

基于综合模糊聚类算法的液体火箭 发动机故障诊断

董周杰, 郭迎清
(西北工业大学 动力与能源学院, 西安 710129)

摘 要: 基于液体火箭发动机正常及故障状况数据的完备程度和数据质量的不断提高,提出一种基于数据驱动的综合模糊聚类算法用于故障诊断。采用模糊 c 均值(FCM)算法对已知正常样本数据进行聚类得到最优的聚类中心,将所得到的聚类中心作为先验样本数据用于传递闭包法最优分类结果的选择从而得到故障检测结果,该算法只需要少量的正常先验样本数据就能快速、准确的检测出故障;随后采用 FCM 算法进行故障分类,可以根据现有的故障数据库进行聚类得到对应的故障类型,并且可以给出故障幅值范围。模型仿真结果表明:该算法对故障的检测率可达 96.8%,故障隔离率达到 94%。某型液体火箭发动机实际试车数据结果表明:该故障诊断算法能够准确及时的检测并隔离出故障。

关 键 词: 液体火箭发动机; 故障诊断; 模糊聚类; 模糊 c 均值 (FCM); 模糊传递闭包法

中图分类号: V233.7 **文献标志码:** A

Fault diagnosis of liquid rocket engine based on comprehensive fuzzy clustering algorithm

DONG Zhoujie, GUO Yingqing
(School of Power and Energy,
Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: Based on the completeness of the normal and fault condition data of liquid rocket engine and the improvement of data quality, a data-driven comprehensive fuzzy clustering algorithm was proposed for fault diagnosis. The fuzzy c -means (FCM) algorithm was used to cluster the known normal sample data to obtain the optimal clustering center, and the obtained cluster center was used as a-priori sample data to select the optimal classification result of the closure method to obtain the fault detection result. Only a small amount of normal prior sample data were required to quickly and accurately detect the fault; then the FCM algorithm was used to classify the fault, and the corresponding fault type can be clustered according to the existing fault database, and the range of fault amplitudes can be given. The simulation results showed that the detection rate of the algorithm was up to 96.8% and the fault isolation rate was 94%. The actual test data of a liquid rocket engine show that the fault diagnosis algorithm can detect and isolate faults accurately and timely.

Key words: liquid rocket engine; fault diagnosis; fuzzy clustering; fuzzy c -means (FCM); fuzzy transfer closure

收稿日期:2019-11-09
作者简介:董周杰(1995—),男,硕士生,主要从事液体火箭发动机故障诊断研究。

引用格式:董周杰,郭迎清. 基于综合模糊聚类算法的液体火箭发动机故障诊断[J]. 航空动力学报,2020,35(6):1326-1334. DONG Zhoujie, GUO Yingqing. Fault diagnosis of liquid rocket engine based on comprehensive fuzzy clustering algorithm[J]. Journal of Aerospace Power, 2020, 35(6):1326-1334.

随着我国液体火箭发射次数的日益增多,以及新一代液体火箭动力系统的研制成熟,积累了丰富的发动机飞行数据和试车数据。此外,利用目前较为成熟的液体火箭发动机故障仿真系统,可以产生发动机正常状况与各类故障状况下的仿真数据。这些丰富的液体火箭发动机飞行数据、试车数据和仿真数据,为研究基于数据分析的发动机故障诊断方法提供了基础^[1]。

模糊聚类算法便是一种基于数据的故障诊断算法,模糊聚类分析是一种采用模糊数学对事物按一定的模糊界限进行描述和分类的“软划分”方法,一般根据样本本身的属性、样本间的亲疏程度和相似性来定义模糊关系,并在此基础上根据一定的隶属度来确定聚类关系,即用模糊数学的方法定量地确定样本间的模糊关系,从而实现客观且准确的聚类分析^[2-4]。

目前,模糊聚类分析方法广泛运用于图像处理^[5]、模式识别^[6]、神经网络训练^[7]、故障诊断^[1,8-9]等领域,均取得了满意的效果和客观的效益,具有重要的理论与实际应用价值。耿辉等^[1]将聚类分析理论应用于液体火箭发动机故障诊断,其方法的诊断结果对所使用的数据质量和完备程度有极大的依赖性;李艳军^[8]将模糊 c 均值聚类算法用于液体火箭发动机的起动过程故障检测,其中涉及检测阈值的选取,检测效果受人为影响因素比较大;Lee 等^[9]研究了基于卡尔曼滤波器和故障因子相结合的液体火箭发动机故障诊断算法,其对模型的依赖性比较强;Zhou 等^[10]提出了一种基于特征加权模糊聚类(weighting fuzzy compactness and separation, WFCS)的多传感器数据航空发动机故障诊断的新方法,该方法的性质是在线性故障数据上实现的,并不能适用于非线性故障数据的故障诊断。

本文采用模糊 c 均值聚类和模糊传递闭包法结合的综合模糊聚类诊断算法对液体火箭发动机进行故障诊断,引入先验知识使原本无监督学习的聚类算法变为有监督的故障诊断算法,实现液体火箭发动机的故障诊断,诊断的结果严格按照聚类的定义,避免人为设定检测阈值等参数的影响,定量计算样本数据之间的模糊关系,从而达到故障的检测和隔离。

1 模糊 c 均值聚类(FCM)

1.1 FCM 原理

模糊 c 均值聚类(fuzzy c -means, FCM)算法,

也称为模糊 ISODATA (iterative selforganizing data analysis)方法,是一种基于目标函数优化的数据聚类方法,通过优化目标函数得到各样本对所有聚类中心的隶属度,从而决定样本的类属以达到自动分类的目的。FCM 算法是从硬聚类目标函数的优化中导出的,具体是由 Bezdek^[11]和 Dunn^[12]在 Ruspini^[13]提出的模糊划分概念基础上推广得出的,它把聚类问题归结为一个非线性规划问题,然后采用交替优化策略求解无监督的分类问题。究其本质,FCM 算法是一种局部搜索的爬山法,通过聚类原型 $\mathbf{P}$ 和模糊划分矩阵 $\mathbf{U}$ 之间的交替迭代来寻找局部最优解。

模糊聚类目标函数描述如下:

$$J_m(\mathbf{U}, \mathbf{P}) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m (d_{ik})^2, \quad \mathbf{U} \in \mathbf{M}_{fc} \quad (1)$$

其中 n 为样本个数, c 是聚类数量, m 为平滑因子,最佳取值^[14]为 $[1.5, 2.5]$, u_{ik} 为第 k 个样本属于第 i 类的隶属度, d_{ik} 为第 k 个样本与第 i 个聚类中心的距离,定义为

$$d_{ik} = \|x_k - p_i\|_2 \quad (2)$$

$\mathbf{M}_{fc}$ 为模糊划分空间,具体表达式如下:

$$\mathbf{M}_{fc} = \{\mathbf{U} | u_{ik} \in [0, 1], \forall i, k; \sum_{i=1}^c u_{ik} = 1, \forall k\} \quad (3)$$

聚类的准则为取 $J_m(\mathbf{U}, \mathbf{P})$ 的极小值,约束条件为 $\sum_{i=1}^c u_{ik} = 1$, 且模糊划分矩阵 $\mathbf{U}$ 各列都是独立的,因此存在如下等值条件:

$$\min\{J_m(\mathbf{U}, \mathbf{P})\} = \sum_{k=1}^n \min\left\{\sum_{i=1}^c (u_{ik})^m (d_{ik})^2\right\} \quad (4)$$

可以用拉格朗日乘数法求解

$$F = \sum_{i=1}^c (u_{ik})^m (d_{ik})^2 + \lambda \left(\sum_{i=1}^c u_{ik} - 1 \right) \quad (5)$$

最优化的一阶必要条件为

$$\begin{cases} \frac{\partial F}{\partial \lambda} = \sum_{i=1}^c u_{ik} - 1 = 0 \\ \frac{\partial F}{\partial u_{ik}} = m(u_{ik})^{m-1} d_{ik}^2 + \lambda = 0 \end{cases} \quad (6)$$

由上述方程可以解得($d_{ik} > 0$ 时)

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{2}{m-1}}} \quad (7)$$

其中 $\exists d_{ik} = 0$ 时, $u_{ik} = 1, \forall j \neq i, u_{jk} = 0$ 。

用同样的方法可以得到 $\min \{J_m(U, P)\}$ 对应的第 i 个聚类中心的值

$$p_i = \frac{\sum_{k=1}^n (u_{ik})^m x_k}{\sum_{k=1}^n (u_{ik})^m} \tag{8}$$

若样本集 X 、聚类数目 c 和平滑因子 m 已知, 并知道初始模糊划分矩阵 U 或初始化聚类原型 P , 便可通过迭代算法确定相应的最佳模糊划分矩阵和聚类中心, 并得到相应的聚类结果。

1.2 FCM 步骤

FCM 算法可以初始化模糊划分矩阵 U 或初始化聚类原型 P 来作为迭代的初始值, 当初始化 P 时, 最终聚类结果会知道每一类的标签, 若初始化 U , 则无法得知相对应的标签。初始化 P 进行 FCM 算法步骤如下:

- 1) 初始化: 初始化聚类原型 $P^{(0)}$, 待聚类样本集 X , 设置迭代停止阈值 ϵ 为 10^{-5} , 给定平滑因子 m 为 1.5, 最大迭代次数 $b_{\max}$ 为 200, 迭代计数器 $b=0$;
- 2) 根据式(7)计算模糊划分矩阵 $U^{(b)}$;
- 3) 根据式(8)更新聚类中心矩阵 $P^{(b+1)}$;
- 4) 如果 $\|P^{(b)} - P^{(b+1)}\|_2 \leq \epsilon$, 或者达到迭代最大次数, 则算法迭代停止, 否则令 $b=b+1$, 转向步骤 2);
- 5) 待隔离样本隶属于所有类别的最大值小于 0.5 时, 则认为未知的故障类型; 否则, 其最大值对应的类别即为该样本隶属的类别。

采用 FCM 算法进行故障隔离的程序流程图如图 1 所示。

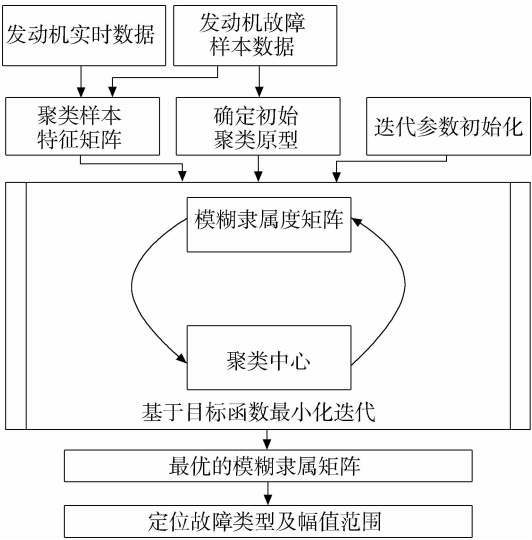


图 1 FCM 算法故障隔离流程图

Fig. 1 Fault isolation flowchart based on FCM

2 模糊传递闭包法

2.1 模糊传递闭包法原理

模糊传递闭包法就是利用样本和属性之间的模糊关系构建模糊相似矩阵 R , 然后由模糊相似矩阵求出传递闭包 $t(R)$, 再依据不同的分割值 λ 对 $t(R)$ 进行聚类, 让 λ 由 1 向 0 变小, 可以得到不同分类数时的分类结果, 可以看出原有的模糊传递闭包法是一种无监督的学习方法, 且不同的分隔值, 会得到不同的聚类结果, 难以获得一个最优的聚类结果。本文引入先验知识(已知正常样本)进行聚类, 选择已知正常样本被聚为一类的分类情况为最终的分类结果, 通过观察此时待测样本是否与已知正常样本聚类为一类来达到检测故障的目的。

2.2 模糊传递闭包法步骤

- 1) 构造特性指标矩阵

每一个样本 x_i 有 m 个特性指标, 即 x_i 可由如下 m 维特性指标向量表示

$$x_i = (x_{i1}, x_{i2}, \dots, x_{im})^T, \quad i = 1, 2, \dots, n \tag{9}$$

其中 x_{ij} 表示第 i 个样本的第 j 个特征指标, 则 n 个样本的所有特性指标构成特性指标矩阵, 记作

$$X = [x_1, x_2, \dots, x_n]^T \tag{10}$$

- 2) 数据标准化处理(构造模糊矩阵)

对原始数据进行无量纲处理, 将矩阵中的元素数值压缩在 $[-1, 1]$ 区间内, 进而构建模糊矩阵。

平移-标准差变换

$$x'_{ik} = \frac{x_{ik} - \bar{x}_k}{s_k} \tag{11}$$

其中 $\bar{x}_k$ 表示第 k 个样本属性的数据平均值, s_k 第 k 个样本属性的数据标准差。

平移-极差变换

$$\dot{x}_{ik} = \frac{x'_{ik} - \bar{x}_k}{\max_{1 \leq i \leq n} \{x'_{ik}\} - \min_{1 \leq i \leq n} \{x'_{ik}\}} \tag{12}$$

- 3) 构造模糊相似矩阵

采用绝对值倒数法进行模糊相似矩阵构造^[15]

$$r_{ij} = \begin{cases} 1 & i = j \\ \frac{c}{\sum_{k=1}^m |x_{ik} - x_{jk}|} & i \neq j \end{cases} \tag{13}$$

其中 c 为适当选取的正数, 使 $r_{ij} \in [0, 1]$ 。

- 4) 模糊等价矩阵(求传递闭包)

从 $R = (r_{ij})_{n \times n}$ 出发, 根据模糊关系合成运算反复自乘, 依次计算出 $R^2, R^4, \dots$, 当第一次出现

$R^k \circ R^k = R^k$ 时,得到传递闭 $t(R) = R^k$,其中 $k \leq [\log_2 n] + 1, n$ 为样本个数。

假设 Q, R 分别是 $U \times V, V \times W$ 上的两个模糊关系,其模糊关系合成运算公式如下:

$$(Q \circ R)(u, w) = \bigvee_{v \in V} [Q(u, v) \wedge R(v, w)] \tag{14}$$

5) 模糊截断矩阵

令分割值 λ 由 1 下降至 0,使用 λ 对传递闭包 $t(R)$ 进行聚类,获得在该分割值 λ 情况下的模糊截断矩阵 $t(R)_\lambda = [r'_{ij}(\lambda)]_{n \times n}$,然后对 $t(R)_\lambda$ 进行分析,得到的分类就是 λ 水平上的等价分类。 λ 从 1 变化到 0,可以得到具有动态性的、论域从细到粗的模糊分类。

聚类的原则如下:

$$r'_{ij}(\lambda) = \begin{cases} 1 & r'_{ij} \geq \lambda \\ 0 & r'_{ij} < \lambda \end{cases} \tag{15}$$

对于 $x_i, x_j \in X$,若 $r_{ij}(\lambda)' = 1$,则在 λ 水平上, x_i 和 x_j 可以归为一类。

分割值 λ 人为选取的不同,聚类结果便有所不同,为了使其能用于实时故障诊断,可以令 $X = [\underbrace{x_1, x_2, \dots, x_{n-1}}_{\text{normal}}, \underbrace{x_n}_{\text{test}}]^T$,即由 $n-1$ 个先验正常样本数据和待测样本构成特性指标矩阵,最终选取先验征程样本数据被聚为一类的结果为最终的分类结果,并以此来诊断待测样本是否为故障数据,基于传递闭包法的故障检测流程图如图 2 所示。

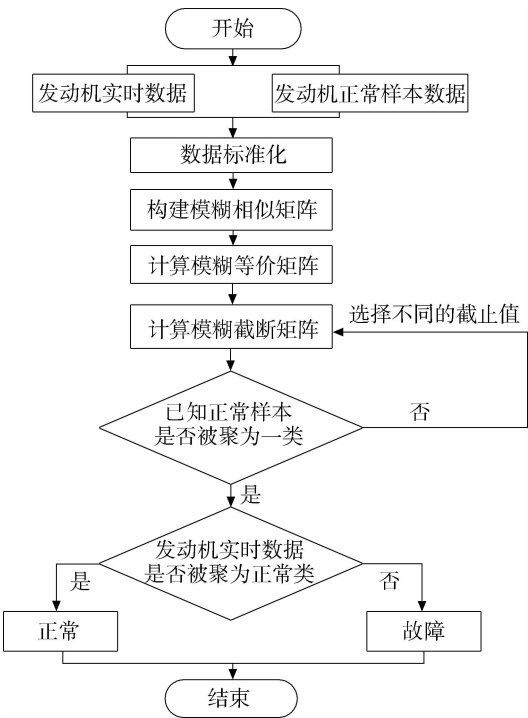


图 2 故障检测流程图

Fig. 2 Flowchart of fault detection procedure

由于正常先验样本的引入,先验样本的质量和数量会直接影响改进的模糊传递闭包法的检测结果和效率,一方面,若已知先验样本之间差距过大,则可能会导致故障样本被聚类为正常类;反之,则可能会使原本正常的样本被聚类为故障类;另一方面,随着先验样本数据的数量的增多,检测效率会有所下降。

2.3 参数寻优

由文中第 2.2 节可知,改进的模糊传递闭包法的检测结果与所选取的先验正常样本的质量和数量有密切的关系,为此采用基于初始化模糊划分矩阵 U 的 FCM 算法对正常样本库按照不同聚类个数获取对应的聚类中心,将其作为先验正常样本数据用于传递闭包法的故障检测,选取最优的故障检测结果对应的聚类中心作为最终故障检测的先验正常样本数据。

由于参数寻优时所用的检测故障数据集的类别不平衡性,因此采用 F1-Score(F_1)作为综合反映整体的故障检测指标^[15-16],其表达式如下所示:

$$F_1 = \frac{2 \times P \times R}{P + R} \tag{16}$$

其中 P 和 R 分别表示精准率和召回率,具体计算参照文献[15]中的性能参数计算公式。

基于 FCM 聚类算法对传递闭包法进行参数寻优的程序流程图如图 3 所示。

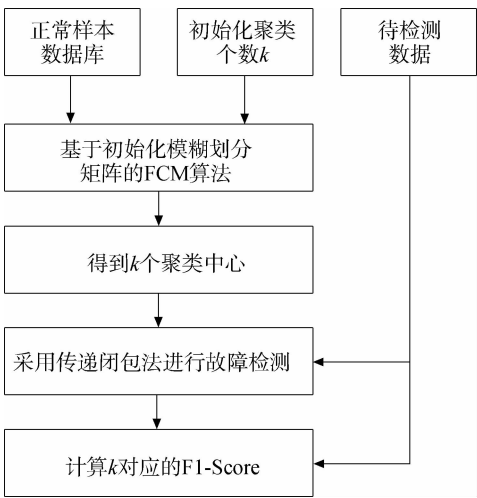


图 3 参数寻优流程图

Fig. 3 Flowchart of parameter optimization

3 综合模糊聚类

模糊传递闭包法和模糊 c 均值聚类算法两种聚类算法有各自的优缺点,两种算法的优缺点对比如表 1 所示。

表 1 两种聚类算法应用的优缺点对比

Table 1 Comparison of advantages and disadvantages of two clustering algorithms for application

算法	优点	缺点
模糊传递闭包法	没有迭代过程,计算速度快; 可以发现新的故障类型	分类精度不够高;需要靠专家知识决策出最优的 分类结果;无法给出对于各种故障的隶属度
模糊 c 均值聚类算法	可以给出待测样本对于各种已知故障的 隶属度;分类精度较高	迭代算法速度比较慢;分类结果的优劣性对 初始化设置依赖较高

为了充分利用两种模糊聚类算法的优点,提出综合模糊聚类诊断算法对液体火箭发动机进行故障诊断。本文将两种算法结合使用,首先利用模糊传递闭包法进行数据的异常检测,检测故障的发生,当发生故障时,再运用模糊 c 均值聚类算法对故障数据进行聚类,从而达到故障隔离的目的。

采用综合模糊聚类算法对液体火箭发动机进行故障诊断,一方面可以提高故障检测的效率,另一方面不仅能够定位故障类型,还可以给出大致的故障幅值范围,且随着试验数据质量和数量的不断完备,该算法的故障定位将越来越准确,故障幅值范围估计也会越来越精确。

4 仿真结果分析

4.1 仿真模型

本文根据已有的液体火箭发动机模型来进行故障诊断的研究,传感器测点参数主要如表 2 所示,主要包含甲烷涡轮泵转速、液氧涡轮泵转速、甲烷泵后压力、液氧泵后压力、副系统甲烷喷嘴前压力、副系统液氧喷嘴前压力、主系统甲烷喷嘴前压力、主系统液氧喷嘴前压力、燃气发生器压力、燃气发生器温度和燃烧室压力。

故障模拟采用故障因子注入法,主要模拟液体火箭发动机主的故障类型包括阀门开度故障、泵气蚀、涡轮效率损失和管路泄露。根据所搭建的液体火箭发动机模型定义各类故障时故障因子范围,其中阀门开度、气蚀、效率、流量等故障因子偏离标准值(所默认标准值均为 1)3%就认为发生故障,流阻故障因子偏离 5%认为发生故障,具体如表 3 所示。

表 2 液体火箭发动机传感器测点参数
Table 2 Liquid rocket engine sensor measuring point parameters

输出参数名称	符号
甲烷涡轮泵转速/(r/min)	n_i
液氧涡轮泵转速/(r/min)	n_o
甲烷泵后压力/MPa	p_{ef}
液氧泵后压力/MPa	p_{eo}
主系统甲烷喷嘴前压力/MPa	p_{go}
主系统液氧喷嘴前压力/MPa	p_{co}
燃气发生器压力/MPa	p_g
燃气发生器温度/K	T_g
燃烧室压力/MPa	P_c

表 3 液体火箭发动机故障因子
Table 3 Liquid rocket engine fault factor

故障因子	符号	正常范围	故障范围
氧副控阀阀门开度异常故障因子	F_{pov}	$0.97 \leq F_{pov} \leq 1$	$0 < F_{pov} < 0.97$
氧主控阀阀门开度异常故障因子	F_{mov}	$0.97 \leq F_{mov} \leq 1$	$0 < F_{mov} < 0.97$
甲烷泵汽蚀故障因子	$F_{fcavitation}$	$0.97 \leq F_{fcavitation} \leq 1$	$0 < F_{fcavitation} < 0.97$
甲烷涡轮效率故障因子	$F_{fefficiency}$	$0.97 \leq F_{fefficiency} \leq 1$	$0 < F_{fefficiency} < 0.97$
甲烷管路流阻故障因子	$F_{fresistance}$	$1 \leq F_{fresistance} \leq 1.05$	$1.05 < F_{fresistance} \leq 2$
甲烷管路流量故障因子	F_{fmass}	$0.97 \leq F_{fmass} \leq 1$	$0 < F_{fmass} < 0.97$
氧泵汽蚀故障因子	$F_{ocavitation}$	$0.97 \leq F_{ocavitation} \leq 1$	$0 < F_{ocavitation} < 0.97$
氧涡轮效率故障因子	$F_{oefficiency}$	$0.97 \leq F_{oefficiency} \leq 1$	$0 < F_{oefficiency} < 0.97$
氧管路流阻故障因子	$F_{oresistance}$	$1 \leq F_{oresistance} \leq 1.05$	$1.05 < F_{oresistance} \leq 2$
氧管路流量故障因子	F_{omass}	$0.97 \leq F_{omass} \leq 1$	$0 < F_{omass} < 0.97$

4.2 模糊传递闭包法参数寻优

4.2.1 参数寻优参数选择

正常样本数据集:分别令表 3 中 10 个故障因子在正常取值范围内偏离标准值(所默认标准值均为 1)0.1,每偏移 0.001 取模型输出为一个正常样本,其中每种故障产生 10 个正常样本数据,共 100 组正常样本;

待检测样本集:按照表 3 中故障因子的取值范围,每偏移 0.001,取模型输出为一个待检测样本,每种故障产生 1000 组,共 10 000 组待检测样本数据,用于参数寻优的样本数据个数统计分布如图 4 所示。

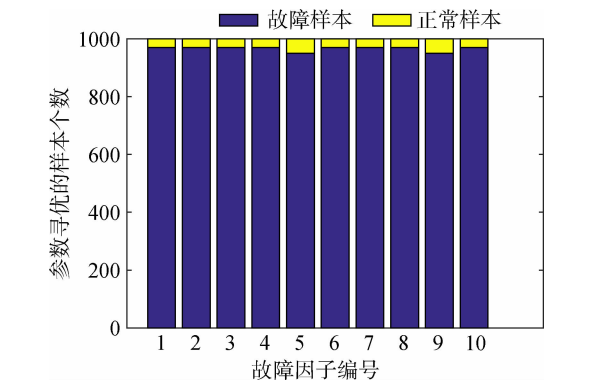


图 4 参数寻优的样本数据个数统计图
Fig.4 Statistical distribution of sample data for parameter optimization

通过参数寻优的样本数据个数统计分布图可以看出每种故障情况下对应的正常和故障样本数据个数差别很大,故用 F1-score 的值来评价最终故障检测性能的优劣。

聚类个数 k :从 2 到 20,每取一个值按照图 3 程序流程进行计算相应的 F1-Score 故障检测指标。

4.2.2 参数寻优结果分析

在给定聚类个数 k 的情况下,分别根据每种故障因子对应的 1000 组检测样本检测结果计算对应 F1-Score,求 10 个故障因子下 F1-Score 的平均值,记作 F1-Score-mean,其随聚类个数 k 的变化如图 5 所示。

由图 5 结果可知,将模糊传递闭包法用于故障检测的所计算的 F1-Score-mean 指标都很高,表明模糊传递闭包法能够有效、准确的检测出故障,当聚类个数 k 为 4 时对应的 F1-Score-mean 值最大,又考虑到算法的实时性,综合选取该种情

况下的最优聚类中心作为模糊传递闭包法故障诊断的先验已知样本数据。

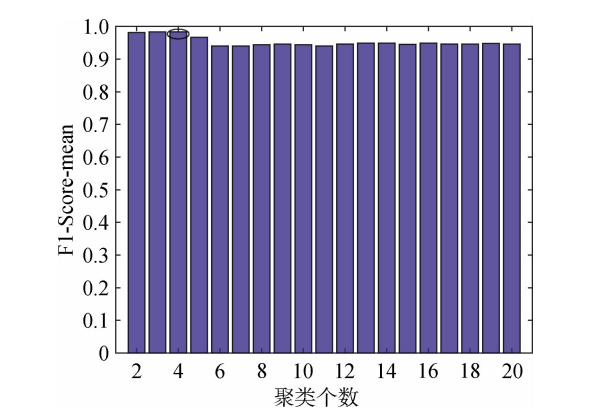


图 5 F1-Score-mean 随聚类个数变化图
Fig.5 Bar chart of F1-Score-mean with changing the number of clusters

4.3 故障检测结果分析

设定在 5 s 时,在液体火箭发动机模型的故障注入部分注入氧副控阀阀门开度异常故障(对应表 3 中故障因子 1),故障幅值为 0.75,仿真时间为 10 s,步长为 0.001 s。

其中模型输出参数 n_f 随时间的变化曲线如图 6 所示,在 0.743 s,输出迭代稳定,此时输出为正常数据,当 5 s 加入故障后,5.408 s 迭代稳定,此时为 F_{pov} 故障等级为 0.75 时的 n_f 故障数据。

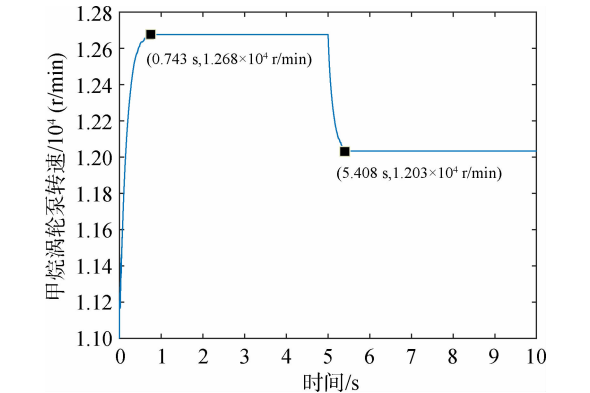


图 6 输出参数 n_f 变化曲线
Fig.6 Curve of output parameter n_f

分别选取 4 s 和 8 s 时刻获取对应的动态聚类结果图,分别如图 7 和图 8 所示。其中 λ 列代表选取的不同截止值,对应截止值下的分类数记为 C ,中间一行数据是对应的样本编号,其中 1~4 为正常样本编号,5 为当前时刻的待测样本编号。

通过图 7 和图 8 的聚类结果图,可以看出当

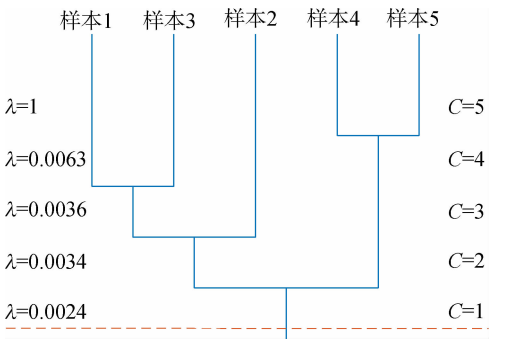


图 7 4 s 时对应的动态聚类结果图

Fig. 7 Graph of dynamic clustering result at four second

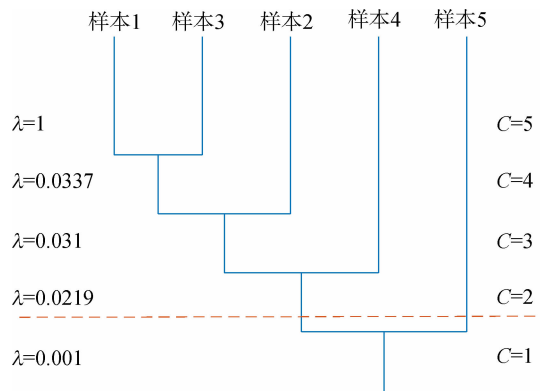


图 8 8 s 时对应的动态聚类结果图

Fig. 8 Graph of dynamic clustering result at eight seconds

待检测样本正常时,待检测样本会在先验正常样本被聚为一类之前与其中的先验正常样本聚为一类,而当待检测样本为故障样本时,先验正常样本被聚为一类时,待检测样本仍单独为一类,基于此能够实现液体火箭发动机实时故障检测。

4.4 故障隔离结果分析

4.4.1 故障隔离数据选择

考虑到要尽可能的将已知的故障样本数据聚类为应属于自己的一类,可以通过适当地降低平滑因子 m 的值来降低模糊隶属关系,因此最终选取 m 的值为 1.25,其余参数设定与文中第 1.2 节中的参数初始化一致。

发动机故障样本数据:参照表 3 中前 4 种故障,分别选取故障因子范围在 0.7~0.9 前闭后开区间,每隔 0.01 取对应的模型输出为一个先验故障样本数据,其中每种故障对应 20 组故障样本,共 80 组故障样本数据。

聚类原型的选取:对每种故障的 20 组样本数据,每 10 个样本数据取平均值作为一个聚类原型,其中每种故障对应 2 个故障原型,共 8 个聚类原型。

测试数据的选取:针对每种故障按式(17)随机生成 50 个故障因子的值,将每个故障因子对应下模型输出参数作为测试数据,其中每种故障对应随机生成的 50 个测试数据,共 200 个测试数据^[17]。

$$F = F_{\min} + (F_{\max} - F_{\min}) \times \text{rand}(0,1) \tag{17}$$

式中 F 表示故障因子的值, $F_{\min}$, $F_{\max}$ 分别表示选取故障因子范围的最小值和最大值,函数 rand 产生一个均匀分布的随机数。

4.4.2 结果分析

选取故障 1(对应表 3 中的氧副控阀阀门开度异常故障)随机生成的故障因子值为 0.8558 的对应的测试数据进行故障隔离生成的隶属度矩阵图如图 9 所示,其中横坐标表示故障样本编号,其中 1~80 依次对应文中第 4.4.1 节生成的已知发动机故障样本数据,81 号样本属于待隔离样本,纵坐标表示对应样本隶属于该类的隶属度,其值越大表明该样本隶属于该类的可能性越大。

从图 9 中可以看出,当将故障因子 1 的幅值设置为 0.8558 时,FCM 算法能够很好地将其聚类为故障 1 的类型,而且还聚类在其对应故障幅值 0.8~0.9 区间内,隔离结果表明 FCM 算法能够很好地对液体火箭发动机故障类型进行分类并且能够给出故障的大概幅值范围。

为了整体表现其聚类效果的好坏,可以绘制如图 10 的分类轮廓图,轮廓值是衡量样本点与它所属聚类相较于其他聚类的相似程度,取值范围为 $[-1,1]$,值越大表明该点更匹配其所属聚类,而不与相邻的聚类匹配。如果大多数点具有高轮廓值,则聚类解决方案是合适的 如果许多点具有低或负轮廓值,则聚类解决方案可能具有太多或太少的聚类。

从分类轮廓图中,可以看出 2、5、7 类中大多数 silhouette 值都接近 0.8,表明其分类效果好;其他类中大多数的值都大于 0.5,可以认为分类较为准确;其中没有出现负值的情况,表明没有被认为的错误分类,可以整体看出该聚类结果比较理想。

对所有测试样本数据进行隔离检验结果统计

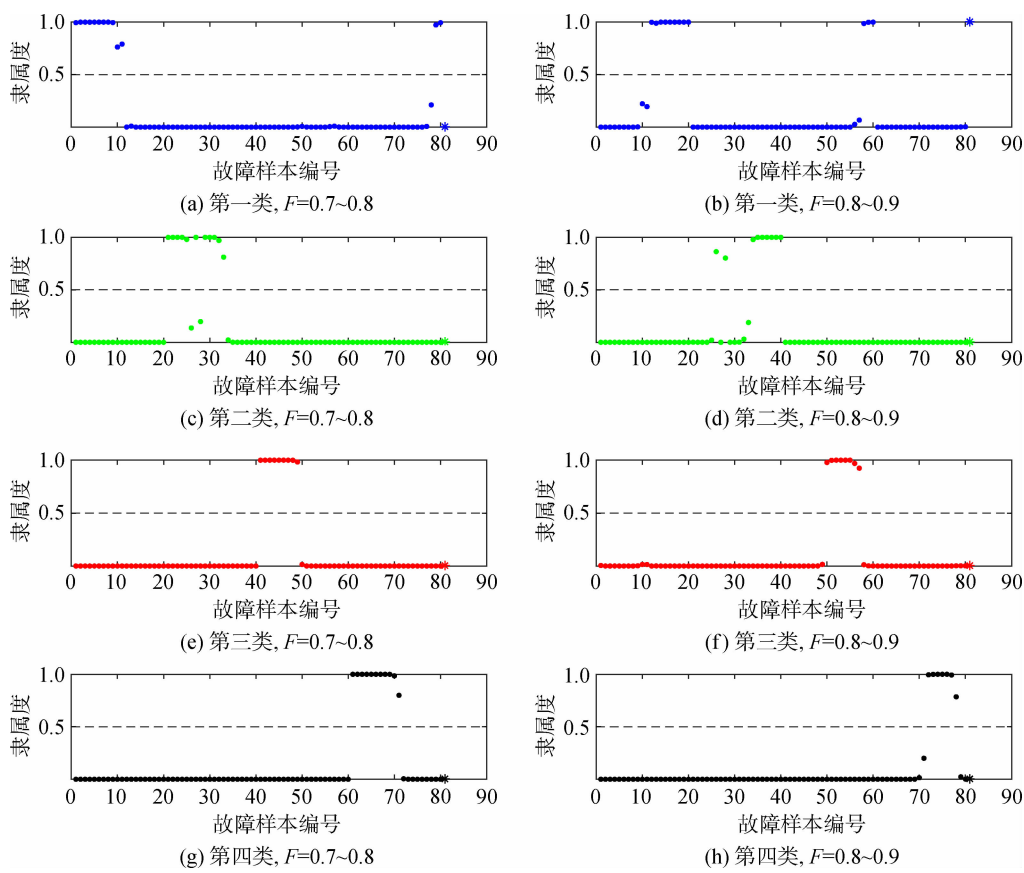


图 9 最终故障隔离的隶属度矩阵图

Fig. 9 Membership matrix value for the final fault isolation

如表 4 所示,可以看出该 FCM 算法对于故障的隔离准确率以及故障幅值划分准确率都比较高,其中故障 2 中故障隔离准确率为 100%而故障幅值划分准确率比较低,是由于随机生成的故障因子 2 的值分布在 0.8 附近的值比较多;而故障 4 的故障隔离准确率只有 84%,同时其故障幅值划分准确率却达到 100%,这表明凡是被正确隔离为故障 4 类型的样本均被准确地划分到相应的幅值区。

表 4 故障隔离检验结果统计

Table 4 Statistical table of fault isolation test

故障类型	故障隔离 准确率/%	故障幅值范围 划分准确率/%
1	100	98.00
2	100	80.00
3	92	89.13
4	84	100.00
统计	94	91.78

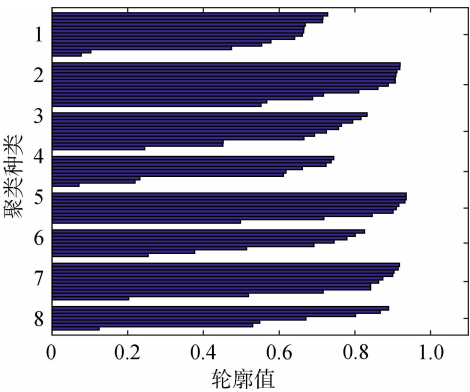


图 10 分类轮廓图

Fig. 10 Graph of classification silhouette

5 结 论

本文通过开展有关模糊聚类算法在液体火箭发动机故障诊断方面的研究工作得到以下结论：
1) 所提出的基于数据驱动的综合模糊聚类故障诊断算法只需要少量的先验正常样本数据就能快速、准确的检测出故障的发生。

2) 综合模糊聚类算法不仅能够识别故障类型,而且可以得到大致的故障幅值范围,并且可以随时添加已知故障库,具有良好的推广性。

3) 基于综合模糊聚类算法的液体火箭发动

机故障诊断过程中不涉及人为的阈值选取等对检测结果影响因素比较大的过程,可以很好地进一步推广应用于工程实践中去。

参考文献:

[1] 耿辉,张翔,张素明,等. 一种基于聚类分析的液体火箭发动机稳态过程故障诊断方法[J]. 火箭推进,2014,40(5):86-91.
GENG Hui,ZHANG Xiang,ZHANG Suming,et al. Cluster analysis based diagnosis method for fault in steady process of liquid propellant rocket engine[J]. Journal of Rocket Propulsion,2014,40(5):86-91. (in Chinese)

[2] 陈水利,李敬功,王向公. 模糊集理论及其应用[M]. 北京:科学出版社,2005.

[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] 李鸿吉. 模糊数学基础及实用算法[M]. 北京:科学出版社,2005.

[5] 雷涛,张肖,加小红,等. 基于模糊聚类的图像分割研究进展[J]. 电子学报,2019,47(8):1776-1791.
LEI Tao,ZHANG Xiao,JIA Xiaohong et al. Research progress on image segmentation based on fuzzy clustering[J]. Acta Electronica Sinica,2019,47(8):1776-1791. (in Chinese)

[6] 王冰,王微,胡雄,等. 基于 GG 模糊聚类的退化状态识别方法[J]. 仪器仪表学报,2018,39(3):21-28.
WANG Bing,WANG Wei,HU Xiong,at el. Degradation recognition method based on GG fuzzy clustering[J]. Journal of Instrumentation,2018,39(3):21-28. (in Chinese)

[7] 何勇,项利国. 基于模糊聚类的 BP 神经网络模型研究及应用[J]. 系统工程理论与实践,2004,24(2):79-82.
HE Yong,XIANG Liguu. Research and application of BP neural network model based on fuzzy clustering[J]. System Engineering Theory and Practice,2004,24(2):79-82. (in Chinese)

[8] 李艳军. 新一代大推力液体火箭发动机故障检测与诊断关键技术研究[D]. 长沙:国防科学技术大学,2014.
LI Yanjun. . Research on key technologies for fault detection and diagnosis of new generation high thrust liquid rocket engine[D]. Changsha: National University of De-

fense Technology,2014. (in Chinese)

[9] LEE K,CHA J,KO S,et al. Fault detection and diagnosis algorithms for an open-cycle liquid propellant rocket engine using the Kalman filter and fault factor methods[J]. Acta Astronautica,2018,150(1):15-27.

[10] ZHOU Y,ZUO H F,HE J. Aero-engine fault diagnosis using a feature weighting fuzzy clustering algorithm[J]. International Journal of Control and Automation, 2017, 10(5):161-168.

[11] BEZDEK J C. Pattern recognition with fuzzy objective function algorithms[J]. Advanced Application in Pattern Recognition,1983,25(3):442.

[12] DUNN J C. A graph theoretic analysis of pattern classification via Tamura's fuzzy relation[J]. IEEE Transactions on Systems Man and Cybernetics,1974,4(3):310-313.

[13] RUSPINI E H. A new approach to clustering[J]. Information and Control,1969,15(1):22-32.

[14] 高新波,谢维信. 模糊 c 均值聚类算法中加权指数 m 的研究[J]. 电子学报,2000,28(4):80-84.
GAO Xinbo,XIE Weixin. A study of weighting exponent m in a fuzzy c -means algorithm[J]. Acta Electronica Sinica,2000,28(4):80-84. (in Chinese)

[15] 王新洲,舒海翅. 模糊相似矩阵的构造[J]. 吉首大学学报,2003,24(3):37-41.
WANG Xinzhou,SHU Haichi. Construction of fuzzy similarity matrix[J]. Journal of Jishou University,2003,24(3):37-41. (in Chinese)

[16] 王攀,陈雪娇. 基于堆栈式自动编码器的加密流量识别方法[J]. 计算机工程,2018,44(11):140-147.
WANG Pan,CHEN Xuejiao. ASE-based encrypted traffic identification method [J]. Computer Engineering, 2018, 44(11):140-147. (in Chinese)

[17] 王修岩,李萃芳,高铭阳,等. 基于 SVM 和 SNN 的航空发动机气路故障诊断[J]. 航空动力学报,2014,29(10):2493-2498.
WANG Xiuyan,LI Cuifang,Gao Mingyang,et al. Fault diagnosis of aero-engine gas path based on SVM and SNN [J]. Journal of Aerospace Power, 2014, 29(10):2493-2498. (in Chinese)

(编辑:张 雪)