

K 平面聚类算法的模糊改进及其鲁棒性研究

朱林 王士同 潘永惠 韩斌
(江南大学信息工程学院 无锡 214122)

摘要: 该文针对 K 平面聚类算法 KPC (K-Plane Clustering)对噪声点敏感的缺陷, 通过引入隶属度约束函数, 推导出鲁棒的改进分割 K 平面聚类算法 IFP-KPC(Improved Fuzzy Partitions for K-Plane Clustering), 并利用 Voronoi 距离对 IFP-KPC 算法的鲁棒性进行了合理解释。实验结果表明 IFP-KPC 算法较之于 KPC 算法具有更好的聚类效果。

关键词: K 平面聚类; 改进模糊分割; Voronoi 距离; 鲁棒性

中图分类号: TP391.4

文献标识码: A

文章编号: 1009-5896(2008)08-1923-05

Improved Fuzzy Partitions for K-Plane Clustering Algorithm and Its Robustness Research

Zhu Lin Wang Shi-tong Pan Yong-hui Han Bin

(School of Information Engineering, Southern Yangtze University, Wuxi 214122, China)

Abstract: A new robust Improved Fuzzy Partitions for K-Plane Clustering (IFP-KPC) algorithm is proposed. The proposed algorithm can reduce the sensitivity of the k-plane clustering algorithm to noises in real datasets. Also the distances to the Voronoi cell are used to give a reasonable explanation for the robustness of IFP-KPC. Experimental results demonstrate the effectiveness of IFP-KPC.

Key words: K-plane clustering; Improved fuzzy partitions; Voronoi distance; Robustness

1 引言

聚类分析是无监督模式识别中一种重要的方法, 已广泛应用于数据挖掘、图像处理、计算机视觉、生物信息和文本分析。聚类算法就是将分布未知的数据进行分类, 目的是寻找隐藏在数据中的结构, 并按照某种相似程度的度量, 尽可能地使具有相同性质的数据归于同一类。Bradley和Mangasarian提出了一种新颖的K平面聚类算法KPC (K-Plane Clustering)^[1], 其思想是通过用超平面替代传统K均值算法的中心点, 最小化样本点到最近超平面平方距离之和求取k个原型超平面。该方法在混合系统识别^[2]、稀疏成分分析^[3]、基因聚类^[4]、谱聚类^[5]和信号稀疏表达与应用的优化算法^[6]等方面取得初步应用。实验发现, KPC算法可以对未知样本集进行有效聚类, 但同时和数据噪声点较敏感, 算法鲁棒性较差。为此, 本文提出了改进分割的K平面聚类算法IFP-KPC (Improved Fuzzy Partitions for K-Plane Clustering), 通过在目标函数中引入隶属度约束, 得到意义更趋明晰的隶属度迭代公式, 使得IFP-KPC算法具有更好的鲁棒性, 提高了聚类准确性。

2 K 平面聚类算法(KPC)

KPC 算法用 c 个超平面替代传统 c 均值聚类算法中的中心点作为聚类原型, KPC 算法的目标是找到一组超平面 $P = \{p_1, p_2, \dots, p_c\}$, 使得每个样本点到距其最近的超平面的平方距离之和最小, w_i 和 γ_i 分别表示超平面 p_i 的单位化法向量和截距。聚类目标函数定义为

$$J_{KPC} = \sum_{i=1}^c \sum_{j=1}^n \|w_i^T x_j - \gamma_i\|^2 \quad (1)$$

KPC 算法就是用迭代方法求解, 当算法收敛, 就得到各类超平面的单位化法向量、截距和样本集的划分^[1]。KPC 算法能对未知样本集进行有效聚类, 但在实验中也发现其最显著的缺陷是对数据噪声很敏感, 算法鲁棒性较差。从图 1 可以看出 KPC 算法的聚类结果会受到噪声点的影响。

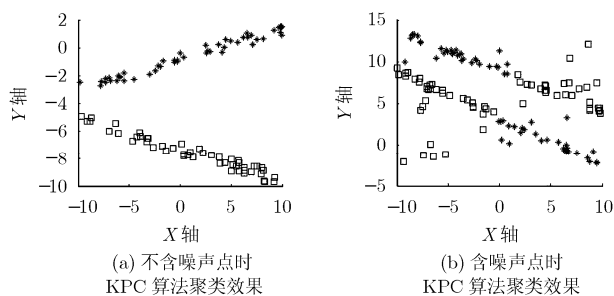


图 1 KPC 算法的聚类效果

2006-12-26 收到, 2007-07-16 改回

国家 863 项目(2006AA10Z313), 国家自然科学基金(60225015), 国防应用基础研究基金项目(A1420461266)和 2005 年教育部科学研究重点基金项目(105087)资助课题

3 改进分割的 K 平面聚类算法(IFP-KPC)

3.1 基于竞争学习理论的隶属度目标函数的构造

竞争学习的学习规则分为WTA (Winner-Take-All)与WTM(Winner-Take-More)两种, 分别称为“硬”竞争学习(hard competitive learning)与“软”竞争学习(soft competitive learning)^[7,8]。WTA规则存在一个重要问题: 对于不同初始值的节点, 学习过程中可能存在死节点(dead nodes)或不充分利用(under-utilization)的现象。为此, 研究人员提出了WTM规则, 通过引入模糊隶属度等方法, 削弱了学习对节点初始值的依赖, 但同时也造成数据的获胜节点不确定, 使得部分节点可能偏离其实际的类数据。文献[9]提出了惩罚对手的竞争学习(RPCL): 对于数据集中的每个样本, 不仅其获胜节点以一定的学习速率向该数据集“靠近”, 而且其竞争对手以更小的学习速率(惩罚速率)被“推离”该数据集, 从而减小了对学习过程的干扰。

基于惩罚对手的竞争学习(RPCL)的思想, 本文综合KPC与模糊KPC算法, 进一步提出了改进分割的K平面聚类算法IFP-KPC。通过引入隶属度约束函数, 构造新的目标函数:

$$J_{IFP-KPC} = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^2 \left(\|\mathbf{w}_i^T \mathbf{x}_j - \gamma_i\|^2 \right) + \sum_{j=1}^n a_j \sum_{i=1}^c u_{ij} (1 - u_{ij}) \quad (2)$$

对于一个单独样本点 \mathbf{x}_j , 隶属度约束函数:

$$f(u_1, u_2, \dots, u_c) = \sum_{i=1}^c u_i (1 - u_i) \text{ 有如下结论:}$$

定理 1 对于隶属度约束函数 $f(u_1, u_2, \dots, u_c)$, 当 $u_i (i = 1, 2, \dots, c)$ 均为 $1/c$ 时, 函数取得最大值 $1 - 1/c$; 当 $u_i (i = 1, 2, \dots, c)$ 其中一个 u_k 为 1, 其它 $u_i (i \neq k)$ 为 0 时, 函数取最小值 0。

定理1证明略。

当 $f(u_1, u_2, \dots, u_c)$ 取最小时, 出现某个 u_k 趋向于 1, 其它的 $u_i (i \neq k)$ 趋向于 0 的情况。图2显示加入隶属度约束函数后, 隶属度的明晰含义被显著提升。

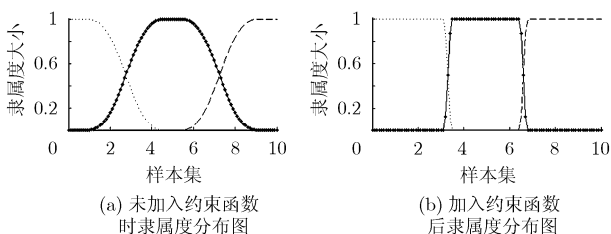


图2 隶属度函数分布图

3.2 IFP-KPC 算法

定理 2 根据拉格朗日优化理论求解式(2), IFP-KPC 算法的目标函数取最小时当且仅当

(1) 超平面 p_i 的截距 γ_i 迭代公式:

$$\gamma_i = \mathbf{w}_i^T \mathbf{X} \mathbf{u}_i^T / \mathbf{u}_i \mathbf{e} \quad (3)$$

其中 $\mathbf{u}_i = (u_{i1}^2, u_{i2}^2, \dots, u_{in}^2)$, $\mathbf{e} = \underbrace{(1, 1, \dots, 1)}_{n \times 1}^T$ 。

(2) 超平面 p_i 的单位化法向量 $\mathbf{w}_i (i = 1, \dots, c)$ 等于式(4)表示的矩阵特征值问题的最小特征值 $\tilde{\alpha}_i (i = 1, \dots, c)$ 对应的特征向量 $\tilde{\mathbf{w}}_i (i = 1, \dots, c)$:

$$\mathbf{X}(\mathbf{D}_i - \mathbf{S}_i)\mathbf{X}^T \mathbf{w}_i = \tilde{\alpha}_i \mathbf{w}_i \quad (4)$$

$$\text{其中 } \mathbf{D}_i = \begin{pmatrix} u_{i1}^2 & 0 & 0 & 0 \\ 0 & u_{i2}^2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & u_{in}^2 \end{pmatrix}, \mathbf{S}_i = \frac{\mathbf{u}_i^T \mathbf{u}_i}{\mathbf{u}_i \mathbf{e}}.$$

(3) 隶属度迭代公式:

$$u_{ij} = \frac{1}{d^2(\mathbf{x}_j, \mathbf{w}_i, \gamma_i) - a_j} \bigg/ \sum_{k=1}^c \frac{1}{d^2(\mathbf{x}_j, \mathbf{w}_k, \gamma_k) - a_j} \quad (5)$$

对于式(5), 当 $a_j > d^2(\mathbf{x}_j, \mathbf{w}_i, \gamma_i)$, $d^2(\mathbf{x}_j, \mathbf{w}_i, \gamma_i) - a_j$ 出现负值, 隶属度 u_{ij} 也会为负。故定义 $a_j = \min\{d^2(\mathbf{x}_j, \mathbf{w}_s, \gamma_s) | s \in \{1, \dots, c\}\} - \eta$, ($\eta > 0$), 代入式(5)从而得到:

$$u_{ij} = \frac{1}{d^2(\mathbf{x}_j, \mathbf{w}_i, \gamma_i) - \min_{1 \leq s \leq c} d^2(\mathbf{x}_j, \mathbf{w}_s, \gamma_s) + \eta} \bigg/ \sum_{k=1}^c \frac{1}{d^2(\mathbf{x}_j, \mathbf{w}_k, \gamma_k) - \min_{1 \leq s \leq c} d^2(\mathbf{x}_j, \mathbf{w}_s, \gamma_s) + \eta} \quad (6)$$

式(6)中, 对 η 的选择影响模糊划分的模糊程度, 同FCM算法中的控制聚类结果模糊程度的模糊指数 m 类似。 η 值越大模糊程度越高; η 值越小模糊程度越低, 越趋进于硬聚类。当 η 趋于无穷时, 样本点对每一类的隶属度均为 $1/c$ 。所以需要选取一个合适的 η 值, 当 η 取较小值时, IFP-KPC 算法基本稳定且能得到较好的聚类效果。实验中将对 η 的选择进行讨论。

下面给出改进分割的 K 平面聚类算法 IFP-KPC 算法的完整描述:

- 步骤 1 初始化聚类数目 c , 阈值 ε , 模糊程度常数 η , 最大迭代次数 T , 随机产生 u_{ij}^1 ;
- 步骤 2 根据式(4)计算超平面 p_i 的单位化法向量 \mathbf{w}_i^l, l 为迭代次数;
- 步骤 3 根据式(3)计算超平面 p_i 的截距 γ_i^l ;
- 步骤 4 根据式(5)计算样本的隶属度值 u_{ij}^{l+1} ;
- 步骤 5 如果 $\|u_{ij}^{(l+1)} - u_{ij}^{(l)}\| < \varepsilon$ 或者 $l > T$ 则停止, 算法结束; 否则 $l = l + 1$, 跳转步骤 2;
- 步骤 6 当算法收敛, 得到各类超平面 p_i 的单位化法向量 \mathbf{w}_i , 截距 γ_i 和模糊隶属度 u_{ij} 。

3.3 Voronoi 距离对 IFP-KPC 算法的鲁棒性解释

Voronoi图是对平面的一种分割, 由连接两相邻点直线的垂直平分线所构成的连续多边形组成。对于通过点集 $p = \{p_1, p_2, p_3, \dots\}$ 分割后的Voronoi图, 样本点将按照最邻近

原则分配到区域 p_i 。在 Voronoi 图中计算数据对象 \mathbf{x}_j 到第 i 类的 Voronoi 距离, 只要选择 \mathbf{x}_j 到第 i 类全部分割超平面的距离中绝对值最小的即可。即 \mathbf{x}_j 到第 i 类的 Voronoi 距离 $d_v(\mathbf{x}_j, p_i)$ [10] 为

$$d_v(\mathbf{x}_j, p_i) = \left| \min_{1 \leq s \leq c} (\mathbf{x}_j - \mathbf{h}_s)^T \mathbf{n}_s \right| \quad (7)$$

其中 $\mathbf{h}_s = (\mathbf{p}_s + \mathbf{p}_i) / 2$, $\mathbf{n}_s = \mathbf{p}_i - \mathbf{p}_s$, $1 \leq s \leq c$, $s \neq i$ 。对于式(7)定义的 Voronoi 距离, 文献[11]有如下结论:

定理 3 对于平面上通过点集 $p = \{p_1, p_2, p_3, \dots\}$ 分割后的 Voronoi 图, Voronoi 距离可以表示为 $d_v(\mathbf{x}_j, p_i) = \frac{1}{2} \cdot \left(d^2(\mathbf{x}_j, p_i) - \min_{1 \leq s \leq c} d^2(\mathbf{x}_j, p_s) \right)$ 。

通过将 Voronoi 距离与 IFP-KPC 隶属度迭代公式(6)比较可以发现, 如果将 Voronoi 距离替换模糊 KPC 算法得到的隶属度 u_{ij} 迭代公式中的欧式距离, 将得到与 IFP-KPC 算法 ($\eta = 0$ 时) 相同的隶属度迭代公式。同时文献[11]指出 Voronoi 距离和欧式距离相比, 对数据噪声点具有更少的敏感度和更好的鲁棒性, 充分说明本文提出的 IFP-KPC 算法具有鲁棒性。

4 仿真实验及实验分析

为了比较 IFP-KPC 算法的性能, 基于实验数据集分别采用 FCM 算法, KPC 算法和 IFP-KPC 算法进行仿真试验, 并用约当指数(Rand Index) [12] 评价其聚类质量。

实验 1 二维空间数据加入噪声点实验 实验 1 包含 3 类每类 50 个样本点的二维空间数据集, 每个类别的数据按照直线方程 $y = \mathbf{w}_1 x + \mathbf{w}_0$ 产生, 为了保证 3 种算法比较的准确性和公平性, 每次聚类实验都随机产生超平面的法向量 \mathbf{w}_1 和截距 \mathbf{w}_0 , 其中 $\mathbf{w}_1 \in (-1, 1)$, $x \in (-10, 10)$, $\mathbf{w}_0 \in (-10, 10)$, 并在 y 上附加范围在 $(-0.5, 0.5)$ 内的随机噪声扰动。通过在平面内加入不同数目的噪声点测试算法抗噪性能。IFP-KPC 算法中采用不同大小的参数 η 来测试其对聚类结果的影响。表 1 给出了 3 种算法分别进行 100 次聚类实验后得到的 Rand Index 的均值。

从表 1 中可以看出, IFP-KPC 算法受噪声点影响较小, 显示 IFP-KPC 算法比 KPC 算法有较好的抗噪性能。同时 $\eta = 1$ 时在所有的 η 取值中聚类效果较好。实验结果表明当 η

取较小值时 ($\eta < 10$), IFP-KPC 算法基本稳定且均能得到较好的聚类效果。FCM 算法不能对超平面分布的数据进行有效地聚类。

实验 2 高维空间大量数据实验 为了验证 IFP-KPC 算法在高维空间大量数据情况下的可扩展性, 实验 2 首先对 10 维空间大量数据情况下加入不同数目的噪声点测试 3 种算法的抗噪性能。每次实验数据集包含 2 类每类 200 个样本点的高维空间数据, 每类按照直线方程 $x_{10} = \mathbf{w}_1 x_1 + \mathbf{w}_2 x_2 + \dots + \mathbf{w}_9 x_9 + \mathbf{w}_0$ 产生。同样为了保证 3 种算法比较的准确性和公平性, 每次聚类实验都随机产生法向量 $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_9$ 和截距 \mathbf{w}_0 , 其中 $[\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_9] \in (-1, 1)$, $[x_1, x_2, \dots, x_9] \in (-10, 10)$, $\mathbf{w}_0 \in (-10, 10)$, 并在 x_{10} 上附加范围在 $(-0.5, 0.5)$ 内的随机噪声扰动。通过在 10 维空间加入不同数目的噪声点测试算法的抗噪性能。表 2 给出了 3 种算法进行 100 次聚类实验后得到的 Rand Index 的均值。

从表 2 可以看出, 在高维情况下 IFP-KPC 算法与 KPC 算法相比受噪声点影响较小, 具有较好的抗噪性能。综合表 1 和表 2 的实验结果可以得到, η 的选择应限定在一定范围内, 一般要根据样本到各个类别超平面的距离大小情况而决定, 否则若 η 太大会产生较差的聚类结果。特别当 η 趋于无穷时, 会出现样本点对每一类的隶属度均为 $1/c$ 的情况。一般情况下, 当 η 取较小值时, IFP-KPC 算法 Rand Index 基本稳定且能得到较好的聚类效果, η 的取值为所有样本点到超平面的最大距离的 0.01 到 0.2 左右较为合适。

实验 2 进一步对 3 种算法在 5 维到 30 维之间选择 6 种不同维数的数据集进行测试。每次实验的数据集由 2 类每类 200 个样本点的高维空间数据组成, 每类数据按照直线方程 $x_d = \mathbf{w}_1 x_1 + \mathbf{w}_2 x_2 + \dots + \mathbf{w}_{d-1} x_{d-1} + \mathbf{w}_0$ 产生, d 表示数据集维数。同样为了保证 3 种算法比较的准确性和公平性, 每次聚类实验都随机产生法向量 $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{d-1}$ 和截距 \mathbf{w}_0 , 其中 $[\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{d-1}] \in (-1, 1)$, $[x_1, x_2, \dots, x_{d-1}] \in (-10, 10)$, $\mathbf{w}_0 \in (-10, 10)$ 并在 x_d 附加范围在 $(-0.5, 0.5)$ 内的随机噪声扰动。实验中在超平面加入 20 个噪声点。在 IFP-KPC 算法的实验中, 参数 η 随着样本维数的不同而自适应变化, 当维数为五维时, η 取 4; 其它维数情况时, η 取 10。表 3 给出了 3 种算法进行 100 次实验后得到的 Rand Index 均值。

表 1 二维情况时 3 种算法对噪声点的 Rand Index 结果比较

| 噪声数目 | IFP-KPC 算法中 η 的数值 | | | | | | | KPC | FCM |
|------|------------------------|--------|--------|--------|---------------|--------|--------|---------------|---------------|
| | 0.0001 | 0.001 | 0.01 | 0.1 | 1 | 10 | 100 | | |
| 5 | 0.7711 | 0.8060 | 0.7760 | 0.7845 | 0.8073 | 0.7539 | 0.6195 | 0.7580 | 0.6507 |
| 10 | 0.7750 | 0.7602 | 0.7851 | 0.7729 | 0.7836 | 0.7676 | 0.6189 | 0.7463 | 0.6488 |
| 15 | 0.7486 | 0.7278 | 0.7567 | 0.7779 | 0.7723 | 0.7437 | 0.6083 | 0.7364 | 0.6488 |
| 20 | 0.7221 | 0.7245 | 0.7204 | 0.7258 | 0.7531 | 0.7566 | 0.6147 | 0.7231 | 0.6422 |
| 30 | 0.7205 | 0.7227 | 0.7385 | 0.7268 | 0.7383 | 0.7356 | 0.6283 | 0.7177 | 0.6407 |

表2 10维情况时3种算法 Rand Index 结果比较

| 算法 | 噪声数目 | | | | | | | | | |
|--------------------|------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| | 0 | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | |
| IFP-KPC 中 η 值 | 1 | 0.9088 | 0.8545 | 0.8474 | 0.7822 | 0.7363 | 0.7496 | 0.6825 | 0.6589 | 0.6273 |
| | 10 | 0.9173 | 0.8993 | 0.8422 | 0.8263 | 0.7786 | 0.7697 | 0.6958 | 0.6731 | 0.6655 |
| | 20 | 0.8872 | 0.8457 | 0.8346 | 0.8026 | 0.7447 | 0.7382 | 0.6828 | 0.6505 | 0.6288 |
| KPC | | 0.8784 | 0.8512 | 0.8042 | 0.7872 | 0.7208 | 0.6831 | 0.6615 | 0.6461 | 0.5936 |
| FCM | | 0.5493 | 0.5444 | 0.5436 | 0.5379 | 0.5374 | 0.5367 | 0.5349 | 0.5320 | 0.5299 |

表3 不同维数情况时3种算法 Rand Index 结果比较

| | 样本维数 | | | | | |
|---------|--------|--------|--------|--------|--------|--------|
| | 5 | 10 | 15 | 20 | 25 | 30 |
| IFP-KPC | 0.8700 | 0.7787 | 0.6624 | 0.6140 | 0.5696 | 0.5315 |
| KPC | 0.8338 | 0.7246 | 0.6039 | 0.5452 | 0.5162 | 0.5123 |
| FCM | 0.5419 | 0.5347 | 0.5252 | 0.5221 | 0.5114 | 0.5034 |

从表3可以看出,在高维空间大量数据情况下,IFP-KPC算法与KPC算法相比表现出更好的聚类性能。但同时也发现KPC算法和IFP-KPC算法聚类结果均受到样本空间维数的影响,随着样本维数的增加,Rand Index越来越小。由于IFP-KPC算法和KPC算法等大多数的聚类算法都基于距离机制,因此这些算法只适合于低维数据集,而在高维空间里,高维数据通常非常稀疏,在一定的数据分布条件下,对于给定目标,最近和最远邻居节点间距离几乎相等,使得传统相似性度量在高维空间中变得没有意义,导致常规的聚类算法失效^[13]。

5 结束语

本文根据K平面聚类算法KPC,提出了改进分割的K平面聚类算法IFP-KPC。通过引入隶属度约束函数,重新构造目标函数,根据拉格朗日优化理论推导出新的迭代学习公式;并利用Voronoi距离对IFP-KPC算法的鲁棒性进行了合理的解释。应当指出,IFP-KPC算法比KPC算法具有更好的鲁棒性,但其也存在不少局限性,在聚类初始化和聚类原型适用性等方面还存在很大改进的空间,在高维空间大量数据情况下还受到一定的约束。在下一步的工作中,将对IFP-KPC算法做更深入的研究,提高其性能,并把本文的思想方法应用于图像分割、文本分类、基因聚类学等领域。

参考文献

- [1] Bradley P S, and Mangasarian O L. K-plane clustering. *Journal of Global Optimization*, 2000, 16(1): 23-32.
- [2] Mojtaba Tabatabaei-Pour, Karim Salahshoor, and Behzad Moshiri. A modified k-plane clustering algorithm for identification of hybrid systems. *Proceedings of the 6th World Congress on Intelligent Control and Automation*, Dalian, China, June 2006: 1333-1337.
- [3] Washizawa Y and Cichocki A. On line k-plane clustering learning algorithm for sparse component analysis. 2006 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP2006, Toulouse, May 2006: 681-684.
- [4] Dhillon I S, Marcotte E M, and Roshan U. Diametrical clustering for identifying anti-correlated gene clusters. *Bioinformatics*, 2003, 19(13): 1612-1619.
- [5] Fischer I and Polard J. New methods for spectral clustering. Technical Report No. IDSIA-12-04, Dalle Molle Institute for Artificial Intelligence, Switzerland, 2004: 1-24. <http://www.idsia.ch/idsiareport/IDSIA-12-04.pdf>.
- [6] Georgiev P G, Theis F, and Cichocki A. Optimization algorithms for sparse representations and applications. *Nonconvex Optimization and Its Applications*, 2006, 82(6): 1-17.
- [7] Zhang Y J and Liu Z Q. Self-splitting competitive learning: A new on-line clustering paradigm. *IEEE Trans. on Neural Network*, 2002, 13(2): 369-380.
- [8] Wu Shuan-hu, Alan Wee-Chung Liew, and Yan Hong, et al. Cluster analysis of gene expression data based on self-splitting and merging competitive learning. *IEEE Trans. on Information Technology in Biomedicine*, 2004, 8(1): 5-15.
- [9] Xu Lei, Krzyzak A, and Oja E. Rival penalized competitive learning for clustering analysis. *RBF Net and Curve Detection. IEEE Trans. on Neural Network*, 1993, 4(4): 636-649.
- [10] Hoppner F. Fuzzy shell clustering algorithms in image processing: fuzzy c-rectangular and 2-rectangular shells.

- IEEE Trans. on Fuzzy Systems*, 1997, 5(4): 599-613.
- [11] Hoppner F and Klawonn F. Improved fuzzy partitions for fuzzy regression models. *Journal of Approximate Reasoning*, 2003, 32(2): 85-102.
- [12] Rand W. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 1971, 66(3): 846-850.
- [13] 杨颖, 韩忠明, 杨磊. 兴趣子空间挖掘算法在高维数据聚类中的应用. *计算机工程*, 2007, 33(2): 12-17.
- Yang Y, Han Z M, and Yang L. Application of interesting subspace mining algorithm in high-dimensional data clustering. *Computer Engineering*, 2007, 33(2): 12-17.
- 朱 林: 男, 1983 年生, 硕士生, 研究领域为图像处理、模式识别.
- 王士同: 男, 1964 年生, 教授, 研究领域为模式识别、人工智能、生物信息学.
- 潘永惠: 男, 1972 年生, 博士生, 副教授, 研究领域为模式识别、人工智能.
- 韩 斌: 男, 1968 年生, 博士生, 副教授, 研究领域为人工智能、模式识别.