

文章编号:1001-9081(2008)05-1197-03

## 一种遗传模糊聚类算法及其应用

宋 娇, 葛临东

(信息工程大学 信息工程学院, 郑州 450002)

(test0826@test0826@yahoo.com.cn)

**摘要:** 研究一种基于遗传算法(GA)的模糊聚类方法, 即将遗传算法得到的聚类中心作为模糊 C-均值(FCM)聚类算法初值, 这样既可以克服 FCM 算法对初始中心敏感的缺点, 也可以解决遗传算法只能找到近似解的问题。将算法用于通信信号的星座聚类, 根据聚类有效性函数自适应地确定聚类中心, 并完成信号类型的识别。仿真实验证明, 当存在较小的定时误差时, 算法对 PSK 和 QAM 信号仍然是有效的。

**关键词:** 遗传算法; 模糊 C-均值; 星座聚类

**中图分类号:** TP391.4    **文献标志码:**A

## Fuzzy clustering algorithm based on genetic algorithm and its application

SONG Jiao, GE Lin-dong

(Institute of Information Engineering, Information Engineering University, Zhengzhou Henan 450002, China)

**Abstract:** A fuzzy clustering arithmetic based on Genetic Algorithm (GA) was proposed. Genetic algorithms instructed to choose the initial cluster centers firstly, and clustering was processed by Fuzzy C-Mean Clustering (FCM). It not only overcomes the sensitivity to initial centers as FCM, but also solves the problem of approximate solution as Genetic Algorithm. This method was applied to constellation clustering and modulation recognition. The simulation results show that the method is feasible to PSK and QAM signal even when timing error exists.

**Key words:** Genetic Algorithm (GA); Fuzzy C-Mean (FCM); constellation clustering

## 0 引言

遗传算法是模拟自然界遗传机制和生物进化论而形成的一种搜索最优解的算法。它以适应度为目标函数, 使用群体搜索技术, 通过对当前群体施加选择、交叉、变异等一系列遗传操作, 从而产生出新一代的群体, 并逐步使群体进化到包含或接近最优解的状态。

模糊 C-均值聚类算法(Fuzzy C-Mean, FCM)是一种典型的动态聚类算法, 在图像处理、数据挖掘、人工智能等方面均有较多应用。但是普通的模糊聚类是一种局部搜索算法, 如果初始值选择不当, 它就会收敛到局部极小点上。所以本文先运用遗传算法得到聚类中心, 然后用改进的模糊聚类算法得到最优解。这样将两者结合起来进行处理, 既克服了 FCM 算法对初始值敏感和容易陷入局部最优解的缺陷, 又能解决遗传算法只能找到近似最优解的问题, 从而得到理想的结果。

信号类型识别是现代通信领域中的一个关键技术, 在数字幅相调制信号的识别中, 可以对信号星座图进行处理, 将形状识别与调制识别相结合, 获取关键特征。本文以此为依据, 将基于遗传算法的模糊聚类用于通信信号的星座图重构上, 进而实现调制类型的识别。在聚类中心数的调整过程中进行聚类有效性分析, 自适应地确定聚类中心的个数。该算法减小了初始聚类中心的随机性, 有一定的抗噪声能力, 对于存在一定定时误差的星座, 仍能够进行星座重构识别。

## 1 基于遗传算法的模糊聚类算法

### 1.1 FCM 算法<sup>[1]</sup>

首先简单介绍 FCM 算法。它是用隶属度函数来确定每

个数据属于某个聚类的程度。设样本序列的总样本数为  $n$ , 聚类数目为  $c$  ( $2 \leq c \leq n-1$ )。FCM 算法通过优化模糊目标函数得到每个样本点对聚类中心的隶属度, 其目标函数定义为:

$$J_m(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m d_{ij}^2 \quad (1)$$

其中  $\{\mathbf{x}_i, i = 1, 2, \dots, n\}$  为样本集合,  $\mathbf{v} = (v_1, v_2, \dots, v_c)$  为聚类中心矩阵,  $\mathbf{u}$  为隶属度矩阵,  $u_{ij}$  为样本  $\mathbf{x}_i$  隶属于中心  $v_i$  的隶属度, 满足  $0 \leq u_{ij} \leq 1, \forall i, j$ ,  $\sum_{i=1}^c u_{ij} = 1, \forall j$ ,  $d_{ij} = \|\mathbf{x}_i - v_i\|$  采用欧氏距离度量,  $m \in (1, \infty)$  为权重指数, 一般  $m = 2.0$ 。

样本的最佳模糊 C-划分即是  $J_m$  的最小方差稳定点( $\mathbf{u}, \mathbf{v}$ )。由条件  $\min\{J_m(\mathbf{u}, \mathbf{v})\}$  得到模糊隶属度和类中心矢量的迭代式:

$$u_{ij} = \begin{cases} \sum_{k=1}^c \left( \frac{d_{ik}}{d_{jk}} \right)^{2/(m-1)}, & d_{ij} > 0, \\ 1, & d_{ij} = 0 \end{cases}, \quad v_i = \frac{\sum_{j=1}^n u_{ij}^m \mathbf{x}_j}{\sum_{j=1}^n u_{ij}^m} \quad (2)$$

FCM 算法整个计算过程就是以目标函数  $J_m$  的梯度下降方向为迭代方向, 反复修改聚类中心和分类矩阵的过程。但算法容易陷入局部极值点, 对初始聚类中心和样本输入次序非常敏感, 聚类结果具有很强的随机性。

### 1.2 遗传算法<sup>[2,3]</sup>

用遗传算法求解聚类问题, 首先要解决三个问题:

- 1) 如何将聚类问题解编码到基因串中;
- 2) 如何构造适应度函数来度量每条基因串对聚类问题的适应度, 即如果某条基因串的编码代表着良好的聚类结果, 则其适应度就高, 否则, 其适应度低;

收稿日期: 2007-11-17; 修回日期: 2008-01-23。 基金项目: 总装备部重点科研基金项目(6130320)。

作者简介: 宋娇(1982-), 女, 辽宁大连人, 硕士研究生, 主要研究方向: 通信信号调制识别; 葛临东(1946-), 男, 山东济南人, 教授, 博士生导师, 主要研究方向: 软件无线电、信号分析与处理。

3) 如何选择各个遗传算子,即如何进行选择、交叉、变异。

下面对本文采用的方法分别加以简单说明。

### 1.2.1 编码

以各聚类中心的编码为个体,采用浮点数多参数级联编码。设样本数据维数为  $d$ (星座聚类中,  $d = 2$ ),类别中心数为  $c$ ,则编码长度为  $d \times c$ ,个体  $\mathbf{a}$  表示为:

$$\mathbf{a} = \underbrace{a_{11} a_{12} \cdots a_{1d}}_{m_1} \underbrace{a_{21} a_{22} \cdots a_{2d}}_{m_2} \cdots \underbrace{a_{c1} a_{c2} \cdots a_{cd}}_{m_c} \quad (3)$$

### 1.2.2 适应度函数

对于 FCM 算法,其最优聚类结果对应目标函数的最小值,即聚类效果越好,则目标函数越小,而此时适应度应越大。因此,这里个体的适应度函数可以借助 FCM 算法的目标函数  $J_m$  来定义为:

$$f = 1/J_m \quad (4)$$

### 1.2.3 非线性排序选择

排序选择能够在一定程度上避免进化的早熟收敛,本文采用非线性排序选择。首先根据个体的适应度进行从大到小排序,排在前面(适应度较大)的个体有较多的被选择机会,其选择概率函数如下:

$$eval(\mathbf{a}_i) = q(1 - q)^{rank_i - 1} \quad (5)$$

其中,  $rank_i$  是个体  $\mathbf{a}_i$  的排序号,参数  $q \in [0, 1]$  表示算法隐含的选择压力,本文取  $q = 0.1$ 。

### 1.2.4 交叉和变异

文中选用均匀算术交叉算子和均匀变异算子:

$$\begin{cases} \mathbf{a}'_i = b \cdot \mathbf{a}_j + (1 - b) \cdot \mathbf{a}_i \\ \mathbf{a}'_j = b \cdot \mathbf{a}_i + (1 - b) \cdot \mathbf{a}_j \\ \mathbf{a}'_l = U_{\min}^l + r \cdot (U_{\max}^l - U_{\min}^l) \end{cases} \quad (6)$$

其中,  $\mathbf{a}_i, \mathbf{a}'_i$  分别表示父代与子代个体,  $b$  为线性组合系数,  $\mathbf{a}'_l \in [U_{\min}^l, U_{\max}^l]$  为交叉产生的子代个体的变异点,  $r$  为  $[0, 1]$  内均匀分布的随机数。

### 1.2.5 群体的更新策略

群体进化过程中,当个体的某个聚类中心偏离数据分布区域时,所有样本点都可能认为不属于该类别,则解无效。选择合适的遗传算法更新策略可以剔除无效解。

从数学的角度讲,允许父代中的优良个体进入下一轮竞争确保了最优解的迭代稳定性,而将子代中最差的个体提前淘汰出局加速了寻优过程。因此,在进入下一轮进化前,用父代中的有效解替换子代中的无效解,然后实施精英保留策略<sup>[4]</sup>,这种群体更新方式能够避免无效解的扩散,保证算法收敛。

综上,遗传聚类算法具体步骤如下:

- 1) 初始化遗传算法参数(最大进化代数  $T_{\max}$ 、种群规模、交叉概率、变异概率)和聚类中心数  $c$ ;
- 2) 进化代数计数器置零  $t = 0$ ,随机产生  $K$  个个体的初始种群  $P(t)$ ;
- 3) 计算个体目标函数,进行适应度评价;
- 4) 遗传操作:选择,交叉,变异;
- 5) 无效解判定,群体更新,  $t = t + 1$ ;
- 6) 若  $t > T_{\max}$ ,则聚类结束,否则转至 3),继续进化。

### 1.3 聚类有效性分析

聚类有效性在聚类分析中有重要的研究价值,它用来度量各聚类之间的独立性和聚类中模式的紧密性。文献[5]详细叙述了描述聚类性能的指标。在该基础上,我们假定聚类数为  $c$ ,  $\mathbf{v}_i$  表示第  $i$  个聚类中心,聚类有效性函数  $S$  定义为<sup>[6]</sup>:

$$S = \frac{sep}{comp} = \frac{c \cdot \min_{i, j} \|\mathbf{v}_i - \mathbf{v}_j\|^2}{\sum_{i=1}^c \sum_{j=1}^n \delta_{ij} \|\mathbf{x}_j - \mathbf{v}_i\|^2} \quad (7)$$

其中,  $sep$  表示聚类的分离性;  $comp$  表示聚类的紧密性,体现了  $c$  个聚类的平均离散度。聚类越独立,  $sep$  越大,  $S$  也越大,聚类越紧密,  $comp$  越小,  $S$  也越大;这样最大化  $S$  就代表了一个有效的最优划分。参量  $\delta_{ij} = \begin{cases} 1, & u_{ij} > u_{kj}, i \neq k \\ 0, & \text{其他} \end{cases}$  用以剔除距离聚类中心较远、对聚类的紧密性贡献不大的数据的影响。

### 2 聚类算法在信号识别中的应用

数字调制信号能够被它的星座图表述,因此在接收端通过恢复的星座图可以对信号进行识别。本文以此为依据,将前面描述的算法应用于信号星座图重构,进而实现信号类型的识别。图 1 分别显示了信噪比为 20 dB 的八进制移相键控(8 Phase-Shift Keying, 8PSK)和八进制正交振幅调制(8 Quadrature Amplitude Modulation, 8QAM)信号的星座图(图中横坐标表示信号的同相分量,纵坐标表示信号的正交分量),显然,星座图的聚类中心可以确定信号的调制方式。图 1 也给出了存在定时误差时的信号星座图,此时星座点更分散,不利于实现正确聚类,但在一定的误差范围内,算法仍然有效。

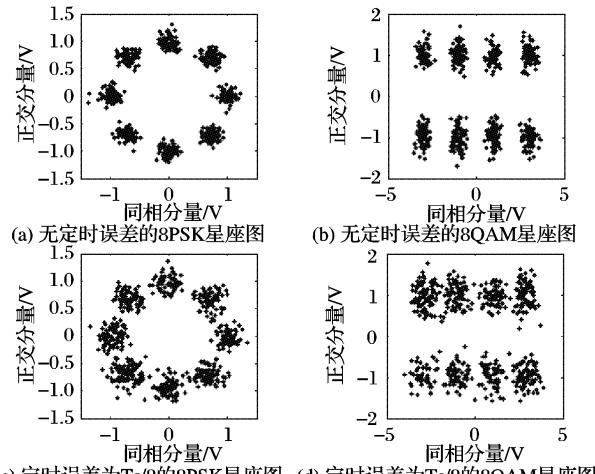


图 1 信号星座图

对于星座聚类问题,由于缺少先验知识,聚类中心数是未知的,因此算法必须确定聚类中心数。文献[6]提出了一种用于信号识别的聚类算法,根据聚类有效性自适应地确定聚类中心数,即“在聚类过程中,当  $c$  未达到最优解时,  $S$  随着  $c$  的增加而增大,趋向于紧密,当  $c$  到达最优值而继续增大时,  $S$  将会由最大值而减小。因此取  $S$  随  $c$  的增加而成为最大值时的  $c$  值作为聚类数”。但实验发现,并不是所有情况都满足该结论(具体分析见第 3 节)。针对上述问题,本文在原算法的基础上对聚类中心数的判定进行了改进,如图 2 所示。

与原算法相比,算法增加了一个判定条件: $S(t - 1)/S(t) > T$ ,用来判断当前最大值是有效性函数  $S$  随聚类中心数  $c$  的抖动,还是有效性函数  $S$  已经达到最大,避免得到错误的聚类中心数。其中  $T$  是判决门限,可以通过大量仿真实验得到。

后续处理则可以根据聚类中心数  $c$  判断信号的调制阶数。相同调制阶数也可能属于不同的信号类型,如图 1 中,两种信号都将得到  $c = 8$  的聚类中心数,还需要进一步判断。因此再根据幅度聚类结果判断信号所属类型,若幅度聚类结果为  $c = 1$ ,则判为 PSK 信号,否则判为 QAM 信号。

综上,聚类算法在信号识别中的应用步骤可以简要描述如下:

- 1) 对数据作归一化处理;
- 2) 用图 2 所示算法确定聚类中心数;

3) 对信号幅度聚类,确定信号属于PSK信号还是QAM信号。

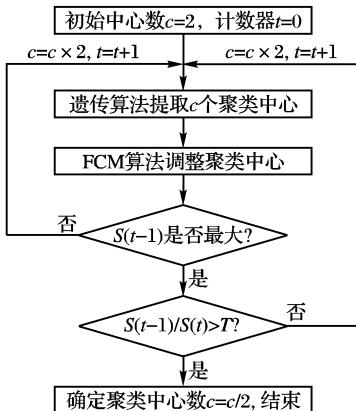


图2 遗传模糊聚类算法在星座重构中的应用框图

### 3 仿真结果与分析

以两种常用信号8QAM、8PSK为待识别信号,考查算法的有效性。我们假定已进行了载波同步,剩余载波相位在小范围抖动,由于固定相差只是导致星座图发生相应的旋转,而星座点相对位置不发生改变,为便于说明可假定固定相差为零。实验中PSK和QAM信号符号速率为1200 Bps,载波频率为1800 Hz。在没有特别申明情况下,各信号的信噪比均为12 dB。FCM算法中,加权系数 $m=2$ ,最大迭代次数 $R=50$ 。

图3显示了8QAM信号分别采用本文算法和文献[6]算法时 $S$ 随 $c$ 变化曲线。在 $c=8$ 时, $S$ 取得最大值。但当 $c$ 未达到最优解时, $S$ 并不是随着 $c$ 的增加而增大,因此若采用原算法将得出聚类中心数 $c=2$ 的错误结论,而改进算法则会通过增加的判定条件将 $c=2$ 判为函数的抖动,有效地避免了这一问题。另外,本文采用遗传算法初始化FCM算法,其有效性要略好于使用减法聚类算法初始化的原算法。

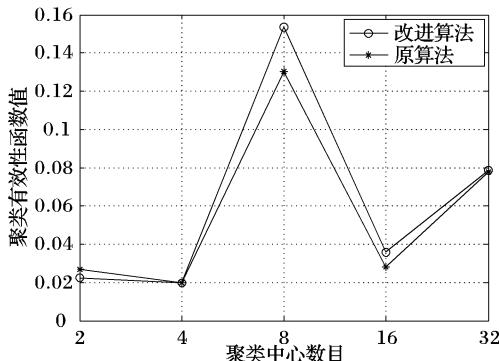


图3 聚类有效性函数与聚类中心数关系曲线

表1给出了8QAM信号应用本文算法得到的聚类中心,

结果与理论中心很接近,证明了本算法对星座恢复是有效的。表2是信号类型识别率的对比结果。实验中,每次采用600个观测符号,进行100次测试。从结果中可以看出,本文算法对8PSK信号的效果与原算法相比相差不大,原因主要是8PSK信号几乎可以保证当 $c$ 未达到最优解时, $S$ 随着 $c$ 的增加而增大,而8QAM信号的 $S$ 随着 $c$ 容易发生抖动,因此改进算法的识别率有较大提高。

表1 本文算法求解的8QAM信号的聚类中心

$c$	理论值	聚类结果	$c$	理论值	聚类结果
$v_1$	(-3, 1)	(-3.024, 1.041)	$v_5$	(-3, -1)	(-2.977, -1.015)
$v_2$	(-1, 1)	(-0.990, 0.989)	$v_6$	(-1, -1)	(-0.990, -1.053)
$v_3$	(1, 1)	(0.955, 1.003)	$v_7$	(1, -1)	(1.007, -1.035)
$v_4$	(3, 1)	(2.964, 1.035)	$v_8$	(3, -1)	(3.024, -0.965)

表2 本文算法与原算法的信号正确识别率 %

定时误差	8PSK		8QAM	
	原算法	改进算法	原算法	改进算法
0	100	100	100	100
$T_s/16$	100	100	82	98
$T_s/8$	98	100	68	82
$3T_s/16$	92	91	32	81
$T_s/4$	33	36	34	33

### 4 结语

本文将遗传算法与FCM算法结合起来,有效地弥补了二者各自的缺点,并将算法应用到星座聚类。算法能自适应地确定聚类中心,进而实现信号调制类型的识别。对存在一定码元定时误差的情况,算法仍能达到较好的识别效果。文中算法适用于幅相调制信号,需要载波和位同步,目前该领域已有较多的研究成果,为算法的应用提供了保证。

#### 参考文献:

- [1] 张智星,孙春在,(日)水谷英二. 神经——模糊和软计算[M]. 西安:西安交通大学出版社, 2000.
- [2] 吴月娴. 调制自动识别技术研究[D]. 郑州:信息工程大学, 2007.
- [3] 高新波. 模糊聚类分析及其应用[M]. 西安:西安电子科技大学出版社, 2004.
- [4] 周明,孙树栋. 遗传算法原理及应用[M]. 北京:国防工业出版社, 2001.
- [5] XIE X L, BENI G. A validity measure for fuzzy clustering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991, 13(8): 841–847.
- [6] 叶健,吴月娴,葛临东. 基于高效自适应聚类算法的调制识别研究[J]. 计算机工程与设计, 2007, 28(3): 506–508.

(上接第1196页)

#### 参考文献:

- [1] 黄震宇,沈祖和. 解一类非线性极大极小问题的熵函数方法[J]. 科学通报, 1996, 41(17): 1550–155.
- [2] 李兴斯. 一类不可微优化问题的有效解法[J]. 中国科学: A辑, 1994, 24(4): 371–377.
- [3] 欧宜贵,邓谋杰,洪世煌. 一类极大极小优化问题的信赖域算法[J]. 工程数学学报, 2004, 21(z2): 41–57.
- [4] 薛毅. 求解 Minimax 优化问题的 SQP 方法[J]. 系统科学与数学, 2002, 22(7): 355–364.
- [5] KENNEDY J, EBERHART R C. Particle swarm optimization[C]// Proceedings of IEEE International Conference on Neural Networks.

Piscataway, NJ: IEEE Press, 1995: 1942–1948.

- [6] SHI Y, EBERHART R. A modified particle swarm optimizer[C]// Proceedings of the IEEE International Conference on Evolutionary Computation. Piscataway, NJ: IEEE Press, 1998: 69–73.
- [7] PARSOPOULOS K E, VRAHATIS M N. On the computation of all global minimizes through particle swarm optimization [J]. IEEE Transactions on Evolutionary Computation, 2004, 8(3): 211–223.
- [8] 张建科,刘三阳,张晓清. 改进的粒子群算法[J]. 计算机工程与设计, 2007, 28(17): 4215–4219.
- [9] 张晓清,张建科,方敏. 多峰搜索的动态粒子群算法[J]. 计算机应用, 2005, 25(11): 2668–2670.