

# 一种含影响力因子的自适应 C 均值算法

金 健<sup>1</sup> 黄国兴<sup>1</sup> 丁建国<sup>1</sup> 胡永涛<sup>2</sup>

(华东师范大学 上海 200062)<sup>1</sup> (国家反计算机入侵和防病毒研究中心 上海 201204)<sup>2</sup>

**摘 要** 含影响力因子的硬聚类算法 (HCMef) 在对多于两类规模不均的样本进行聚类时,影响力因子指数对聚类结果影响很大,取值不当会引起类消失。影响力因子指数在 0 附近穷举,并利用影响力因子指数较大时类规模会在较少训练次数内消失的现象,降低计算量。提出了含影响力因子的自适应 C 均值聚类策略 (AHCMef),并介绍了两阶段聚类方法,进一步提高聚类成功率和执行效率。结果显示,最佳影响力因子指数出现在成功聚类时的较大值附近,聚类效果从该点起随影响力因子指数的减小而降低。对 HCMef 算法应用于多类规模不均样本情况时,影响力因子指数的选取具有指导意义。

**关键词** 聚类,类规模,C 均值,影响力因子

## Adaptive Hard C-means Algorithm with Effectiveness Factors

JIN Jian<sup>1</sup> HUANG Guo-xing<sup>1</sup> DING Jian-guo<sup>1</sup>, HU Yong-tao<sup>2</sup>

(East China Normal University, Shanghai 200062, China)<sup>1</sup>

(National Research Center of Anti-Computer-Intrusion and Anti-Computer-Virus, Shanghai 201204, China)<sup>2</sup>

**Abstract** The hard c-means clustering algorithm with effectiveness factors (HCMef) will result in disappearance of some cluster population, when the number of clusters is larger than two and the index of the effectiveness factors is inappropriately valued. The index of effectiveness factors is taken values near zero in exhaustive manner and advantages of the cluster disappearance in few epochs are taken to reduce the amount of computation. Proposed an adaptive hard c-means algorithm with effectiveness factors, and introduced a two-phase clustering method to improve the efficiency and success rate of clustering. Results show that, the optimal index point of the effectiveness factors lies near the larger side during the successful clustering, and the clustering results worsen along with the decrease of effectiveness factor index from this point. It's of guiding significance for the selection of effectiveness factor indexes in HCMef algorithm when applied to data sets with multiple unequal sized clusters.

**Keywords** Clustering, Cluster population, C-means, Effectiveness factor

## 1 引言

专家混成系统 (Mixture of Experts, MoE) 是一个模块化的关联竞争学习模型。它基于“分而治之”思想,通过一个门网络 (Gating network) 将一个任务划分成若干个适当的子任务,每个子任务均可由一个较简单的专家网络 (Expert network) 加以解决<sup>[1]</sup>。模块化模糊神经网络 (MFNN) 采用了专家混成系统的结构,采用传统的模糊 C 均值 (FCM) 聚类算法对任务进行划分<sup>[2]</sup>。金健、金龙等人还将 MFNN 应用于气象预报,取得了良好的效果<sup>[3-5]</sup>。但 FCM 算法有个较大的缺陷就是完全基于欧氏距离,没有考虑类内样本的规模<sup>[6-8]</sup>。在对规模有差异,且靠得较近的样本进行聚类时,较小类的中心将向较大类偏移。对于像气象预报这类在高维空间中运作的复杂数据,我们无法预知其客观的内部自然分类情况。因此有必要对类规模不均的情况加以假设,并对

MFNN 中的门网络加以改进。文献[9]介绍了采用 FCMef 算法作为门网络的 MFNNef 算法,进行了一些探索。但对于 FCMef 算法的内部参数及性质仍不清楚。由于 FCM 算法的模糊性,使得对改进算法的聚类结果评价的客观性也难以把握,而且,FCM 算法中的模糊加权指数对聚类结果的影响也很大,而其最佳取值尚没有统一的方法。鉴于以上因素的考虑,文献[10]对 FCM 算法的特例 HCM 算法进行研究,介绍了含影响力因子的硬聚类算法 (HCMef),并在二维空间中对两类规模不均的样本进行试验,取得了非常明显的效果。然而,在随后的进一步研究中,作者发现该算法在处理多于两类规模不均的样本时,其中的影响力因子指数对聚类结果的影响很大;而且,如果该指数取值不当,会引起类规模消失的现象,即,聚类中心存在,但与之对应的类内样本数为 0 或很少。

本文尝试利用“类规模消失”的现象,对影响力因子指数的取值进行研究,希望找到其中的规律。

到稿日期:2008-04-29 本文受上海市科委项目 (065115023) 和国家重点基础研究发展规划 (973) 项目 (2002CB312001) 资助。

金 健 (1979-), 博士研究生,研究方向为知识发现与数据挖掘, E-mail: jyginger@163.com; 黄国兴 (1946-), 博士生导师,研究方向为智能信息系统、知识发现与数据挖掘; 丁建国 (1968-), 博士,研究方向为知识发现与数据挖掘; 胡永涛 (1976-), 副研究员, 硕士,研究方向为信息安全、知识发现和数据挖掘。

## 2 理论和方法

### 2.1 含影响力因子 HCM 算法 (HCMef) 回顾

设样本集  $P = [p_1, p_2, \dots, p_Q]^T$ ,  $p_q \in P^R$  是  $R$  维特征空间中的  $Q$  个样本,  $p_q$  表示第  $q$  个样本。则含影响力因子的 HCM 聚类算法 (HCMef) 的训练过程也就是使目标函数

$$J_1 = \sum_{q=1}^Q \sum_{c=1}^C (u_{c,q}^1 \cdot (\omega_c \cdot d_{c,q})^2) \quad (1)$$

最小化的优化过程。其中,  $C$  是类别数,  $u_{c,q}$  满足约束条件

$$\sum_{c=1}^C u_{c,q} = 1, \forall q \quad (2)$$

它表示样本  $p_q$  是否属于类别  $c$ , 即

$$u_{c,q} = \begin{cases} 1, & p_q \text{ belong to cluster } c \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

$d_{c,q}$  指样本  $p_q$  与第  $c$  类的聚类中心  $v_c$  的距离或不相似度, 通常选用欧氏距离, 即

$$d_{c,q} = \|p_q - v_c\|_2 \quad (4)$$

$\omega_c$  是类别  $c$  的影响力因子, 我们将它定义为

$$\omega_c = \left(\frac{1}{\Omega_c}\right)^\zeta \quad (5)$$

其中,  $\zeta$  是影响力因子指数, 它控制影响力因子的作用强度。特别地, 当  $\zeta = 0$  时,  $\omega_c = 1$ 。此时, HCMef 算法退化为传统的 HCM 算法。因此, HCM 聚类算法实际上是 HCMef 的一个特例。  $\Omega_c$  表示类别  $c$  的规模。类别的规模可以有多种不同的定义, 例如密度、样本数, 或者两者的结合等等。本文中, 我们定义类别  $c$  的规模为

$$\Omega_c = \sum_{q=1}^Q u_{c,q} \quad (6)$$

也就是类别  $c$  包含的样本数。

下面以定理的形式给出 HCMef 算法中隶属度及聚类中心的更新规则。

**定理 1 (HCMef 算法隶属度更新规则)** 聚类中心  $V_{C \times R}$  和影响力因子  $\omega$  已知, 样本  $p_q$  到聚类中心  $v_c$  的距离记为  $d_{c,q}$ , 则 HCMef 算法的隶属度更新规则为

$$u_{c,q} = \begin{cases} 1, & \text{if } (\omega_c \cdot d_{c,q}) = \min_{i \in \{1, 2, \dots, C\}} (\omega_i \cdot d_{i,q}) \\ 0, & \text{otherwise} \end{cases}, \forall c, q \quad (7)$$

证明: HCMef 算法的求解过程是使目标函数式(1)在满足约束式(2)的情况下的最小化过程, 且隶属度的取值为

$$u_{c,q} \in \{0, 1\}, \forall c, q \quad (8)$$

考察目标式(1)中的子式

$$\sum_{c=1}^C (u_{c,q}^1 \cdot (\omega_c \cdot d_{c,q})^2) \quad (9)$$

并结合式(2)和式(9)可知, 对于某样本  $p_q$ , 其对  $C$  个类的隶属度有且只有一个为 1, 其余均为 0。要使式(9)最小化, 只需当  $d_{c,q}$  最小时,  $u_{c,q} = 1$ 。如果对于所有的  $Q$  个样本均满足该规则, 就可以使目标函数式(1)最小化, 即得到式(7)。证毕。

**定理 2 (HCMef 算法聚类中心更新规则)** 若隶属度  $U_{C \times Q}$  已知, 则 HCMef 算法的聚类中心为

$$v_c = \frac{\sum_{q=1}^Q (u_{c,q}^1 \cdot p_q)}{\sum_{q=1}^Q u_{c,q}^1}, \forall c \quad (10)$$

证明: 为了使目标函数式(1)最小化, 使其对第  $c$  类聚类中心的  $r$  分量求微商, 并令其等于 0, 得

$$\begin{aligned} \frac{\partial J_1}{\partial v_{c,r}} &= \frac{\partial \sum_{q=1}^Q \sum_{c=1}^C (u_{c,q}^1 \cdot (\omega_c \cdot d_{c,q})^2)}{\partial v_{c,r}} \\ &= \sum_{q=1}^Q \frac{\partial \sum_{c=1}^C (u_{c,q}^1 \cdot (\omega_c \cdot d_{c,q})^2)}{\partial v_{c,r}} \\ &= \sum_{q=1}^Q \frac{\partial (u_{c,q}^1 \cdot (\omega_c \cdot d_{c,q})^2)}{\partial v_{c,r}} \\ &= \omega_c^2 \cdot \sum_{q=1}^Q (u_{c,q}^1 \cdot \frac{\partial d_{c,q}^2}{\partial v_{c,r}}) \\ &= \omega_c^2 \cdot \sum_{q=1}^Q (u_{c,q}^1 \cdot \frac{\partial \sum_{i=1}^R (v_{c,i} - p_{q,i})^2}{\partial v_{c,r}}) \\ &= 2\omega_c^2 \cdot \sum_{q=1}^Q (u_{c,q}^1 \cdot (v_{c,r} - p_{q,r})) = 0, \forall c, r \quad (11) \end{aligned}$$

进一步得出

$$\sum_{q=1}^Q (u_{c,q}^1 \cdot v_{c,r}) - \sum_{q=1}^Q (u_{c,q}^1 \cdot p_{q,r}) = 0, \forall c, r \quad (12)$$

从而计算出

$$v_{c,r} = \frac{\sum_{q=1}^Q (u_{c,q}^1 \cdot p_{q,r})}{\sum_{q=1}^Q u_{c,q}^1}, \forall c, r \quad (13)$$

写成向量形式即为式(10)。证毕。

HCMef 算法是个迭代过程(见算法 1)。

#### 算法 1 HCMef 聚类算法

- 1 begin
- 2 初始化参数: 类别数  $C$ , 最大训练次数  $K_u$ , 训练中误差  $\epsilon_u$ , 影响力因子指数  $\zeta$ , 隶属度矩阵初始值  $U$
- 3 for  $k \leftarrow 1$  to  $K_u$  do
- 4 将隶属度  $U$  作为已知, 根据定理 2 更新聚类中心  $V$ , 根据式更新影响力因子  $\omega$
- 5 将隶属度  $U$  和聚类中心  $V$  作为已知, 求出目标函数值  $J^{(k)}$
- 6 将聚类中心  $V$  作为已知, 根据定理 1 更新隶属度矩阵  $U$
- 7 if  $k > 1$  且  $|J_m^{(k)} - J_m^{(k-1)}| < \epsilon_u$  then break
- 8 end
- 9 end

### 2.2 自适应 HCMef 策略 (AHCMef)

文献[10]对二维二类不同规模的样本集进行的试验表明, HCMef 算法具有较好的适应性和稳定性。对于边界较明显的分类, HCMef 算法可以比传统的 HCM 算法更快找到类中心; 对于边界不太明显的分类, HCMef 算法可以比传统的 HCM 算法找到的中心更准确, 而且所需时间跟 HCM 算法差不多, 甚至更少。

然而, 在对 HCMef 算法的进一步研究中我们发现, 按照目前类别规模的定义, 如果影响力因子指数过小, 则 HCMef 算法退化成 HCM 算法, 影响力因子失去作用; 而如果影响力因子指数过大, 则会出现规模较小类别的消失现象, 表现为所有样本对该类隶属度为 0。由于类规模定义为类内样本数, 因此同一样本集各类规模会因类别数的不同而不同, 不同样本数的各样本集中各类规模更会表现出较大差异。

类别规模对影响力因子起直接作用, 而影响力因子指数可以起到调节影响力因子的作用。我们通过实验发现, 较合适的影响力因子指数一般不超过 3, 且通常分布在 0 和 1 之间。本文尝试利用不合适的影响力因子指数引起类别消失的现象, 在一个宽范围内自动寻找合适的影响力因子指数, 建立自适应 HCMef 聚类策略(如算法 2 所示)。

## 算法2 自适应 HCMef 聚类策略 (AHCMef)

```

1 begin
2 初始化全局参数:影响力因子指数序列  $\zeta$ ,类内最少样本数  $Q_{\min}$ ,
  隶属度初始化尝试次数  $K_{\text{try}}$ ; HCMef 算法参数:最大训练次数
   $K_u$ ,确保完成的训练次数  $K_{\text{assure}}$ ,训练中止误差  $\epsilon_u$ 
3  for  $\zeta \in \zeta$  (从大到小依次选取) do
4    for  $k_{\text{try}} \leftarrow 1$  to  $K_{\text{try}}$  do
5      获得初始化隶属度矩阵  $U$ 
6      for  $k \leftarrow 1$  to  $K_u$  do
7        将隶属度  $U$  作为已知,根据定理2更新聚类中心  $V$ ,根据
          式(5),(6)更新影响力因子  $\omega$ 
8        将隶属度  $U$  和聚类中心  $V$  作为已知,求出目标函数值
           $J^{(k)}$ 
9        将聚类中心  $V$  作为已知,根据式定理1更新隶属度矩阵
           $U$ 
10       if  $k > K_{\text{assure}}$  then
11         统计  $C$  个类中各类样本数  $[Q_1, Q_2, \dots, Q_C]^T$ 
12         if  $\forall Q_c < Q_{\min}$  then
13           stop_reason = 类别消失
14           break
15         else if  $|J_m^{(k)} - J_m^{(k-1)}| < \epsilon_u$  then
16           stop_reason = 达到最小误差
17           break
18         end
19       end
20     end
21   if  $k = K_u$  then
22     stop_reason = 达到最大训练次数
23   end
24 end
25 end
26 end

```

该策略中有几个需要注意的地方:

第3行,影响力因子指数的选取。如果整个策略要把所有的解(即,  $\text{stop\_reason} \neq$  类别消失)记录下来,则选取次序可以任意。如果策略中找到第一个解以后不再寻找,则建议对影响力因子指数从大到小选取。因为影响力因子指数越大,类别越易消失;反之就是 HCM 算法,通常可以一次便找到解,失去了影响力因子的作用,效果不一定好。

第10行,  $K_{\text{assure}}$  是 HCMef 算法中确保完成的训练次数。由于初始隶属度矩阵的不确定性,有可能在初始的时候算法就满足停止训练的条件,致使训练过程无法继续。而事实上,该局面有可能通过少数几次训练后得以扭转,因此设立此参数加以调节。

第12行,  $Q_{\min}$  是定义的类别内最少样本数。类别的消失不一定要到类内样本数为0时才成立,设立该参数增加“消失”的灵活性。

第14行, break 指跳出循环。当算法完成了确保完成的训练次数后,最少类别数仍未达到预设值,则认为该类别消失了,没有必要再训练下去。及时退出训练可以缩短整个算法的运行时间。

### 2.3 两阶段聚类

HCM 算法同 FCM 算法一样,聚类结果有使各类规模均分的趋势,致使小类中心向大类偏移。而这种偏移虽然导致

了聚类结果的不精确,但 HCM 算法中聚类中心从初始值到训练结束的演变趋势还是正确的。如果在 HCMef 算法中也采用跟 HCM 算法相同的初始化方法,则必然增加其寻找正确类中心的不确定性。因此,我们考虑两阶段聚类策略:先使用传统的 HCM 算法进行粗聚类,将此聚类结果用于 HCMef 算法的初始化。所以可对算法2作如下修改:第2行,初始化完成后,用 HCM 算法进行  $K_{\text{try}}$  次聚类,并记录各次训练结束时的隶属度矩阵  $\tilde{U}_{C \times Q \times K_{\text{try}}}$ ;第5行,初始化  $U$  从第2行取得的  $K_{\text{try}}$  次隶属度矩阵中依次选取。

## 3 实验及分析

### 3.1 数据集

现实应用中的数据集往往含有不可控制的噪声,它们适合在测试算法的健壮性时使用,而在测试算法的有效性时,使用这些数据反而会掩盖算法的真实效果。为了能考察本文算法的真实性能,本文采用的实验数据为人工数据。

表1 数据集正态分布概率参数

类别号	均值	方差	样本数
1	(0, 0°)	$\sqrt{10}$	500
2	(11, -30°)	$\sqrt{7}$	350
3	(10, -90°)	$\sqrt{5}$	250
4	(9, 30°)	$\sqrt{3}$	150
5	(8, 90°)	$\sqrt{2}$	100

文献[10]已对二维两类不同规模的数据集进行了试验,本文尝试对多类不同规模的数据集进行聚类试验。本文实验数据集中包含5个不同规模的类别,5类数据均按标准正态概率分布随机生成,随机数产生的各种参数如表1所示。其中,均值使用了极坐标表示方式。该数据集在二维平面上的分布如图1所示。

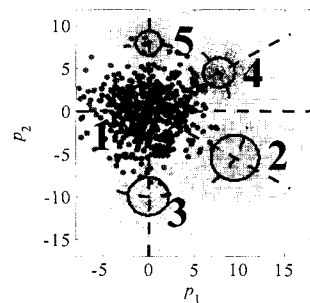


图1 实验数据集在二维空间中的分布

注:  $p_1$  和  $p_2$  分别表示坐标的第1和第2维;5个实心圆是以各类均值(虚线的交点)为中心,方差为半径画出;实心圆旁的数字表示类别编号。

### 3.2 实验参数及结果分析

本实验中用到的参数及取值如表2所示,分两次进行试验。第1次实验(1-AHCMef)中,对算法2第5行采用随机数进行隶属度初始化;第2次实验(2-AHCMef)采用两阶段聚类策略对隶属度进行初始化。其中,两阶段聚类策略中,进行初始隶属度使用的 HCM 算法参数为  $K_u = 100, \epsilon_u = 1e-6$ 。

表2 AHCMef 实验参数设置

参数名	取值	参数意义
C	5	类别数
$\zeta$	[1,0.9,...,0.1,0]	影响力因子指数序列
$Q_{\min}$	50	类内最少样本数
$K_{\text{try}}$	50	尝试不同隶属度次数
$K_u$	100	HCMef 算法最大训练次数
$K_{\text{assure}}$	3	HCMef 算法确保完成的训练次数
$\epsilon_u$	1e-6	HCMef 算法中止训练误差

实验结果有 4 项:

第 1 项,统计各种影响力因子指数取值下,1-AHCMef 和 2-AHCMef 成功聚类的次数。这里的“成功聚类”是指类别未消失,也即,HCMef 算法中止训练的原因为“达到中止训练误差”或“达到最大训练次数”。另外还有一种情况也认为是聚类失败,请参考实验结果第 3 项中的说明。本项结果如图 2 所示。

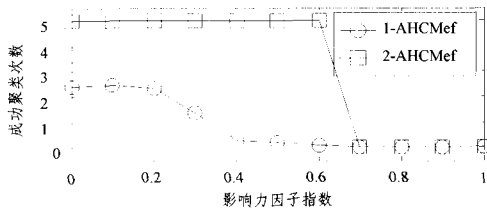


图 2 成功聚类次数比较

从图 2 可见,AHCMef 算法中,聚类成功率随着影响力因子指数的增大而降低,甚至会降为 0。1-AHCMef 实验的聚类成功率下降趋势较平缓,即使是影响力因子指数减小为 0,即退化成传统的 HCM 算法,其聚类成功率也不高,大约是 50%。可见,初始隶属度对 HCMef 算法的影响非常大。而采用两阶段聚类的 2-AHCMef 实验的聚类成功率明显高于 1-AHCMef。

第 2 项,统计成功聚类时各影响力因子指数取值下 HCMef 算法的平均训练次数。结果如图 3 所示。

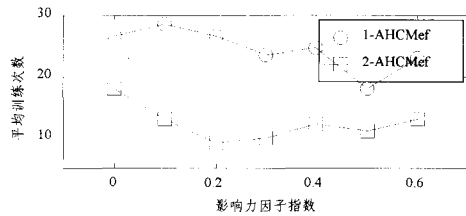


图 3 成功聚类时平均训练次数比较

图 3 显示,2-AHCMef 实验的平均训练次数普遍少于 1-AHCMef。尽管 2-AHCMef 在算法开始前执行了  $K_{\text{try}}$  次 HCMef 过程,但总体复杂度仍和 1-AHCMef 相同。对照片算法 2,1-AHCMef 执行 HCMef 算法的次数为  $K_{\zeta} \cdot K_{\text{try}}$ ,其中, $K_{\zeta}$  指影响力因子指数序列  $\zeta$  的长度,而 2-AHCMef 执行 HCMef 算法的次数为  $K_{\zeta} \cdot K_{\text{try}} + K_{\text{try}} = (K_{\zeta} + 1) \cdot K_{\text{try}}$ 。所以,两者的算法复杂度在同一数量级。因此,2-AHCMef 的收敛速度要大于 1-AHCMef,效率较高。

第 3 项,统计成功聚类时各影响力因子指数取值下各次 HCMef 聚类结果中的各类中心与原始数据产生时的均值(我们认为是客观上的类中心)的误差,以欧氏距离来度量。结果如图 4 所示。

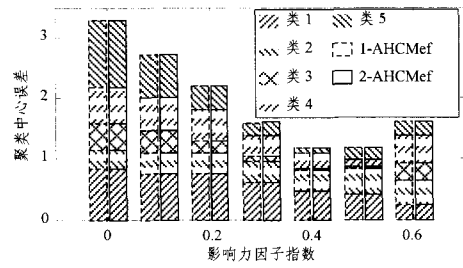


图 4 成功聚类时各聚类中心误差比较

由于 HCMef(包含 HCM)算法聚类得到的中心次序是随机的,而在统计实验结果时为了计算平均值需要将它们的次序统一。原始数据均值代表的类中心矩阵表示为  $\hat{V}_{C \times R}$ ,聚类所得的乱序的类中心表示为  $V_{C \times R}$ 。本文采用的统一类中心次序的方法为:分别求出均值类中心  $\tilde{v}_i$  与乱序类中心  $v_i$  的误差最小者  $c_i$ 。若  $c_i \neq c_j, \forall i, j | i, j \in \{1, 2, \dots, C\}, i \neq j$ ,则认为聚类成功,但精度不确定;否则,则认为聚类失败。当聚类成功时, $c_i$  的值表示乱序的类中心  $v_{c_i}$  与均值类中心  $\tilde{v}_i$  最接近,应作为第  $i$  类。

第 4 项,统计成功聚类时各影响力因子指数取值下各次 HCMef 聚类结果中各类错分样本数占总样本数的百分比。第  $c$  类错分样本数包括两种样本:一种是应该是第  $c$  类的样本却划分到了其它类;另一种是本该是其它类的样本却被划分给了第  $c$  类。实验结果如图 5 所示。

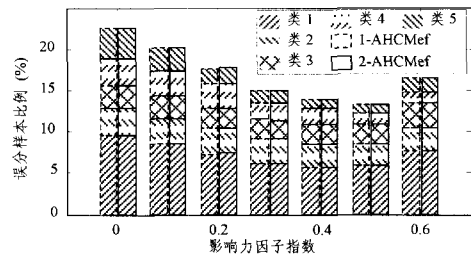


图 5 成功聚类时各类别误分样本比例比较

第 3 项结果反映了聚类中心的准确性。由于 HCMef 算法是专门针对样本类别规模不均的情况设计,因此,不同于传统的 HCM 算法,只要聚类中心准确,则类别划分也准确,在 HCMef 算法中,会出现类中心准确,但其类内样本已消失的现象。因此,HCMef 算法的准确性还需要考察类别划分情况,即第 4 项结果。由图 4 和图 5 可见,聚类中心和类别误分情况的变化趋势基本一致,即,误差随影响力因子指数的增大而减小,在影响力因子指数取值为大于 0.5,并在聚类失败前又有小幅度上升。另外,1-AHCMef 和 2-AHCMef 的误差几乎完全相同。这说明只要能成功聚类,普通的 AHCMef 算法和两阶段式 AHCMef 算法取得的结果是相同的。

结束语 本文对含影响力因子的自适应 C 均值算法(AHCMef)进行研究,并对随机数式隶属度初始方法和两阶段式隶属度初始方法对聚类结果产生的影响作了比较分析。结果显示,含影响力因子的自适应 C 均值算法对多类规模不均的样本的聚类效果随影响力因子指数的变化呈现较一致的规律性。即,聚类成功率随影响力因子指数的减小而提高,且最佳影响力因子指数出现在成功聚类时的较大值者附近。因此,在实际操作中可不必对影响力因子指数用穷举方式进行试探,而是从大到小有选择地进行,可进一步降低复杂度,提

高效率。但最佳影响力因子指数的最佳定位仍需要作进一步研究。另外,通过比较我们还发现,聚类成功的情况下,两种形式的 AHCMef 算法的聚类效果基本相同,但两阶段聚类方式的 AHCMef 算法在成功率和效率方面更优越。C 均值聚类是模糊 C 均值聚类的一种特例,因此今后可对含影响力因子的自适应 C 均值作进一步的推广研究。

### 参考文献

[1] Jacobs R. Adaptive mixtures of local experts[J]. Neural Computation, 1991, 3(1):79-87  
 [2] Langari R, Wang L. Fuzzy models, modular networks, and hybrid learning[J]. Fuzzy Sets and Systems, 1996, 79(2):141-150  
 [3] 金龙,林熙,金健,等. 模块化模糊神经网络的数值预报产品释用预报研究[J]. 气象学报, 2003, 61(1):78-84  
 [4] 王晶晶,金健,邱新法. 模块化模糊神经网络训练策略研究[J]. 南京气象学院学报, 2007, 30(2):274-278  
 [5] Jin L, Jin J, Yao C. A Short-term Climate Prediction Model

Based on a Modular Fuzzy Neural Network [J]. Advances in Atmospheric Sciences, 2005, 22(3):428-435

[6] Bensaid A M, Hall L O, Bezdek J C, et al. Partially Supervised Clustering for Image Segmentation [J]. Pattern Recognition, 1996, 29(5):859-871  
 [7] 刘小芳,曾黄麟,吕炳朝. 部分监督加权模糊 C-均值算法的聚类分析 [J]. 计算机仿真, 2005, 22(003):114-116  
 [8] 钮永莉,陈水利, Niu Y, 等. 模糊 C 均值算法的改进 [J]. 模糊系统与数学, 2004, 18:304-308  
 [9] Jin J, Wang J-J, Huang G-X, et al. Study on Influence of Effectiveness Factors on Modular Fuzzy Neural Network Based Prediction Model[C]//Sixth World Congress on Intelligent Control and Automation, Conference Proceedings (WCICA 2006). Dalian, China, 2006  
 [10] 金健,黄国兴,梁道雷. 二维空间中硬聚类算法影响力因子的作用研究[J]. 计算机科学, 2006, 33(10):182-185

(上接第 217 页)

小的系数的连乘积,然后再逐次放大。如果有  $n$  个较小系数,前  $n-1$  个取为 2,最后一个为大于 1 小于 2 的正整数。例如当  $k=5$  时,我们可以分解为  $k=2 \times 2 \times 1.25$ ,这样我们需要进行 3 次插值。

### 4 实验结果

我们通过截取一幅大图像中的模糊细节部分(见图 4 用白框选择的部分)进行放大,来检测本文算法的效果。通过比较最近邻插值方法、双线性插值方法、双三次插值和本文算法,结果表明我们提出的算法克服了经典插值后图像的边缘锯齿和模糊效应,使修正优化后的图像边界清晰(图 5)。

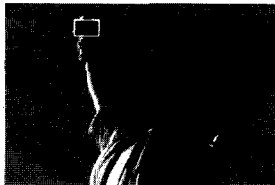
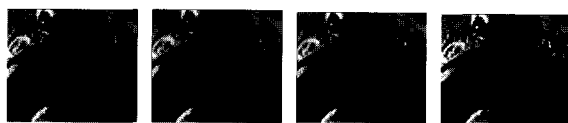


图 4 待放大的源图像



(a)最近邻插值结果 (b)双线性插值结果 (c)双三次插值结果 (d)本文算法结果

图 5 局部被放大后得到的图像

**结束语** 经典插值算法的平滑功能使放大后的图像边缘区域变宽,产生模糊和锯齿效应。本文针对此缺陷,充分挖掘经典插值算法放大后图像自身所包含的信息,提出了一种基于切向量控制细分格式的图像插值算法,它不仅可以放大图像到任意倍数,而且可以较好地保持图像的边界特征,解决了传统方法中出现的插值所得图像边界模糊,有阶梯块状的问题。

### 参考文献

[1] Thvenaz P, Blu T, Unser M. Image Interpolation and Resam-

pling. Handbook of medical imaging, Orlando, FL, USA: Academic Press, Inc., 2000

[2] Watson D F. Contouring: A Guide to the Analysis and Display of Spatial Data. New York: Pergamon Press, 1992  
 [3] Smith M R, Nichols S T. Efficient algorithms for generating interpolated (zoomed) mr images. Magnetic Resonance in Medicine, 1988, 1:156-171  
 [4] Unser 1988, Aldroubi A, Eden M. Enlargement or reduction of digital images with minimum loss of information. IEEE Transactions on Image Processing, 1995, 4:247-258  
 [5] Lehmann TM, Gonner C, Spitzer K. Survey: Interpolation methods in medical image processing. IEEE Transaction on medical imaging, 1999, 18:1049-1075  
 [6] Meijering E, Zuiderveld K J, Viergever M A. Image reconstruction by convolution with symmetrical piecewise nth order polynomial kernels. IEEE Transactions on Image Processing, 1999, 8:192-201  
 [7] Caselles V, Morel J M, Sbert C. An axiomatic approach to image interpolation. IEEE Transactions on Image Processing, 1998, 7:376-386  
 [8] Jensen K, Anastassiou D. Subpixel edge localization and the interpolation of still images. IEEE Transactions on Image Processing, 1995, 4:285-295  
 [9] Leung W Y V, Bones P J, Lane R G. Statistical interpolation of sampled images. Optical Engineering, 2001, 40:547-553  
 [10] Hou H, Andrew H. Cubic splines for image interpolation and digital filtering. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1978, 26:508-517  
 [11] Hu M, Tan J Q. Adaptive oscillatory rational interpolation for image processing. Journal of Computational and Applied Mathematics, 2006, 195:46-53  
 [12] Hong S-H, Park R-H, Yang S J, et al. Image interpolation using interpolative classified vector quantization. Image Vis. Comput., 2007, doi:10.1016  
 [13] Dyn N, Levin D, Gregory J A. A 4-point interpolatory subdivision scheme for curve design. Computer Aided Geometric Design, 1987, 4:257-268