, 宋振明. 基于系统 $L(X)$ 的几种近似推理[J]. 模糊系统与数学, 1998, 12(2): 55-60
 [17] Xu Y, Ruan D, Qin K Y, et al. Lattice-Valued Logic-An alternative approach to treat fuzziness and incomparability. Studies in Fuzziness and Soft Computing [M]. Berlin: Springer, 2003
 [18] Pei Z, Xu Y. Lattice implication algebra model of a kind of linguistic terms and its inference [C]//Proceeding of the 6th International FLINS Conference. 2004: 93-98
 [19] Chen S W, Xu Y, Ma J. A linguistic truth-valued uncertainty reasoning model based on lattice-valued logic. Fuzzy Systems and Knowledge Discovery [C]//International Conference, FSKD2005. LNAI 3613, 2005: 276-283
 [20] Meng D, Jia H D, XU Y. Framework of six linguistic lattice-valued evaluation system [C]// International Conference on Intelligent Systems and Knowledge Engineering. 2006
 [21] Xu Y, Chen S W, Ma J. Linguistic truth-valued lattice implication algebra and its properties [C]//Proceeding of CESA2006 Conference. 2006: 1413-1418
 [22] Lai J J, Xu K J, Xu Y, et al. Linguistic truth-valued lattice value propositional logic $\ell P(X)$ (in press)

(上接第 212 页)

[3] Yang Jian, Jin Zhong, Yang Jing-Yu, et al. Frangi: Essence of kernel Fisher discriminant: KPCA plus LDA. Pattern Recognition, 2004, 37(10): 2097-2100
 [4] Billings S A, Lee K L. Nonlinear Fisher discriminant analysis using a minimum squared error cost function and the orthogonal least squares algorithm. Neural Networks, 2002, 15(2): 263-270
 [5] Chen S, Hong X, Harris C J. Sparse kernel regression modeling using combed locally regularized orthogonal least squares and D-optimality experiments design. IEEE Trans. on Automatic Control, 2003, 48(6): 1029-1036
 [6] Mika S, Rätsch G, Weston J, et al. Fisher discriminant analysis with kernels//Hu Y H, Larsen J, Wilson E, et al. Neural Networks for Signal Processing IX. IEEE, 1999: 41-48
 [7] Mika S, Smola A J, Schölkopf B. An improved training algorithm for kernel fisher discriminants//Jaakkola T, Richardson T, eds. Proceedings AISTATS. Morgan Kaufmann, 2001: 98-104
 [8] Xu Y, Yang J Y, Lu J F. An efficient kernel-based nonlinear regression method for two-class classification // Proceedings of 2005 International Conference on Machine Learning and Cybernetics. Guangzhou, China, 2005: 4442-4445
 [9] Xu J, Zhang X, Li Y. Kernel MSE algorithm: A unified frame-

work for KFD, LS-SVM and KRR//Proceedings of the International Joint Conference on Neural Networks (IJCNN-2001). Washington, D. C., 2001: 1486-1491
 [10] Xu Yong, Zhang D, Song Fengxi, et al. A method for speeding up feature extraction based on KPCA. Neurocomputing, 2007, 70 (4/6): 1056-1061
 [11] Xu Y, Yang J Y, Lu J, et al. An efficient renovation on kernel Fisher discriminant analysis and face recognition experiments. Pattern Recognition, 2004, 37 (10): 2091-2094
 [12] Xu Y, Zhang D, Jin Z, et al. A fast kernel-based nonlinear discriminant analysis for multi-class problems. Pattern Recognition, 2006, 39(6): 1026-1033
 [13] Duda R, Hart P. Pattern Classification and Scene Analysis. New York: Wiley, 1973
 [14] Song Fengxi, Yang Jingyu, Liu Shuhai. Pattern recognition based on the minimum norm minimum-squared error classifier//The Eighth International Conference on Control, Automation, Robotics and Vision. Kunming, China, December 2004
 [15] 边肇祺, 张学工. 模式识别. 北京: 清华大学出版社, 2000
 [16] 徐勇, 陆建峰, 金忠, 等. 改进的 KMSE 方法及其应用. 模式识别与人工智能, 2007, 20(3): 394-398
 [17] 俞正光, 等. 线性代数. 北京: 清华大学出版社, 2005