

基于二阶模糊聚类算法的雷达目标距离像识别

彭翔,周代英

(电子科技大学 电子工程学院,成都 611731)

(pengxiang_2000@163.com)

摘要: 针对模糊C-均值(FCM)算法敏感于聚类中心初始值的缺点,提出一种基于二阶模糊聚类方法。该方法利用传递闭包(TC)算法无初始化的优点,先对样本集按一定分类水平进行划分,选取若干类,求得这些类的样本均值作为FCM算法的初始聚类中心。一方面能够获得理想的聚类中心初始值,同时还能通过分类水平值来优化聚类中心数和聚类中心,避免局部最优,克服一致性聚类。利用该算法对三类飞机目标的实测一维距离像数据进行了识别实验,实验结果表明,基于二阶模糊聚类方法的识别率比FCM有了明显的改善。

关键词: 模糊C-均值;二阶模糊聚类;传递闭包;分类水平;距离像

中图分类号: TP391.41 **文献标志码:** A

Range profile target recognition based on second-order fuzzy clustering

PENG Xiang, ZHOU Dai-ying

(College of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu Sichuan 611731, China)

Abstract: Concerning that fuzzy C-means (FCM) algorithm was sensitive to the initial value of cluster centers, a fuzzy clustering method based on Second-Order Fuzzy Clustering (SOFC) was proposed, which took advantage of Transitive Closure (TC) algorithm's non-initialization, the samples were classified according to certain classification level firstly, and then the number of class was selected. The sample means of these classes was used to initialize the FCM algorithm's cluster centers. On one hand, it can obtain good initial value of cluster centers; on the other hand, the level of value through the categories can optimize the number of cluster centers and cluster centers, avoiding local optimum and overcoming the consistency of clustering. The algorithm was used to recognize three types of aircraft targets based on one-dimensional image data. The experimental results show that the recognition of SOFC algorithm gets obvious improvement than FCM algorithm.

Key words: Fuzzy C-Means (FCM); Second-Order Fuzzy Clustering (SOFC); transitive closure; classification level; range profile

0 引言

雷达目标一维距离像是雷达立体视角内目标体上强散射点回波在雷达视线方向上的投影,反映了目标散射点在纵向距离上的分布情况,含有丰富的目标几何形状、结构等固有特征信息,这使利用一维距离像来进行目标识别成为可能。而且,一维距离像较容易获得,对雷达平台有很好的适应性,所以利用其进行目标识别已经成为现代战争环境中的主要识别方法之一^[1]。但是一维距离像敏感于姿态角的变化,以至相同的目标在不同姿态角下有完全不同的一维距离像,而不同的目标一维距离像可能相近,这种模糊性给识别带来了困难。目前基于一维距离像的识别方法主要有以Li和Yang为代表的距离像匹配方法,以Rothwell等为代表的特征变换方法,以Potter和Chiang为代表的目标散射中心匹配方法及以Jacobs为代表的统计识别方法^[2]。

模糊集概念能够对概念较模糊的事物进行更好地描述,将其引入到聚类分析中,就产生了模糊C-均值(Fuzzy C-Means, FCM)、传递闭包(Transitive Closure, TC)、最大树(Maximal Tree)等算法。其中FCM算法是一种能自动对数据样本进行分类的方法,通过优化目标函数得到每个样本点对类中心的隶属度,从而决定样本点的归属。基于上述简洁、高效的性质,FCM算法被广泛地研究和应用^[3-6];但是,FCM算法也存在着诸多缺陷,如对初始值敏感、属于局部最优搜索、

不能保证获得全局最优解等^[7]。

TC算法是一种基于模糊等价关系的聚类分析方法^[8],虽然没有FCM算法简洁、高效,但它对初始值没要求,且可以通过控制其分类水平来改变聚类个数。

为此,本文提出一种二阶模糊聚类(Second-Order Fuzzy Clustering, SOFC)算法,将TC算法得到的聚类结果作为FCM算法的聚类中心初始值,从而克服因其盲目设置聚类初始值而导致的一致性聚类,根据FCM算法聚类效果反过来再指导TC算法划分样本,循环调整分类水平得到近似全局最优的聚类结果。

1 相关算法

1.1 FCM算法^[7]

无标记样本集 $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, $\mathbf{x} \in \mathbf{R}^{m \times n}$ 。其中, $\mathbf{x}_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}^T$, $i = 1, 2, \dots, n$ 。将样本集划分到 c 个类别中,由于引入模糊概念,一个样本对类别的隶属度取值范围不再只限 $\{0, 1\}$,而是区间 $[0, 1]$, n 个样本分别属于 c 个类别的隶属度可用矩阵表示为:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_{11} & \cdots & \mu_{1n} \\ \vdots & & \vdots \\ \mu_{c1} & \cdots & \mu_{cn} \end{bmatrix} \quad (1)$$

其中 μ_{ij} 为第 j 个样本对第 i 类目标的隶属度。

在硬聚类C-均值目标函数的基础上引入隶属度函数,就得到了FCM目标函数 J_f 。

收稿日期:2010-07-05;修回日期:2010-08-18。

作者简介:彭翔(1984-),男,海南万宁人,硕士研究生,主要研究方向:雷达目标识别;周代英(1966-),男,湖南湘潭人,教授,博士,主要研究方向:雷达目标识别。

$$J_f = \sum_{j=1}^c \sum_{i=1}^n [\mu_j(x_i)]^b \|x_i - m_j\|^2 \quad (2)$$

其中: $b \in \mathbf{R}$ 为一加权系数; m_j 为第 j 个聚类中心。其约束条件为:

$$0 \leq \mu_{ij} \leq 1; 1 \leq i \leq n, 1 \leq j \leq c \quad (3)$$

$$\sum_{j=1}^c \mu_{ij} = 1; 1 \leq i \leq n \quad (4)$$

$$0 \leq \sum_{i=1}^n \mu_{ij} \leq n; 1 \leq j \leq c \quad (5)$$

聚类的准则是通过求得最终隶属度矩阵 μ 和聚类中心组 m 使得目标函数 J_f 最小,根据拉格朗日乘数法求得 $\mu_j(x_i)$ 和 m_j 为:

$$m_j = \left(\sum_{i=1}^n [\mu_j(x_i)]^b x_i \right) / \left(\sum_{i=1}^n [\mu_j(x_i)]^b \right); 1 \leq j \leq c \quad (6)$$

$$\mu_j(x_i) = \frac{(1/\|x_i - m_j\|^2)^{1/(b-1)}}{\sum_{k=1}^c (1/\|x_i - m_k\|^2)^{1/(b-1)}}; \quad 1 \leq i \leq n, 1 \leq j \leq c \quad (7)$$

1.2 传递闭包算法^[8]

定义相似系数计算公式为:

$$r_{ij} = \frac{1}{m} \sum_{k=1}^m \exp\left(-\frac{3(x_{ik} - x_{jk})^2}{4S_k^2}\right); \quad i, j = 1, 2, \dots, n \quad (8)$$

$$\text{其中: } S_k = \left[\frac{1}{n} \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2 \right]^{1/2}; k = 1, 2, \dots, m \quad (9)$$

$$\bar{x}_k = \frac{1}{n} \sum_{i=1}^n x_{ik}; k = 1, 2, \dots, m \quad (10)$$

相似矩阵为:

$$R = \begin{bmatrix} r_{11} & \cdots & r_{1n} \\ \vdots & & \vdots \\ r_{n1} & \cdots & r_{nn} \end{bmatrix} \quad (11)$$

计算传递闭包 $t(R)$ 即 x 上的等价矩阵,重复进行关系合成运算^[8]:

$$R(i+1) = R(i) \circ R(i); i = 1, 2, \dots, \infty \quad (12)$$

当 $R(i+1) = R(i)$ 时停止。

设定分类水平 $\alpha (0 \leq \alpha \leq 1)$, 可得到 α 下的强截集 R_α 。在 R_α 中非对角线上取值为 1 的元素下标代表的样本归为一类,其余对角线上元素下标代表样本则自成一类,从而得到样本划分的结果。

1.3 SOFC 算法

SOFC 算法通过引入反馈方法,得到最优 α 。主要有两部分工作:1)怎样利用 TC 算法结果初始化 FCM 算法的聚类中心;2)怎样评价 SOFC 算法,并最终获得最优 α 。

1.3.1 聚类中心的初始化

将 TC 算法得到的各类按其所含样本个数的大小进行降序排序,取前 k 个类,并求得所选类样本均值 m 。

$$m = \{m_1, m_2, \dots, m_k\}; 1 \leq k \leq C \quad (13)$$

$$m_i = \frac{1}{l_i} \sum_{j=1}^{l_i} x_{ij} \quad (14)$$

其中: x_{ij} 为第 i 类的第 j 个样本; l_i 为第 i 类所含样本数; C 为总的类的个数。将求得样本均值 m 作为 FCM 算法初始化聚类中心。

1.3.2 最优分类水平 α 的确定

综合考虑时间代价和搜索精度,设定步长 λ , 给出 FCM 分类评价指标定义和其阈值 ε_0 , 将分类水平 α 从 $[0, 1]$ 遍历,再根据评价指标选取最优 α ,分类评价指标的定义需根据具体问题确定,获得最优 α 的流程如图 1 所示。

由图 1 可以看到,在确定了步长 λ 和分类评价指标阈值 ε_0 后,在某个 α 值下,利用 TC 法将样本进行初始划分,将划分所得类按其所含样本数大小进行降序排序。取前 i 个类,并以这 i 个类的各样本均值作为 FCM 算法的初始聚类中心向量组,后经 FCM 算法可得到最优聚类中心向量组,即可利用 1.4 节中介绍的识别算法进行目标识别。如果将目标识别率定义为分类评价指标,则用在某个 α 值下算得的识别率 ε 同目标识别率 ε_0 进行比较。如果小于目标识别率,则改变 α 值重复上述步骤;否则,存储该 α 值,即 α_{opt} 。

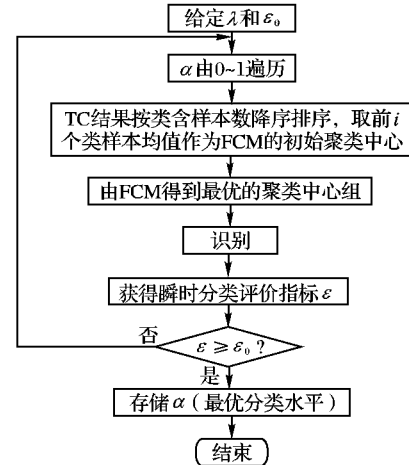


图 1 获取最优 α_{opt} 流程

SOFC 算法的目标是训练出参数 α_{opt} (即最优 α), 然后用其指导 TC 算法分类。 ε_0 一般取主观的期望阈值, 如果其过高可能捕获不到好的 α 。

1.4 识别算法

标记样本集为 $x = \{x_{11}, \dots, x_{n1}, \dots, x_{1c}, \dots, x_{nc}\}$, 即有 c 类目标, 每类目标有 n 个样本。中心向量组为 $m = \{m_{11}, \dots, m_{p1}, \dots, m_{1c}, \dots, m_{pc}\}$, 即有 c 类目标, 每类目标取 p 个聚类中心。定义距离度量:

$$d_{ij} = \text{norm}(x_{ij} - m_j); 1 \leq j \leq c, 1 \leq i \leq s, 1 \leq l \leq p$$

其中: c 为目标个数, s 为识别样本数, p 为每个目标聚类中心个数。则得到 c 个目标距离矩阵, 其中第 j 个目标距离矩阵为:

$$D_j = \begin{bmatrix} d_{11j} & \cdots & d_{s1j} \\ \vdots & & \vdots \\ d_{1pj} & \cdots & d_{spj} \end{bmatrix}; 1 \leq j \leq c \quad (15)$$

则判别式为:

$$r = \arg\min_{1 \leq j \leq c} \left(\sum_{i=1}^p d_{ij} \right); 1 \leq i \leq s \quad (16)$$

本文将目标的识别率作为分类评价指标。

2 仿真实验

为了验证本算法的有效性, 本文利用安-26、奖状、雅克-42 共 3 种飞机的各 4 段连续轨迹实测数据进行识别实验。

每段数据有 260 个 256 维的距离像, 取奇数像共 130 个作为训练样本, 偶数的像作为测试样本, 则在一组实验数据中有训练样本和测试样本各 390 个。

数据预处理如下。

1) 归一化。用一维距离像除以像本身的二范数进行归一化。

2) 距离对准。利用 Fourier 变换的平移不变性, 将一维距离像做 Fourier 变换即可对齐。

2.1 用 FCM 算法对三类飞机进行识别

当 $b = 2, c = 3$ 时, 对每类目标各任意初始化 3 个聚类中

心。用FCM算法分别对3个样本集聚类,在得到最优的聚类中心后,利用1.4节给出的识别算法对待识别样本进行识别。实验结果如表1所示。

目标	数据组			
	1	2	3	4
安-26	88	47	88	0
奖状	98	94	95	100
雅克-42	98	97	97	94

由表1可以看到,FCM算法对第2组和第4组数据奖状和雅克-42识别率很高,但对安-26的识别率却很低,甚至不能识别。

当 $b=2, k=3, \lambda=0.5$ 时,用SOFC算法专门针对低识别率目标进行优化,其他识别效果较好的组则通过参考FCM算法的结果,获得相应分类水平 α ,可得到与FCM算法同样的识别率,实验结果如表2所示。

目标	数据组			
	1	2	3	4
安-26	88	66	88	81
奖状	98	87	95	75
雅克-42	98	94	97	91

从表1和表2可以看到,第2组和第4组数据安-26目标识别率有了很大提高,尤其是第4组数据中由0%提高到81%;但是奖状和雅克-42在使安-26识别率提高的 α 条件下,识别率有所下降。

2.2 用SOFC算法对3类飞机进行识别

为了说明步长对识别率的影响,以第2组数据实验,仍然取前3个样本数最多的TC分类,并取其样本平均作为FCM的聚类中心,其他参数不变。为使实验更贴近实际,此处加入高斯白噪声,信噪比取20 dB,3类目标平均识别率随分类水平变化曲线如图2和图3所示。

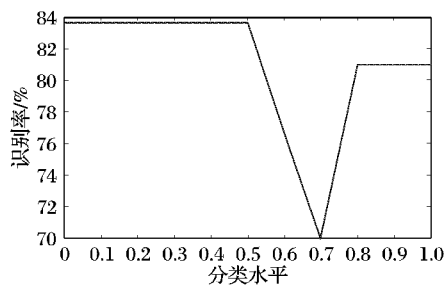


图2 步长为0.1时的平均识别率曲线

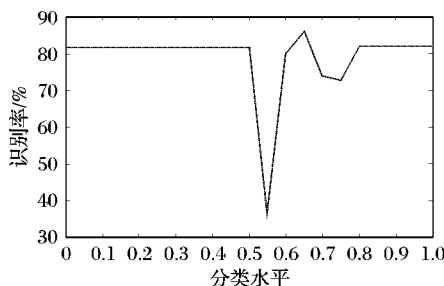


图3 步长为0.05时的平均识别率曲线

由图2和图3可以看到,3类目标平均识别率在步长为0.1和0.05时均有大幅下跌点,且步长越长,可见的下跌幅度越大。但是可以通过选取适当的分类水平来避开这些点,说明该算法有一定抗噪能力。

在信噪比为20 dB情况下取第4组数据进行实验,取分

类水平为0.4、0.7和0.8,其他参数不变,3类目标平均识别率随聚类中心数变化曲线如图4所示。

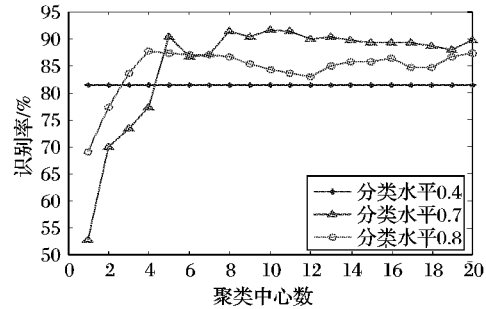


图4 3种分类水平的识别率曲线

由图4可见,当分类水平为0.4时,3类目标平均识别率不随聚类中心个数的增加而改变,因为几乎所有的样本都在TC划分时被分在了一个类模式中,这个类内模式已经包含了该类目标样本大多数信息;当分类水平为0.7时,平均识别率缓慢上升,当中心数为8时触顶缓慢回落并维持在某个水平上,这时样本比较均匀地分布在了少数几个类内模式中,即该类目标样本信息分散到了几个类内模式中;当分类水平为0.8时,平均识别率在中心数为4时就达到峰值,经短暂小幅波动后趋于平稳,这时样本均匀地散落在了许多类内模式中,每个类内模式可能只包含有1至2个样本。综上可见,对聚类中心个数的选择并非越多越好,应该适当选择聚类中心个数。

3 结语

SOFC算法通过恰当地选取初始聚类中心克服了FCM算法敏感于初始值的缺点,而且对传递闭包算法产生的聚类中心个数具有指导意义;另外,它具有很强的针对性,可以大幅提高感兴趣类的识别率,所以SOFC算法能更好地适用于一维距离像分类识别。但是,通过实验2可以看到,对步长和聚类中心数选取对识别率有很大影响。

参考文献:

- [1] 单凯品, 肖怀铁, 朱俊. 基于模糊核C-均值聚类分析的HRRP识别[J]. 电光与控制, 2010, 17(5): 42-45.
- [2] 周代英. 雷达目标一维距离像识别研究[D]. 成都: 电子科技大学, 2001.
- [3] HADJAHMADI A H, HOMAYOUNPOUR M M, AHADI S M. Robust weighted fuzzy C-means clustering [C]// IEEE International Conference on Fuzzy Systems. Washington, DC: IEEE Computer Society, 2008: 305-311.
- [4] CHINTALAPUDI K K, KAM M. A noise-resistant fuzzy C means algorithm for clustering [C]// Fuzzy Systems Proceedings, 1998 IEEE World Congress on Computational Intelligence. Washington, DC: IEEE Computer Society, 1998, 2: 1458-1463.
- [5] 于剑. 论模糊C均值算法的模糊指标[J]. 计算机学报, 2003, 26(8): 968-973.
- [6] 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, 2004, 22(3): 114-116.
- [7] 边肇祺, 张学工. 模式识别[M]. 2版. 北京: 清华大学出版社, 2000: 280-282.
- [8] 彭祖赠, 孙毓玉. 模糊(Fuzzy)数学及其应用[M]. 武汉: 武汉大学出版社, 2007: 147-152.
- [9] COLIN N, MORUZZIS M. Radar target recognition by fuzzy logic [C]// IEEE National Radar Conference. Washington, DC: IEEE Computer Society, 1997: 257-262.
- [10] 周德全. 基于一维距离像的雷达目标识别研究[D]. 南京: 南京理工大学, 1998.
- [11] 李珊. 高分辨一维距离像雷达目标识别方法研究[D]. 长沙: 国防科学技术大学, 2005.