

基于 GFCM 聚类算法的飞机故障诊断方法

鲁卿¹, 冯金富¹, 张佳强¹, 胡俊华¹, 李志强²

(1. 空军工程大学 工程学院, 陕西 西安 710038; 2. 95380 部队, 广东 湛江 524329)

摘要:为解决 FCM 算法对初始值敏感而易于陷入局部极小点的问题, 针对 FCM 算法应用于系统原位测试时小数据量特点, 提出了一种增量方式全局最优模糊 c 均值算法, 进行了收敛速度优化并给出了算法步骤, 机载武器系统信息通道原位故障诊断实验验证了此算法在小数据量情况下可以较好地解决 FCM 算法收敛局部最优的问题。

关键词:聚类; 模糊 c 均值; 全局优化; 故障诊断

中图分类号: TP306; TP18 **文献标识码:** A **文章编号:** 1009-3516(2007)06-0016-03

无监督聚类算法能很好的适应原位测试设备快速挖掘被测系统信号所隐藏的系统状态信息, 判断系统运行正常与否, 准确定位故障部位的要求^[1-4]。在众多的聚类算法中, 基于目标函数的模糊 c 均值(Fuzzy c-Means, FCM)聚类算法以其高效性在工程实际中得到了广泛应用。但 FCM 聚类算法对初始值较为敏感, 在样本群之间数目相差较大的情况下, 算法易陷入局部极小点或鞍点^[5-6]。

为克服聚类算法对初始化敏感的缺点, 许多学者针对其应用领域提出了解决方案。谢维信、高新波等人提出了基于遗传算法的 HCM 算法和 FCM 算法^[6], Klein 提出了采用模拟退火进行聚类分析^[7], Asultan 和 Selim 提出对模糊 c 均值算法中的模糊隶属度矩阵 U 进行模拟退火处理的模糊聚类算法^[8]。本文针对 FCM 算法应用于系统原位测试时小数据量特点, 在分析 FCM 算法迭代局部搜索问题的基础上, 依据文献[9]提出的增量方式假设, 提出了一种增量方式全局最优模糊 c 均值算法并进行了收敛速度优化, 最后将此方法在机载武器系统信息通道原位故障诊断实验中进行了有效性验证。

1 FCM 算法局部极值问题

FCM 聚类算法的出发点是基于对目标函数的优化, 其实质就寻找一组中心矢量, 使各样本到其的加权距离平方和达到最小^[6]。若样本集、聚类类别数和权重系数为已知, 就能通过迭代算法确定最佳模糊分类矩阵和聚类中心, 整个计算过程就是反复修改聚类中心和分类矩阵的过程。由于模糊聚类目标函数是非凸的^[5], 而 FCM 算法又是迭代爬山寻求最优解, 因此很容易陷入局部极值点或鞍点而得不到最优解甚至满意解, 其聚类结果好坏很大程度上取决于初始聚类中心点的选择。图 1 给出了初始值选择不当时由 FCM 算法产生的局部极值问题, 其中, “*”“o”标记点表示两类数据, “x”为聚类中心。

由图 1 可以看出, 陷入鞍点的聚类中心相对于全局最优聚类中心有靠近样本集质心的趋势。一般情况下, 诊断对象都是工作在正常状态, 不正常状态出现机会很少, 其相应的类所包含的对象也就会很

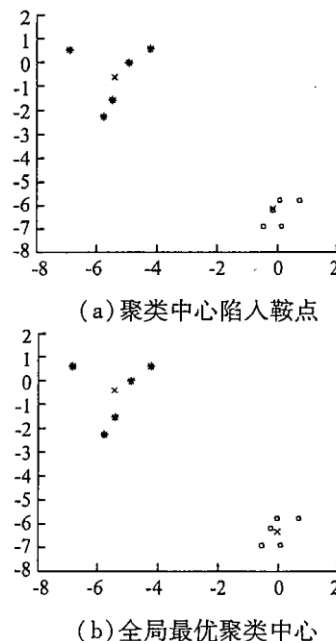


图 1 FCM 算法局部极值问题

收稿日期: 2007-05-29

基金项目: 军队科研基金资助项目

作者简介: 鲁卿(1980-), 男, 湖北监利人, 博士生, 主要从事机载武器系统故障诊断研究;

冯金富(1964-), 男, 江苏泰兴人, 教授, 主要从事机载武器系统总体、仿真与设计研究。

少^[2],在采用FCM算法对系统进行状态监控时,如果算法陷入局部极值点将导致正常态的数据被错误的划分到故障态一类,从而致使误报情况发生。

2 全局优化模糊c均值聚类算法

以上分析了FCM聚类算法是一种基于聚类准则的局部最优搜索方法,其聚类结果好坏很大程度上取决于初始聚类中心的选择。为解决这一问题,首先假定FCM聚类算法是一种局部搜索算法且其聚类结果随初始聚类中心位置不同而异。

在文献[9]中,对于数据集 X 的 k 均值聚类问题有如下假设:如果已知数据集 X 的 $(k-1)$ 类划分问题的 $(k-1)$ 个聚类中心的全局最优解,那么针对数据集 X 的 k 类划分问题,将 $(k-1)$ 类划分问题的 $(k-1)$ 个最优中心位置作为 k 类划分问题的 $(k-1)$ 个聚类中心初始位置,适当的选择第 k 个聚类中心初始位置,通过 k 均值聚类算法进行局部搜索可得到 k 均值聚类问题的全局最优解。

依据上述增量式假设,可提出全局最优模糊c均值(Global Fuzzy c-Means, GFCM)聚类算法。采取在每一步增加一个新的聚类中心增量方式进行优化,在数据量较小情况下,将数据集 X 的质心作为1类模糊划分的最优聚类中心,对于2类模糊划分问题,第1个初始聚类中心位置为1类划分问题的最优聚类中心位置,第2个初始聚类中心位置分别选取为样本集 X 的 n 个样本点,分别采用该 n 组初始聚类中心点进行“模糊2-均值”聚类,取目标函数 J_m 最小对应的一组聚类中心为“模糊2-均值”的最优解,依此类推,直到取得 c 模糊划分的全局最优解。其核心思想概括如下:找到 $(c-1)$ 类模糊划分的最优聚类中心 $(v_1^*(c-1), v_2^*(c-1), \dots, v_{c-1}^*(c-1))$,取 $(v_1^*(c-1), v_2^*(c-1), \dots, v_{c-1}^*(c-1)x_n)$, $(n=1, 2, \dots, n)$ 为模糊 c 类划分的 n 组初始聚类中心,分别采用FCM算法搜索局部最优解,目标函数 J_m 最小对应即为模糊 c 划分的全局最优解。显然,该方法不依赖任何初始条件,提高了聚类精度。

然而,考虑到算法收敛速度,上述方法出现一个问题:每次增加一个新的聚类中心则需运行 n 次FCM算法,由此将导致减慢算法的收敛速度。文献[10]给出了一种“合适”初始聚类中心点的选取方法:选取初始

目标函数 $J_m = \sum_{i=1}^n (\sum_{j=1}^n \|x_i - v_j\|^{2(1-m)})^{1-m}$ 最小对应的一组初始聚类中心。依据该方法,为提高GFCM算法收敛速度,在每一步聚类的 n 个初始状态不进行 n 次FCM算法得到 n 个最终目标函数值 J_m 而进行比较选择,取代该步采用计算 n 个初始状态的目标函数值 J_m ,寻找初始目标函数最小对应的初始状态作为初始聚类中心。由此每次增加一个聚类中心则只需运行1次FCM算法,可加快算法收敛。

以下给出改进的GFCM算法的步骤:①确定聚类类别数 C ,计算数据集质心 v_0 ,以 v_0 为1类模糊划分的全局最优聚类中心;②以 $(v_1^*(c-1), v_2^*(c-1), \dots, v_{c-1}^*(c-1)x_n)$ 为 c 类模糊划分的 n 组初始聚类中心,运行 J_m 计算目标函数,将 $\min J_m$ 对应的一组作为FCM算法初始聚类中心,运行FCM算法,计算全局最优聚类中心;③If $c=C$,结束, or $c=c+1$,转向②。

3 实验与结论

某型飞机机载武器控制系统信息通道实现对空地制导武器搜索瞄准阶段的目标指示,是发射使用导弹的关键环节,该通道发生故障将导致魔球控制失调或目标坐标信息错误等问题。经过大量的数据收集和分析,得到信息通道的若干故障模式,其中选取使用某型空地电视制导武器时的目标指示通道为实验研究对象如图2所示。

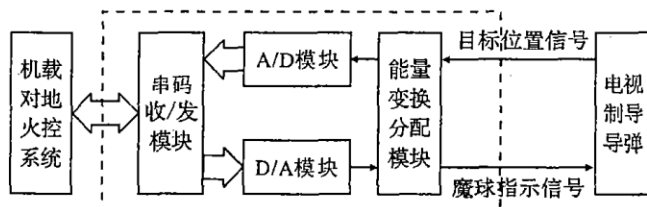


图2 某型电视制导武器信息通道示意图

该通道有4种典型的故障模式,具体如下: F_1 :串码收/发模块故障; F_2 :A/D转换模块故障; F_3 :D/A转换模块故障; F_4 :能量变换分配模块故障。本实验采集系统正常状态 F_0 数据40组和4种已知故障状态的数据各3组并进行归一化处理,其中每组数据中包含两路测试点数据:机载对地火控系统接收的目标位置信号和电视制导导弹接收的魔球指示信号。

分别采用FCM和GFCM两种算法对测试数据进行状态聚类,其中模糊加权幂指数 $m=2$ 、最大迭代次数 $T_{max}=20$ 时,两种算法聚类中心及对应状态如下:

$$V_{fcm} = \begin{bmatrix} 0.5143 & 0.7311 \\ 0.2619 & 0.1101 \\ 0.0578 & 0.0388 \\ 0.1158 & 0.1936 \\ 0.5639 & 0.7239 \end{bmatrix} \sim \begin{bmatrix} F_4 \\ F_2 \\ F_1 \\ F_3 \\ F_0 \end{bmatrix} \quad V_{gfc} = \begin{bmatrix} 0.5429 & 0.7154 \\ 0.1468 & 0.1305 \\ 0.0618 & 0.0358 \\ 0.1141 & 0.2243 \\ 0.2375 & 0.1206 \end{bmatrix} \sim \begin{bmatrix} F_0 \\ F_4 \\ F_1 \\ F_3 \\ F_2 \end{bmatrix}$$

样本点分布及分类结果如图3所示。

由图3可以看出,正常状态的测试数据在采用FCM算法的状态分类中被误分为两类,而 F_2 (A/D转换模块故障)和 F_4 (能量变换分配模块故障)被归为一类,诊断结果错误;采用GFCM的故障诊断结果则与实际测试结果完全一致,很好地克服了FCM算法迭代搜索收敛局部最优的问题,提高了故障诊断的精度。

为解决FCM算法对初始值敏感而易于陷入局部极小点的问题,针对FCM算法应用于系统原位测试时小数据量特点,在分析FCM算法迭代局部搜索问题的基础上,依据文献[7]提出的 k 均值聚类增量式假设,提出了一种全局最优模糊 c 均值算法并进行了收敛速度优化,将此方法在机载武器系统信息通道原位故障诊断实验中进行了有效性验证,结果显示此算法在小数据量情况下可以较好地解决FCM算法收敛局部最优的问题。

参考文献:

- [1] 欧阳晓黎,杨军. 基于模式识别的装备故障诊断方法[J]. 弹箭与制导学报, 2004, 24(2): 86-90.
- [2] 徐章遂,房立清,王希武,等. 故障信息诊断原理与应用[M]. 北京:国防工业出版社, 2000.
- [3] 王锋,孙秀霞. 基于专家系统与神经网络集成的航姿系统故障诊断[J]. 空军工程大学学报:自然科学版, 2005, 6(1): 17-21.
- [4] 王志华. 基于模式识别的柴油机故障诊断技术研究[D]. 武汉:武汉理工大学能源与动力工程学院, 2004.
- [5] Bezdek J C. Pattern Recognition with Fuzzy Objective Function Algorithms[M]. New York: Plenum Press, 1981.
- [6] 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.
- [7] Klein R W, Dubes R C. Experiments in Projection and Clustering by Simulated Annealing. Pattern Recognition, 1989, 22(2): 213-220.
- [8] AL-Sultan K S, Selim S Z. A Simulated Annealing Algorithm for the Clustering Problem. Pattern Recognition, 1991, 24(10): 1003-1008.
- [9] Aristidis L, Nikos V, Jakob J. Verbeek. The Global K-Means Clustering Algorithm. Pattern Recognition, 2003, 36: 451-461.
- [10] Nabil b, Pierre H, Nenad M. Fuzzy J-Means: a new heuristic for fuzzy clustering. Pattern Recognition, 2002, 35: 2193-2200.

(编辑:姚树峰)

A Fault Diagnosis Method of Airplane Based on Global Fuzzy C-Means Clustering Algorithm

LU Qing¹, FENG Jin-fu¹, ZHANG Jia-qiang¹, HU Jun-hua¹, LI Zhi-qiang²

(1. The Engineering Institute, Air Force Engineering University, Xi'an 710038, China; 2. P. L. A. 95380, Zhanjiang 524329, Guangdong, China)

Abstract: In order to solve the problem apt to land in local minimum results for its sensitivity to the initial conditions, in view of character of test-online, this paper proposes a global fuzzy c -means (GFCM) clustering algorithm based on incremental approach to clustering. The converging speed of GFCM is improved by simplifying the algorithm and then the approach of the algorithm is given. The experimental test for unsupervised clustering and fault pattern recognition of the information channels of airborne weapon system is given by using the new GFCM algorithm. The results show that the proposed algorithm is effective in dealing with the aforementioned problem under condition of small data capacity.

Key words: cluster; fuzzy c -means; global optimization; fault diagnosis

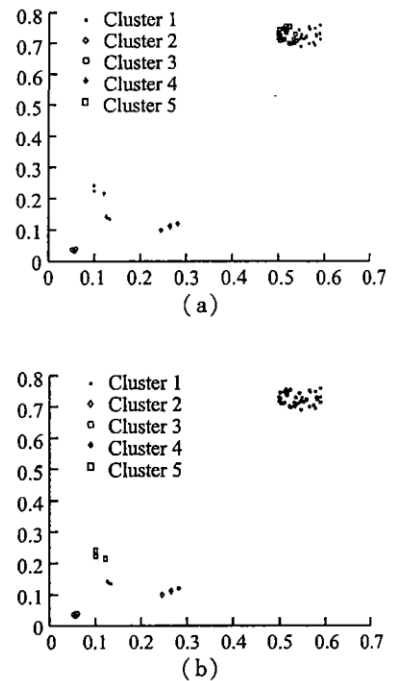


图3 两种算法诊断结果比较