

文章编号: 1001-0920(2007)07-0829-04

# 核空间中的 Xie-Beni 指标及其性能

普运伟<sup>1,2,3</sup>, 金炜东<sup>1</sup>, 朱 明<sup>1,2</sup>, 胡来招<sup>2</sup>

(1. 西南交通大学 信息科学与技术学院, 成都 610031; 2. 电子对抗国防科技重点实验室, 成都 610036; 3. 昆明理工大学 计算中心, 昆明 650093)

**摘要:** 针对核空间中模糊聚类算法的有效性评价问题,以核非线性映射为工具,将 Xie-Beni 指标推广到核 Hilbert 空间,得到其对应的核化形式,并指出该核化指标与  $V_{LL}$  指标的区别和联系.在此基础上,通过比较实验,研究了核化的 Xie-Beni 指标对高斯核宽度和模糊指数  $m$  的稳定特性.结果表明,核化的 Xie-Beni 指标较之  $V_{LL}$  等其他指标具有更好的性能和可靠性,可优先作为核模糊聚类算法的有效性判据.

**关键词:** 核模糊聚类; Xie-Beni 指标; 聚类有效性

**中图分类号:** TP181; TP391.4

**文献标识码:** A

## Kernelized Xie-Beni index and its performance

PU Yun-wei<sup>1,2,3</sup>, JIN Wei-dong<sup>1</sup>, ZHU Ming<sup>1,2</sup>, HU Lai-zhao<sup>2</sup>

(1. School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China; 2. National Electronic Warfare Laboratory, Chengdu 610036, China; 3. Computer Center, Kunming University of Science and Technology, Kunming 650093, China. Correspondent: PU Yun-wei, E-mail: puyunwei@126.com)

**Abstract:** With the help of the nonlinear mapping defined by kernel function implicitly, the noted Xie-Beni index for fuzzy  $c$ -means clustering is generalized into the kernel-defined Hilbert space. And the relation between the kernelized version of Xie-Beni and another index  $V_{LL}$  is also investigated. Then, the performances and the dependencies of these two indices on Gauss kernel-width and fuzzy exponent  $m$  are examined by some benchmark experiments, compared with two previously formulated indices, partition coefficient and partition entropy. The results show the superior performance and reliability of the kernelized Xie-Beni index in comparison to other indices, and it can take priority of being employed as the validity criterion for the kernel fuzzy clustering algorithm.

**Key words:** Kernel fuzzy clustering; Xie-Beni index; Cluster validity

### 1 引言

核模糊  $c$  均值聚类 (KFCM) 算法借助核非线性映射,将原空间中的数据隐式映射到高维核空间,凸显了样本间的特征差异,从而取得更好的聚类结果<sup>[1-3]</sup>.然而,与经典 FCM 一样,KFCM 也需事先确定聚类数,这直接影响了该算法的实际应用.因此,必须引入有效性评价指标,以便能在聚类的同时自动发现数据集中的最佳聚类数目.对于 FCM,人们提出了许多这样的指标<sup>[4]</sup>,但这些指标大多不能直接用于 KFCM 算法.划分系数  $V_{PC}$ ,划分熵  $V_{PE}$  等指标由于仅使用隶属度矩阵,可直接用于 KFCM;而 Xie-Beni 的  $V_{XB}$  等为数众多的指标不仅使用了隶属度矩阵,还使用了数据集的结构信息.这类指标在样

本集被执行非线性映射时,样本间的几何结构也相应发生了改变,因此必须作适当的处理才能用于 KFCM.鉴于 Xie-Beni 指标的高使用率及优异性能<sup>[4,5]</sup>,本文将其推广到核空间,得到与之相应的核化表示  $V_{KXB}$ ,并研究  $V_{KXB}$  与文献[2]提出的  $V_{LL}$  指标间的区别和联系.实验结果验证了  $V_{KXB}$  良好的性能及可靠性.

### 2 核模糊聚类算法

设待聚类样本集为  $X = \{x_1, x_2, \dots, x_N\}$ ,核非线性映射为  $\phi: X \rightarrow \mathcal{H}$ .则 KFCM 聚类的目标函数为<sup>[1-3]</sup>

$$J_m(X; U, V) = \sum_{i=1}^c \sum_{j=1}^N u_{ij}^m \| \phi(x_j) - (v_i) \|^2 =$$

收稿日期: 2006-04-11; 修回日期: 2006-07-15.

基金项目: 国家自然科学基金项目(60572143); 国家电子对抗技术预研基金项目(NEWL51435QT220401).

作者简介: 普运伟(1972—),男,云南通海人,博士生,从事模式识别、电子战信号处理等研究;金炜东(1959—),男,安徽桐城人,教授,博士生导师,从事智能信息处理、优化理论与方法等研究.

$$\sum_{i=1}^c \sum_{j=1}^N u_{ij}^m [ (x_j, x_j) - 2 (x_j, v_i) + (v_i, v_i) ] \quad (1)$$

式中:  $v_i$  为第  $i$  类的类中心;  $(v_i)$  为该类中心在相应核空间中的像,且有

$$(v_i) = \frac{\sum_{k=1}^N u_{ik}^m (x_k)}{\sum_{k=1}^N u_{ik}^m} \quad (2)$$

为最小化式(1),利用核函数与特征空间内积的关系  $(x_i, x_j) = (x_i), (x_j)$ , 不难得到

$$(x_j, v_i) = (x_j), (v_i) = \frac{\sum_{k=1}^N u_{ik}^m K(x_k, x_j)}{\sum_{k=1}^N u_{ik}^m} \quad (3)$$

$$(v_i, v_i) = (v_i), (v_i) = \frac{\sum_{k=1}^N \sum_{s=1}^N u_{ik}^m u_{is}^m K(x_k, x_s)}{(\sum_{k=1}^N u_{ik}^m)^2} \quad (4)$$

设  $V(c)$  为所使用的有效性指标,则 KFCM 算法可描述如下:

Step1: 设置目标函数精度,模糊指数  $m$ ,最大迭代次数  $T$ ,最大聚类数  $C_{max}$ ,并令  $c = 2$ .

Step2: 选择核函数及相应参数,计算核矩阵.

Step3: 随机初始化模糊划分矩阵  $U = [u_{ij}]_{c \times n}$ .

Step4: for  $t = 1 : T$ , 有:

- 1) 利用式(1) 计算目标函数  $J(t)$ ;
- 2) 更新  $U$  阵

$$u_{ij} = \frac{[1/d_T^2((x_j), (v_i))]^{-1/(m-1)}}{[1/d_T^2((x_j), (v_k))]^{-1/(m-1)}};$$

3) 若  $|J(t) - J(t - 1)| < \epsilon$  或  $t = T$ ,则转 Step5;否则,  $t = t + 1$ ,转 1).

Step5: 计算  $V(c)$ .若  $c < C_{max}$ ,则  $c = c + 1$ ,转 Step3; 否则,结束算法,并选择  $V(c)$  为极值(极大或极小由所采用的指标而定),对应的  $c$  为最佳聚类数  $C_{opt}$ ,相应的  $U$  为最佳模糊划分矩阵.

### 3 核空间中的 Xie-Beni 指标

#### 3.1 Xie-Beni 指标的核化形式

Xie-Beni 指标  $V_{XB}$  采用紧致性来评价类内的内聚程度,采用分离性来评价不同类间的隔离程度.其原空间中的定义为<sup>[4]</sup>

$$V_{XB} = \frac{\sum_{i=1}^c \sum_{j=1}^N u_{ij}^m (x_j - v_i)^2}{n(\min_{j,k} (v_j - v_k)^2)} \quad (5)$$

其中:  $\frac{1}{n} \sum_{i=1}^c \sum_{j=1}^N u_{ij}^m (x_j - v_i)^2$  用于评价聚类的紧致性,  $\min_{j,k} (v_j - v_k)^2$  用于评价类间的分离性.因为好的聚类划分应使类间的隔离尽可能大,而类内的散布则尽可能小,因此  $V_{XB}$  取极小值时对应最佳划分.

现将  $V_{XB}$  推广到核 Hilbert 空间,则对应的核化表示  $V_{KXB}$  可写成如下形式:

$$V_{KXB} = \frac{\sum_{i=1}^c \sum_{j=1}^N u_{ij}^m ((x_j) - (v_i))^2}{n(\min_{j,k} ((v_j) - (v_k))^2)} \quad (6)$$

注意到式(1),可知  $V_{KXB}$  的分子部分即为 KFCM 迭代结束时的目标函数值  $J(T)$ ,而分母中的二聚类质心间的距离平方为

$$((v_j) - (v_k))^2 = (v_j, v_j) - 2(v_j, v_k) + (v_k, v_k) \quad (7)$$

显然,  $(v_j, v_j)$  和  $(v_k, v_k)$  可由式(4) 计算,而

$$(v_j, v_k) = (v_j), (v_k) = \frac{\sum_{l=1}^N u_{jl}^m (x_l)}{\sum_{k=1}^N u_{jl}^m}, \frac{\sum_{s=1}^N u_{ks}^m (x_s)}{\sum_{s=1}^N u_{ks}^m} = \frac{\sum_{l=1}^N \sum_{s=1}^N u_{jl}^m u_{ks}^m K(x_l, x_s)}{\sum_{l=1}^N \sum_{s=1}^N u_{jl}^m u_{ks}^m} \quad (8)$$

因此,根据式(1),(4) 和(8) 便可计算得到核空间中的 Xie-Beni 指标  $V_{KXB}$ .

#### 3.2 $V_{KXB}$ 与 $V_{LL}$ 的关系

文献[2] 提出一种模糊核聚类的自适应算法,采用有效性指标  $s$ (本文表示为  $V_{LL}$ ) 来确定数据集的聚类数目.该指标表示为

$$V_{LL} = \frac{\text{comp}}{\text{sep}} = \frac{\sum_{i=1}^c \sum_{j=1}^n u_{ij} \frac{(x_i, x_i) - 2 \sum_{i=1}^T u_{ij} + \sum_{j=1}^T u_{ij}}{n u_{ij}^2}}{\min_{i,j} (\sum_{i=1}^T u_{ij} - 2 \sum_{i=1}^T u_{ij} + \sum_{j=1}^T u_{ij})} \quad (9)$$

式中

$$u_{ij} = \begin{cases} 1, & u_{ij} > u_{ij}, i = l; \\ 0, & \text{Others.} \end{cases}$$

$$i = (i_1, i_2, \dots, i_m),$$

$$j = (j_1, j_2, \dots, j_n),$$

$$i = (i_1, i_2, \dots, i_m)^T.$$

可见,  $u_{ij}$  阵即为  $[u_{ij}^m]$  阵对每行归一的结果,而  $i$  为核矩阵的第  $i$  列.因此,式(9) 中的分母部分

$$\min_{i,j} (\sum_{i=1}^T u_{ij} - 2 \sum_{i=1}^T u_{ij} + \sum_{j=1}^T u_{ij}) = \min_{i,j} ((v_i) - (v_j))^2 \quad (10)$$

也就是说,  $V_{LL}$  和  $V_{KXB}$  在计算分离性时是一致的.再注意到  $u_{ij}$  的作用以及  $(x_j) - (v_i)^2$  和  $(x_i, x_i) - 2 \sum_{i=1}^T u_{ij} + \sum_{j=1}^T u_{ij}$  相等的事实,不难发现两指标在计算类内紧致性时的差别:  $V_{KXB}$  考虑了对每个样本点到任意聚类中心的距离,并以  $u_{ij}^m$  加权;而  $V_{LL}$  仅考虑各样本点到其所属类质心的距离,并以  $1/u_{ij}^2$  加权.因此可以推断,在数据集的聚类结构较明显时,  $V_{KXB}$  和  $V_{LL}$  将具有几乎相近的性能,若在  $u_{ij}$  仅取 1 或 0 的极端情形(硬聚类),则二者将完全相等.而在

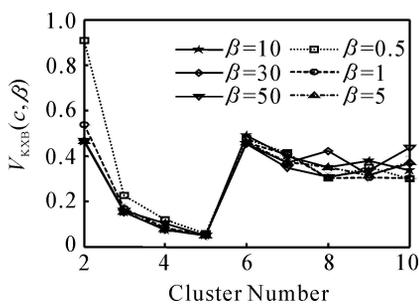
划分不明显时,  $V_{LL}$  的权重  $1/u_{ij}^2$  的放大作用将显著增强,使其紧致性急剧增大,  $V_{LL}$  将认为这种划分结果不甚合理. 相反,由于  $V_{KXB}$  在其紧致性的计算中利用了模糊矩阵的所有信息,样本点到其他类的距离贡献将在  $u_{ij}^m$  的作用下起到适度惩罚的作用. 但样本点到自身类中心的距离也将因为  $u_{ij}^m$  的减小而适度减小,因而整体指标值不会急剧变化. 综合看,  $V_{LL}$  具有使聚类划分倾向于较分明的趋势 ( $V_{PC}$  也有类似行为<sup>[4]</sup>),而  $V_{KXB}$  更能适应多种情形.

#### 4 实验结果与分析

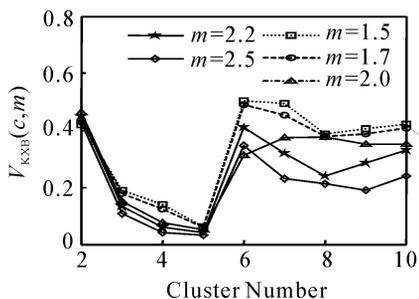
为验证  $V_{KXB}$  的性能和稳定性,下面通过 3 个实验进行测试,并与  $V_{PC}, V_{PE}, V_{LL}$  进行比较. 实验采用高斯核,算法精度 设为  $1e-5$ ,最大迭代次数  $T$  为 100. 为减少陷入局部极值的几率,对于每个聚类数  $c$ ,算法采取 10 次重复并从中选优的策略.

##### 实验 1 五高斯数据集

该数据集为简单球状分布,常用于检验聚类算法和有效性指标的基本性能和稳定性. 数据集的信息可参考文献[3]. 实验中,  $c_{max}$  取为 10. 为考察各指标对核宽度 的敏感特性,实验测试了  $= 0.3 \sim 50$  的范围. 图 1(a) 给出了  $V_{KXB}$  随  $c$  和 变化的性能曲线,其中  $= 1.0$  时各指标的值列于表 1.  $V_{PC}, V_{PE}, V_{LL}$  随  $(c, )$  具有类似于图 1(a) 的变化趋势,不再给出. 此外, KFCM 的性能与模糊指数  $m$  有关,加之  $V_{KXB}$  的表示式 (6) 中本身含有  $m$ ,因此实验也测试了各指标随  $m$  的变化情况 ( $m$  取  $1.5 \sim 2.5^{[6]}$ ),其中 为 5.0 时的实验结果列于表 2,  $V_{KXB}$  随  $(c, m)$  的性



(a)  $m = 2$  时  $V_{KXB}$  随  $c$  的变化情况,



(b)  $= 5$  时  $V_{KXB}$  随  $c, m$  的变化情况,

图 1 五高斯数据集的实验结果

表 1 五高斯数据集的有效性指标值 ( $= 1.0$ )

类数 $c$	$V_{PC}$	$V_{PE}$	$V_{LL}$	$V_{KXB}$
2	0.62	0.56	1.58	0.54
3	0.64	0.64	0.79	0.17
4	0.67	0.66	0.86	0.08
5	<b>0.76</b>	<b>0.53</b>	<b>0.15</b>	<b>0.05</b>
6	0.68	0.68	1.91	0.46
7	0.64	0.78	2.00	0.41
8	0.61	0.87	1.69	0.30
9	0.58	0.96	2.03	0.31
10	0.56	0.99	2.50	0.30

表 2 五高斯数据集最佳聚类数随  $m$  的变化 ( $= 5.0$ )

$m$	$V_{PC}$	$V_{PE}$	$V_{LL}$	$V_{KXB}$
1.5	5	5	5	5
1.6	5	5	5	5
1.7	5	5	5	5
1.8	5	5	5	5
1.9	5	5	5	5
2.0	5	5	5	5
2.1	5	2	5	5
2.2	5	2	5	5
2.3	5	2	5	5
2.4	5	2	5	5
2.5	5	2	5	5

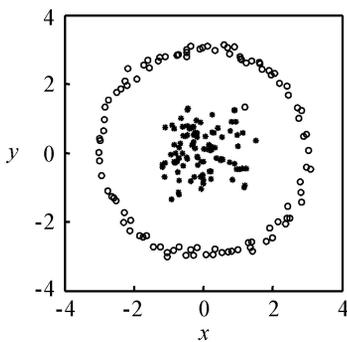
能曲线如图 1(b) 所示.

由图 1(a) 和表 1 的结果可知,在  $= 0.3 \sim 50$  的范围内, KFCM 均能稳定、正确地聚类(实验中无一错分样本),所考察的各有效性指标均正确识别出真实聚类数 5,且随 的变化表现出较好的稳定性. 同时,图 1(b) 和表 2 的有效性指标随  $m$  的依赖关系结果表明,除  $V_{PE}$  外,其余 3 个指标均能在  $m = 1.5 \sim 2.5$  的范围内可靠工作,始终给出正确聚类数 5. 在其他核宽度 的测试中,所得结果与表 2 类似.

##### 实验 2 Ring 数据集

Ring 数据集<sup>[3]</sup> 为非球状结构,由两类组成,分别是半径为 3 的圆环以及边长为 1 的正方形均匀分布数据. 两类的类中心均为  $(0,0)$ ,数据点数各为 100. 同时,两类数据分别被加上方差为 0.1 和 0.3 的高斯白噪声.  $= 0.2$  时的 KFCM 聚类结果如图 2 所示,有效性指标的取值列于表 3.

Ring 数据是验证核聚类算法的著名数据集,经典的 FCM 算法不能对其正确分类<sup>[3]</sup>. 从图 2 可知, KFCM 算法在  $= 0.2$  时具有较好的性能,仅有一个错分样本. 在 依赖关系的实验中发现,对于该数据集, 的有效范围远小于五高斯数据集,为  $0.1 \sim$

图2 Ring数据集的KFCM聚类结果( $\alpha = 0.2$ )表3 Ring数据集的有效性指标值( $\alpha = 0.2$ )

类数 $c$	$V_{PC}$	$V_{PE}$	$V_{LL}$	$V_{KXB}$
2	0.54	0.65	10.87	2.24
3	0.38	1.03	$2.62e+7$	$1.43e+6$
4	0.29	1.31	$3.31e+7$	$1.22e+6$

0.3. 且在该范围内, KFCM 难以形成大于4的聚类. 因此, 表3中仅列出  $c$  为 2, 3, 4 时各指标的值. 由该表可见, 测试的4个指标均检测到真实聚类数2. 对于0.1~0.3范围内的其他  $\alpha$  值, 结果与表3类似. 实验同样测试了各指标随  $m$  的变化情况, 结果显示各指标的取值趋势较为稳定, 也类似于表3, 不再给出. 可见, KFCM 特别适合于本实验这种具有特殊分布结构的数据集, 核化的 Xie-Beni 指标, 包括其余3个指标在  $\alpha \in [0.1, 0.3]$ ,  $m \in [1.5, 2.5]$  的范围内取值趋势均较为稳定.

### 实验3 UCI Wine 数据集

Wine 数据集由3类共178个样本组成, 每个样本13个属性. 3类样本中不存在交迭, 但划分不太明显<sup>[7]</sup>. 表4列出了  $\alpha = 10$ ,  $m = 2$  时有效性指标的值. 由该表可见, 仅有  $V_{KXB}$  检测到真实聚类数3, 而  $V_{PC}$ ,  $V_{PE}$  和  $V_{LL}$  得到的最佳聚类数均为2. 本实验中, 在  $m = 2$  且  $\alpha \in [1, 50]$  的范围内, 错聚样本数均为9(聚类正确率95%), 各指标随  $\alpha$  的取值趋势均类似于表4. 表5进一步给出了  $\alpha = 10$  时各指标的  $m$  依赖性. 可见, 所有指标在本实验中表现出较稳定的  $m$  依赖性.

总结上述3个实验的结果可知, KFCM 能适应

表4 Wine数据集的有效性指标值( $\alpha = 10$ )

类数 $c$	$V_{PC}$	$V_{PE}$	$V_{LL}$	$V_{KXB}$
2	0.62	0.56	1.68	0.50
3	0.50	0.86	2.49	0.41
4	0.37	1.16	176.24	12.26
5	0.30	1.38	26.41	1.15
6	0.25	1.55	2387.13	55.22
7	0.21	1.71	7935.25	110.60
8	0.19	1.84	6688.72	70.29

表5 Wine数据集最佳聚类数随  $m$  的变化( $\alpha = 10$ )

$m$	$V_{PC}$	$V_{PE}$	$V_{LL}$	$V_{KXB}$
1.5	2	2	2	3
1.6	2	2	2	3
1.7	2	2	2	3
1.8	2	2	2	3
1.9	2	2	2	3
2.0	2	2	2	3
2.1	2	2	2	3
2.2	2	2	2	3
2.3	2	2	2	3
2.4	2	2	2	3
2.5	2	2	2	3

多种分布类型的数据集, 给出较为合理的聚类划分. 在所考察的4个有效性指标中, 仅有  $V_{KXB}$  完全正确地识别出3个数据集的真实聚类数, 且结果具有最好的稳定性. 对于  $V_{LL}$  指标, 尽管所检测出的最佳聚类数也不随  $\alpha$  和  $m$  变化, 但却无法正确给出 Wine 数据的真实聚类数目. 这是因为对于五高斯和 Ring 数据集, 尽管分布类型不同, 但类与类之间的划分却较为明显. 而 Wine 数据的划分则不太分明, 因而使  $V_{LL}$  指标失效.  $V_{PC}$  也具有使聚类划分倾向于分明的特点, 在所有测试中性能表现与  $V_{LL}$  相当, 这与3.2节的分析结果一致.  $V_{PE}$  是4个测试指标中唯一表现出对  $m$  依赖性的指标, 其测试性能表现一般. 实际上, 考虑到 KFCM 最小化样本到质心距离平方和的本质, 可以预期建立在同一原则上的 Xie-Beni 指标将有相对较为理想的性能. 因此, 在实际应用 KFCM 时, 可优先选择  $V_{KXB}$  作为聚类有效性指示函数, 以便算法能自动寻求数据集的最佳聚类数目和最佳划分.

## 5 结论

KFCM 是 FCM 在高维特征空间中的推广, 显示出了比 FCM 更优良的性能和更强的适应各种数据分布的能力. 为了定量评价 KFCM 聚类结果的合理性, 本文将著名的 Xie-Beni 指标推广到核空间, 得到了相应的核化表示, 并分析了该表示与其他有效性指标的关系. 包括高斯核宽度  $\sigma$  以及模糊指数  $m$  的依赖性比较实验的结果, 证实了核化 Xie-Beni 指标的良好性能和可靠性.

## 参考文献(References)

- [1] Kim D-W, Lee K Y, Lee D, et al. Evaluation of the performance of clustering algorithms in kernel-induced feature space[J]. Pattern Recognition, 2005, 38(4): 607-611.

(下转第835页)

- Computer Science. Berlin: Springer-Verlag, 2006, 3927: 242-256.
- [2] Gao Y, Lygeros J, Quincampoix M, et al. On the control of uncertain impulsive system: Approximate stabilisation and controlled invariance[J]. Int J Control, 2004, 77(16): 1393-1407.
- [3] Quincampoix M, Seube N. Stabilization of uncertain control systems through piecewise constant feedback [J]. J Mathematics Analysis and Applications, 1998, 218(1): 240-255.
- [4] Cardaliaguet P, Quincampoix M, Saint-Pierre P. Pursuit differential games with state constraints [J]. SIAM J Control and Optimization, 2002, 39(5): 1615-1632.
- [5] Blanchini F. Set invariance in control[J]. Automatica, 1999, 35(11): 1747-1767.
- [6] Milani B E A, Dorea C E T. On invariant polyhedra of continuous-time linear systems subject to additive disturbances[J]. Automatica, 1996, 32(5): 785-789.
- [7] Aubin, J-P. Viability theory[M]. Boston: Birkhauser, 1991.
- [8] Clarke F H, Ledya Yu S, Stern R J, et al. Nonsmooth analysis and control theory[M]. New York: Springer-Verlag, 1998.

## (上接第 828 页)

- [5] Yu M, Wang L, Chu T. Robust stabilization of sampled-data systems [C]. Proc of American Control Conf. Portland, 2005: 3421-3426.
- [6] Cao Y, Sun Y, Cheng C. Delay-dependent robust stabilization of uncertain systems with multiple state delays[J]. IEEE Trans on Automatic Control, 2001, 43(11): 1608-1612.
- [7] Niculescu S I, Souza C E de, Dugard L, et al. Robust exponential stability of uncertain systems with time-varying delays[J]. IEEE Trans on Automatic Control, 1998, 43(5): 743-748.
- [8] Fridman E. New Lyapunov-Krasovskii functionals for stability of linear retarded and neutral type systems[J]. System Control Letters, 2001, 43(4): 309-319.
- [9] Fridman E, Shaked U. An improved stabilization method for linear time-delay systems[J]. IEEE Trans on Automatic Control, 2002, 47(11): 1931-1937.
- [10] Moon Y S, Park P, Kwon W H, et al. Delay-dependent robust stabilization of uncertain state-delayed systems[J]. Int J of Control, 2001, 74(14): 1446-1455.

## (上接第 832 页)

- [2] 李侃, 刘玉树. 模糊核聚类的自适应算法[J]. 控制与决策, 2004, 19(5): 595-597.  
(Li Kan, Liu Yu-shu. Fuzzy kernel clustering self-adaptive algorithm[J]. Control and Decision, 2004, 19(5): 595-597.)
- [3] Wu Z D, Xie W X, Yu J P. Fuzzy C-means clustering algorithm based on kernel method[C]. Proc of the 5th Int Conf on Computational Intelligence and Multimedia Applications. Xi'an: IEEE Computer Society, 2003: 1-6.
- [4] Kim D-W, Lee K H, Lee D. On cluster validity index for estimation of the optimal number of fuzzy clusters [J]. Pattern Recognition, 2004, 37(10): 2009-2025.
- [5] Xu R, Wunsch II D C. Survey of clustering algorithms [J]. IEEE Trans on Neural Networks, 2005, 16(3): 645-678.
- [6] Pal N R, Bezdek J C. On clustering for the fuzzy C-means model[J]. IEEE Trans on Fuzzy Systems, 1995, 3(3): 370-379.
- [7] Newman D J, Hettich S, Blake C L, et al. UCI repository of machine learning databases [DB/OL]. Irvine: University of California, Department of Information and Computer Science. (1995-02-17). <http://www.ics.uci.edu/~mllearn/MLRepository.html>.