ISSN 1001-9081 CODEN JYIIDU 2020-03-10 http://www.joca.cn

文章编号:1001-9081(2020)03-0917-08

DOI:10.11772/j.issn.1001-9081.2019071309

# 基于局部密度的加权一类支持向量机算法 及其在涡轴发动机故障检测中的应用

黄 功,赵永平\*,谢云龙

(南京航空航天大学 能源与动力学院,南京 210016) (\*通信作者电子邮箱 y. p. zhao@163. com)

摘 要:针对基于数据的涡轴发动机故障检测算法的分类性能较差、鲁棒性不强的问题,提出一种改进的加权一 类支持向量机(WOCSVM)算法——基于局部密度的WOCSVM(LD-WOCSVM)算法。首先,对于每个训练样本,选取 以该样本为中心,以全体训练样本中心到距离最远样本之间马氏距离的百分之二为半径的球体内所包含的k个近邻 样本;其次,以该样本到选定的k个训练样本的中心的距离大小来评估该样本为故障样本的可能性,并以此为依据,使 用经过归一化的距离来计算对应样本的权重。针对目前算法不能很好地反映样本分布特点的问题,提出了一种基于 快速聚类的权重计算方法并将其命名为FCLD-WOCSVM。该算法通过求取每个训练样本的局部密度和该样本到高 局部密度的距离两个参数,来确定该样本的分布位置,并利用求得的两个参数来计算该样本的权重。两种算法都是 通过对可能的故障样本分配较小的权重来增强算法的分类性能。为了验证算法的有效性,分别在4个UCI数据集和 T700涡轴发动机上进行仿真实验。实验结果表明,与自适应WOCSVM(A-WOCSVM)算法相比,LD-WOCSVM 算法在 AUC值上提高了0.5%,FCLD-WOCSVM算法在G-mean上提高了12.1%,两种算法可以作为涡轴发动机故障检测候选 算法。

关键词:涡轴发动机;故障检测;一类支持向量机;局部密度;马氏距离

中图分类号:TP181 文献标志码:A

## 文献标志码:A Fault detection for turboshaft engine based on local density weighted one-class SVM algorithm

HUANG Gong, ZHAO Yongping\*, XIE Yunlong

(College of Energy and Power Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing Jiangsu 210016, China)

Abstract: An improved Weighted One Class Support Vector Machine (WOCSVM) algorithm-Local Density WOCSVM (LD-WOCSVM) was proposed to solve the problems of poor classification performance and weak robustness of the data-based turboshaft engine fault detection algorithm. Firstly, for each training sample, k nearest neighbor samples contained in the body of the ball were selected, and the ball was centered on this sample with a radius of 2% of the Mahalanobis distance from the center of all training samples to the farthest samples. Secondly, the distance from this sample to the center of selected ktraining samples was used to evaluate the probability that this sample is a fault sample, and based on this, the normalized distance was used to calculate the weight of the corresponding sample. An algorithm of weight calculation based on rapid clustering namely FCLD-WOCSVM was proposed to deal with the problem that the present algorithms were not able to reflect the characteristics of sample distribution very well. In this algorithm, by obtaining two parameters of the local density of each training sample and the distance from the sample to the high local density, the distribution position of this sample was determined, and the weight of the sample was calculated by using the two obtained parameters. The classification performance of both algorithms was improved by assigning small weights to the possible fault samples. In order to verify the effectiveness of the two algorithms, simulation experiments were carried out on 4 UCI datasets and T700 turboshaft engines respectively. Experimental results show that, compared with Adaptive WOCSVM (A-WOCSVM) algorithm, LD-WOCSVM algorithm improves the AUC (Area Under the Curve) value by 0.5%, and FCLD-WOCSVM algorithm improves the G-mean (Geometric mean) by 12.1%. These two algorithms can be used as candidate algorithms for turboshaft engine fault detection.

Key words: turboshaft engine; fault detection; One-Class Support Vector Machine (OCSVM); local density; Mahalanobis distance



收稿日期:2019-07-30;修回日期:2019-09-08;录用日期:2019-09-09。 基金项目:中央高校基本科研业务费专项(NS2017013)。

作者简介:黄功(1994—),男,湖北孝感人,硕士研究生,主要研究方向:航空发动机建模与控制、航空发动机故障检测、机器学习 算法; 赵永平(1982—),男,河南南阳人,研究员,博士,主要研究方向:航空发动机建模与控制、人工智能; 谢云龙(1996—),男,安徽安庆人, 硕士研究生,主要研究方向:航空发动机建模与控制、机器学习算法。

第40卷

## 0 引言

涡轴发动机是一种将燃料的化学能转化为输出轴功率的 机械装置,主要为直升飞机提供动力。由于其压气机、涡轮等 部件长时间工作在高温、高压和高转速的恶劣环境里,发生故 障在所难免。据统计,在航空发动机发生的各类故障中,气路 故障占比约90%以上,因此对气路故障进行有效的检测和研 究就显得尤为重要<sup>[1]</sup>。

为了能及时有效地对故障进行检测和隔离来保护人民生 命财产安全和降低损失,各国学者和技术人员提出了很多有 用的方法。目前,针对气路故障检测的方法主要有基于模型、 基于数据和基于知识规则等三种方法[2]。其中,基于模型和 基于数据的两种检测方法应用最为广泛。在基于模型的检测 方法中,比较常用的有基于卡尔曼滤波器[3-4]和基于观测器的 方法[5]。基于模型的检测方法在机载实时应用方面为我们带 来了很多好处,但同时也存在着很多不足。该方法要求对发 动机知识和建模过程有较全面的掌握,而且随着发动机设计 水平和制造工艺的日趋复杂,建模不确定性和建模难度也越 来越大,使建立精确的数学模型充满了挑战。与此同时,随着 支持向量机(Support Vector Machine, SVM)、人工神经网络 (Artificial Neural Network, ANN)等人工智能算法的快速发 展,基于数据的故障检测方法在检测效率和检测精度等方面 有了长足的进步。其中,反向传播(Back Propagation, BP)算 法是一种最基础的ANN算法。理论上, BP算法具有任意逼近 能力,但同时也存在譬如收敛速度慢,易陷入局部最小等缺 点。殷锴等<sup>[6]</sup>利用BP算法来对航空发动机进行故障检测,实 验表明,BP算法可作为故障检测的一种备选方案。为了进一 步提高BP算法的检测精度,Qu等<sup>[7]</sup>提出了一种利用遗传算法 (Genetic Algorithm, GA)来优化 BP神经网络的初始权重和阈 值的新算法,并将其用于航空发动机气路故障诊断中,实验证 明,该算法相较于BP算法具有更快的收敛速度和更高的故障 诊断精度。极限学习机(Extreme Learning Machine, ELM)是 一种扩展的单隐含层前馈神经网络(Single-hidden Layer Feedforward Neural Network, SLFNN)。相较于 BP 神经网络, ELM不需要更新隐含层,因而大大缩短了训练时间。王小雷 等<sup>[8]</sup>利用ELM对发动机进行故障检测,取得了较好的效果。 为了消除无效的隐含层节点,廖洪一<sup>[9]</sup>提出一种利用GA来优 化ELM的输入权值矩阵和隐含层偏置值的新算法,相较于标 准的ELM,这一新算法有效地精简了隐含层节点数,提高了模 型的泛化能力。当ELM用于解决故障检测等分类问题时,存 在一种硬间隔缺陷,Zhao等<sup>[10]</sup>提出了一种软极限学习机(Soft ELM, SELM)算法来解决这一问题。SELM算法通过灵活地 为每个样本设定一个目标间隔,再求解一系列正则化的ELM, 有效地提高了算法的分类精度、增强了模型的泛化能力。 SVM 是 Cortes 等[11]于 1995 年提出的一种二分类算法,该算法 通过最小化结构风险来寻找最优分类面,被广泛应用于图像 分割[12-14]、人脸识别[15-17]和文本识别[18-19]等领域。在2012年, Heng等<sup>[20]</sup>使用SVM对航空发动机数据进行故障检测分析,取 得了比较好的效果。最小二乘支持向量机(Least Squares Support Vector Machine, LSSVM)是基于SVM改进而来的,它 通过求解线性方程组而不是求解二次规划问题在很大程度上 降低了算法的训练时间。在2013年,庄夏等[21]提出一种利用 改进的粒子群优化算法来优化LSSVM的惩罚参数和核函数

参数的新算法,将该算法用于故障检测中得到了比较好的 结果。

随着发动机设计水平的提高以及制造工艺的日趋完善, 发动机故障数据获取的难度越来越大,代价也越来越大,因 此,类别不平衡学习(Class Imbalance Learning, CIL)问题在基 于数据的故障检测方法中的重要性日益凸显。在2018年,Xi 等<sup>[22]</sup>提出了一种能解决CIL问题的改进LSSVM算法,在涡扇 发动机数据集中取得了比较好的效果。一类支持向量机 (One Class Support Vector Machine, OCSVM)是一种只需要正 常数据就可建立故障检测分类器的单分类算法,常用于故障 检测<sup>[23-24]</sup>、入侵检测<sup>[25-26]</sup>等。加权一类支持向量机(Weighted OCSVM, WOCSVM)是通过对正常样本分配较大的权值,对 可能的故障样本分配较小的权值,来最大限度地降低故障样 本对分类器的影响,从而达到增强算法鲁棒性的目的,因此, 对权重计算规则的制定就显得尤为重要。文献[27]提出了一 种自适应加权一类支持向量机(Adaptive WOCSVM, A-WOCSVM)算法,该算法能有效地对样本中的离域点进行检 测,但也存在着不能很好地描述样本边界的问题。为此,本文 提出一种基于局部密度的加权一类支持向量机(Local Density WOCSVM, LD-WOCSVM)算法。在该算法中,笔者首 先考虑到马氏距离能排除特征之间相关性的干扰这一特性, 在求解样本上近邻的过程中,使用马氏距离而不是欧氏距离 来度量,并且选用当前样本k近邻的中心而不是整个数据集 的中心来计算样本权重。此外,笔者通过改进一种快速聚类 算法来识别可能的故障样本,并对故障样本分配较小的权重 而提出了一种新算法,为方便叙述,将这一新算法命名为 FCLD-WOCSVM。在4个基准数据集上的仿真结果表明,与 A-WOCSVM 算法相比, LD-WOCSVM 算法和 FCLD-WOCSVM 算法在故障检测中能取得更好的结果,证明了本文提出算法 的有效性。

1 基于局部密度的加权一类支持向量机

## 1.1 加权一类支持向量机

OCSVM是Schölkopf等<sup>[28]</sup>于1999年提出的一种单分类算法,只需要正常数据即可建立分类器,能很好地解决CIL问题。为了能进一步提高OCSVM算法的鲁棒性,学者们提出了各种改进措施,WOCSVM算法即是其中的一种。本文是在文献[27]所提算法的基础上做出改进的。

给定训练样本集 $D = \{x_i\}_{i=1}^{N}, x_i \in \mathbb{R}^n, x$ 是一个测试样本, 且 $x \in \mathbb{R}^n$ 。存在从 $\mathbb{R}^n$ 到希尔伯特空间 $\chi$ 的非线性映射 $\varphi$ 使得  $\varphi(x_i) \in \chi$ ,在希尔伯特空间存在超平面 $w \cdot \varphi(x_i) - \rho = 0$ 将原 点和映射样本完全分开。其中,w为超平面法向量, $\rho$ 为偏移 量<sup>[29]</sup>,映射样本代表着正常样本,原点代表着故障样本。引入 松弛因子 $\xi_i \ge 0$ 可使算法具有某种程度的鲁棒性。用 $\omega_i$ 表示 样本 $x_i$ 对应的权重,且 $\omega_i \in [0, 1], 则$  WOCSVM 的优化目标可 写为

$$\min_{\boldsymbol{w},\rho} \frac{1}{2} \boldsymbol{w}^{\mathrm{T}} \boldsymbol{w} - \rho + \frac{1}{\nu N} \sum_{i=1}^{N} \omega_i \boldsymbol{\xi}_i$$

s.t.  $w^{\mathsf{T}}\varphi(\mathbf{x}_i) \ge \rho - \xi_i$ ;  $\xi_i \ge 0$ ,  $\forall i = 1, 2, \dots, N$  (1) 其中 $w \in \chi, \rho \in \mathbb{R}$ .  $\nu \in (0, 1]$ 为预先定义的正则化参数, 是被 允许的最大误警率。引入拉格朗日乘子 $\alpha_i \ge 0, \beta_i \ge 0$ ,则有:



919

$$L(\boldsymbol{w},\boldsymbol{\xi},\boldsymbol{\rho},\boldsymbol{\alpha},\boldsymbol{\beta}) = \frac{1}{2} \boldsymbol{w}^{\mathrm{T}} \boldsymbol{w} - \boldsymbol{\rho} + \frac{1}{\nu N} \sum_{i=1}^{N} \omega_{i} \boldsymbol{\xi}_{i} - \sum_{i=1}^{N} \alpha_{i} \left( \boldsymbol{w}^{\mathrm{T}} \boldsymbol{\varphi}(\boldsymbol{x}_{i}) - \boldsymbol{\rho} + \boldsymbol{\xi}_{i} \right) - \sum_{i=1}^{N} \beta_{i} \boldsymbol{\xi}_{i} \quad (2)$$

由Karush-Kuch-Tucker条件可得:

$$\begin{cases} \frac{\partial L}{\partial \boldsymbol{w}} = 0 \rightarrow \boldsymbol{w} = \sum_{i=1}^{N} \alpha_i \boldsymbol{\varphi}(\boldsymbol{x}_i) \\ \frac{\partial L}{\partial \boldsymbol{\xi}} = 0 \rightarrow \alpha_i = \frac{1}{\nu N} \omega_i - \beta_i \\ \frac{\partial L}{\partial \boldsymbol{\rho}} = 0 \rightarrow \sum_{i=1}^{N} \alpha_i = 1 \end{cases}$$
(3)

将式(3)代入式(2),可得原优化目标的对偶形式为: min  $\alpha^{T}Q\alpha$ 

s.t. 
$$0 \le \alpha_i \le \frac{1}{\nu N} \omega_i, \sum_{i=1}^N \alpha_i = 1$$
 (4)

其中: $\boldsymbol{\alpha} = [\alpha_1, \alpha_2 \cdots, \alpha_N]^T, Q$ 为训练样本数据组成的核矩阵,  $Q(i,j) = K(\boldsymbol{x}_i, \boldsymbol{x}_j)$ 。本文使用高斯核函数 $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = exp(-||\boldsymbol{x}_i - \boldsymbol{x}_j||^2/(2\sigma^2))$ 作为非线性映射函数, $\sigma$ 为核函数的参数。进一步解算式(4),可求得 $\boldsymbol{\alpha} = [\alpha_1, \alpha_2 \cdots, \alpha_N]^T$ 。继而可求得以下决策函数:

$$f(\boldsymbol{x}) = \operatorname{sign}\left(\sum_{i \in SV_s} \alpha_i K(\boldsymbol{x}, \boldsymbol{x}_i) - \sum_{i, j \in SV_s} \alpha_j K(\boldsymbol{x}_i, \boldsymbol{x}_j)\right)$$
(5)

其中*SVs*表示支持向量的索引。对于测试样本x,如果f(x)的 值为1,则说明x为正常样本;如果f(x)的值为-1,则说明x为 故障样本。由式(4)可知,权重 $\omega_i$ 越小,则对应的拉格朗日乘 子 $\alpha_i$ 越趋近于零,说明样本 $x_i$ 与分类边界的距离越远,对分类 边界的影响越小。因此,只要对潜在的故障样本分配较小的 权重,即可起到增强算法鲁棒性的作用。

## 1.2 基于局部密度的加权方式

现将文献[27]的权重求取方式简述如下。

设C为数据集的几何中心,则样本 $x_i$ 与C之间的欧氏距 离为:

$$d(\mathbf{x}_i, \mathbf{C}) = \left\| \varphi(\mathbf{x}_i) - \frac{1}{N} \sum_{j=1}^N \varphi(\mathbf{x}_j) \right\|^2 = \sqrt{K(\mathbf{x}_i, \mathbf{x}_i) - \frac{2}{N} \sum_{j=1}^N K(\mathbf{x}_i, \mathbf{x}_j) + \frac{1}{N^2} \sum_{j,k=1}^N K(\mathbf{x}_j, \mathbf{x}_k)} \quad (6)$$

令 $\hat{d}(\mathbf{x}_i, \mathbf{C})$ 表示 $\mathbf{x}_i$ 与 $\mathbf{C}$ 之间正则化的距离,则:

$$\hat{d}(\boldsymbol{x}_{i}, \boldsymbol{C}) = \frac{d(\boldsymbol{x}_{i}, \boldsymbol{C})}{\max_{i=1, 2, \cdots, N} d(\boldsymbol{x}_{i}, \boldsymbol{C})}$$
(7)

则样本 $x_i$ 对应的权重 $\omega_i$ 可由式(8)求得:

$$p_i = e^{-\frac{(d(x_i, C))^2}{2}}$$
 (8)

由上述表述可知,文献[27]所提算法的权重求取方式存 在一定缺陷。当训练集的样本构成比较复杂,比如存在多个 小范围聚集的情况,则样本与训练集中心的距离并不能衡量 该样本与数据集边界的位置关系,也就是采用式(8)来求取权 重并不科学。为此,本文提出一种基于局部密度的加权方式。 其基本原理可简述如下。

 $x_i$ (*i* = 1, 2, …, *N*)表示训练数据集**D**的第*i*个样本,集合 { $x_{i,1}, x_{i,2}, ..., x_{i,k}$ }表示离样本 $x_i$ 最近的*k*个训练样本,按照距 离从小到大的顺序排列。即 $x_{i,i}$ (*j* = 1, 2, …, *k*)  $\in$  *KNN*( $x_i$ ),其 中, KNN为K-Nearest Neighbor的缩写。考虑到数据的真实分布特点, k值一般取为以当前样本 $x_i$ 为中心, 以 $L \times 2\%$ 为半径的球体内所包含训练样本的个数, 其中L为全体训练样本中心到距离最远的训练样本之间的马氏距离。用 $d(x_i, x_{i,j})$ 表示样本 $x_i$ 和样本 $x_i$ 之间的距离。 $\bar{x}_i$ 是样本 $x_i$ 的k个邻近样本的

中心,即
$$\bar{\mathbf{x}}_{i} = \frac{1}{k} \sum_{j=1}^{k} \mathbf{x}_{i,j^{\circ}} \quad$$
则:  
$$d(\mathbf{x}_{i}, \bar{\mathbf{x}}_{i}) = \left\| \varphi(\mathbf{x}_{i}) - \frac{1}{k} \sum_{j=1}^{k} \varphi(\mathbf{x}_{i,j}) \right\|^{2} = \sqrt{K(\mathbf{x}_{i}, \mathbf{x}_{i}) - \frac{2}{k} \sum_{j=1}^{k} K(\mathbf{x}_{i}, \mathbf{x}_{i,j}) + \frac{1}{k^{2}} \sum_{j,m=1}^{k} K(\mathbf{x}_{i,j}, \mathbf{x}_{i,m})} \quad (9)$$

令 $\hat{d}(\mathbf{x}_i, \bar{\mathbf{x}}_i)$ 表示样本 $\mathbf{x}_i$ 与其近邻中心 $\bar{\mathbf{x}}_i$ 之间的正则化距离,则:

$$\hat{d}(\boldsymbol{x}_i, \bar{\boldsymbol{x}}_i) = \frac{d(\boldsymbol{x}_i, \bar{\boldsymbol{x}}_i)}{\max_{i=1, 2, \cdots, N} d(\boldsymbol{x}_i, \bar{\boldsymbol{x}}_i)}$$
(10)

由于故障样本与正常样本的形成机制不同,一般情况下 故障样本均远离正常样本。当 $\hat{d}(\mathbf{x}_i, \bar{\mathbf{x}}_i)$ 越大,说明样本 $\mathbf{x}_i$ 与其 他样本的距离越大,则 $\mathbf{x}_i$ 为故障样本的可能性越大,此时需要 对它分配较小的权值。则令

$$\omega_i = 1 - (\hat{d}(\boldsymbol{x}_i, \bar{\boldsymbol{x}}_i))^2 / 2 \tag{11}$$

式(11)即为基于局部密度的 WOCSVM 算法的权重求取 方式。

## 1.3 基于马氏距离的 k 近邻样本的求取

在 1.2 节中,求取离样本  $\mathbf{x}_i$ 最近的 k 个训练样本  $\{\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \dots, \mathbf{x}_{i,k}\}$ 时,考虑到训练数据集中可能掺杂有无法确 定特征之间是否有联系的故障样本这一情况,如果采用传统 的欧氏距离来度量,则无法保证计算的科学性。为此,本文提 出用马氏距离(Mahalanobis Distance, MD)来度量两样本之间 距离的远近。

马氏距离是统计学家马哈拉诺比斯定义的一种距离,它 与欧氏距离相比,能有效地消除各变量之间的关联性并且和 数据的量纲没有关系。给定训练样本集**D**,对于第*i*个样本**x**<sub>i</sub> 和第*j*个样本**x**<sub>i</sub>之间的马氏距离可由下式计算得出:

$$d_{ii}^{M} = \sqrt{(\boldsymbol{x}_{i} - \boldsymbol{x}_{j})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_{i} - \boldsymbol{x}_{j})}$$
(12)

其中*S*表示总体协方差矩阵。由于样本协方差是总体协方差 的无偏估计,为便于计算,常用样本协方差矩阵代替总体协方 差矩阵。当*S*为单位矩阵时,马氏距离变为欧氏距离。

为方便叙述,将此算法简称为LD-WOCSVM。基于以上 分析,可将算法流程简述如下。

步骤1 输入训练数据集 $\{x_i\}_{i=1}^N$ ,核函数参数 $\sigma$ 和正则 化参数 $\nu_{\circ}$ 

步骤2 由式(12)计算样本的k个邻近样本。

步骤3 由式(11)求得权重,代入式(4),由序列最小最优(Sequential Minimal Optimization, SMO)算法求解式(4),继而由式(5)求得决策函数。

步骤4 由式(5)预测新样本是正常还是故障。

## 2 基于快速聚类的LD-WOCSVM

为了能进一步增强算法的鲁棒性,先采用一种快速聚类 分析算法来获取样本分布特点,再通过对可能的故障样本分 配较小的权重来实现自适应加权。基于快速聚类算法的权重



第40卷

求取如下。

Rodriguez 等<sup>[30]</sup>在 2014 年提出了一种很简洁的快速聚类 算法,该算法的参数很容易确定,通过分析,该算法的复杂度 为 $O(N^2)$ 。本文通过对该算法进行进一步改进,可以快速准 确地识别出一部分故障样本,从而使更大限度地提高算法鲁 棒性成为可能。该算法通过定义 $\rho_i$ 和 $\delta_i$ 两个参数来衡量对应 的样本 $\mathbf{x}_i$ 是否为故障样本。其中, $\rho_i$ 表示数据点 $\mathbf{x}_i$ 的局部密 度, $\delta_i$ 表示数据点 $\mathbf{x}_i$ 到高局部密度的距离。 $\rho_i$ 定义为:

$$\rho_i = \sum \chi(d_{ij} - d_c) \tag{13}$$

其中: 当x < 0时, $\chi(x) = 1$ ; 当 $x \ge 0$ 时, $\chi(x) = 0$ 。 $d_c$ 为截断距 离(超参数), $d_{ij}$ 为样本 $x_j$ 到样本 $x_i$ 的距离。则 $\rho_i$ 表示到样本  $x_i$ 的距离小于 $d_c$ 的样本的个数。对于 $d_c$ 值的选取,一种常用 的方法是选择的 $d_c$ 应能使平均每个样本点的附近样本个数为 数据集总样本个数的1%~2%。在本文中,使用k近邻算法获 取 $d_c$ 的值,具体步骤如下。

*N*为数据集总样本个数,*s* = *N*×1%取整数获取*s*值。用  $x_{i,s}(i = 1, 2, ..., N)$ 表示 $x_i$ 的第*s*个近邻样本, $d(x_i, x_{i,s})$ 表示样 本 $x_i$ 和样本 $x_{i,s}$ 之间的欧氏距离。使用Q(x)来表示数据集  $\{x_i\}_{i=1}^{N}$ 中每个样本与该样本第*s*个近邻样本之间距离的集 合,即 $Q(x) = \{d(x_i, x_{i,s})\}_{i=1}^{N}$ 。则 $d_i$ 的计算公式如下:

$$\begin{aligned} d_{c} &= \frac{\min(Q(\boldsymbol{x})) + \max(Q(\boldsymbol{x}))}{2} \end{aligned} \tag{14} \\ 动于参数\delta_{i},其计算公式为: \\ \delta_{i} &= \min_{i:p_{i}>p_{i}} (d_{ij}) \end{aligned} \tag{15}$$

可知, $\delta_i$ 表示数据点 $\mathbf{x}_i$ 到高局部密度的最小距离。对于 局部密度最大的点,设置 $\delta_i = \max_j (d_{ij})$ 。通过上述分析计算, 可以得到数据集中所有样本对应的 $\rho_i$ 和 $\delta_i$ 。从对参数定义的 规则可以得到,拥有较大 $\rho_i$ 值和较大 $\delta_i$ 的样本,可以认为是训 练集中某个类簇的中心;而对拥有较小 $\rho_i$ 值和较大 $\delta_i$ 的样本, 认为是训练集中存在的故障样本(噪声);对拥有较大 $\rho_i$ 值和 较小 $\delta_i$ 的样本,认为是分布在距离最近的类簇中心的周围;对 拥有较小 $\rho_i$ 值和较小 $\delta_i$ 的样本,认为分布在距离该样本最近 的类簇的边缘。为了使故障样本分配到更小的权重,令

$$\boldsymbol{\omega}_{i} = \mathrm{e}^{-(\delta_{i}/(\rho_{i} + \Delta))^{2}} \tag{16}$$

其中: $\Delta$ 是一个很小的正数,目的是为了避免出现分母为零的 情况,一般取 $\Delta$ =10<sup>-6</sup>。为了便于叙述,将此算法简称为 FCLD-WOCSVM。基于以上分析,可将算法FCLD-WOCSVM 的流程简述如下。

步骤1 输入训练数据集 $\{x_i\}_{i=1}^{N}$ ,核函数参数 $\sigma$ 和正则 化参数 $\nu_{\circ}$ 

步骤2 由式(13)和(15)计算 $\rho_i$ 和 $\delta_i$ ,再由式(16)求得 权重。

步骤3 将权重代入式(4),由SMO算法求解式(4),继而 由式(5)求得决策函数。

步骤4 由式(5)预测新样本是正常还是故障。

3 实验分析

为了验证本文所提的LD-WOCSVM和FCLD-WOCSVM两种算法的有效性,本文将从以下两个方面加以论述。首先选用4个常用的基准数据集,用来测试本文所提算法的分类性能,同时参与比较的还有文献[27]中的A-WOCSVM算法、文

献[31]中的WOCSVM算法、文献[32]中的加权支持向量数据 描述(Weighted Support Vector Data Description, WSVDD)算 法、标准 OCSVM 算法、文献 [33] 中的局部异常因子(Local Outlier Factor, LOF)算法和文献[34]中的孤立森林(Isolation FOREST, IFOREST)算法。仿真结果表明,本文所提算法相 较于另6种算法在分类性能上均有所增强。其次,将LD-WOCSVM 和 FCLD-WOCSVM 两种算法用于某型涡轴发动机 故障检测,实验结果表明,本文所提算法能够构建性能较好的 分类器,可作为故障检测的候选算法。本文采用G-mean (Geometric mean)值和 AUC(Area Under the Curve)值作为最 终的评价指标,两种指标的值越大,则说明故障检测性能越 好。为了减少样本随机选取对分类结果的影响,对每种算法 均独立运行50次并取其平均值和方差作为最终结果。如前 所述,本文采用高斯核函数作为非线性映射函数。核函数参 数 $\sigma$ 和正则化参数 $\nu$ 通过5折交叉验证进行优化,参数 $\sigma$ 和 $\nu$ 的搜索区间分别为{2<sup>-10</sup>, 2<sup>-9</sup>, ..., 2<sup>-4</sup>}和{10<sup>-4</sup>, 10<sup>-3</sup>, ..., 10<sup>3</sup>}。 本文仿真实验是在 Intel Core i5-7400 CPU @ 3.00 GHz 的 Windows 10上进行的,采用 Matlab R2017a 编程。

3.1 基准数据集测试

3.1.1 仿真设置

耒

本文选取 UCI(University of California Irvine)数据集中的 4个常用数据集的基本信息在表1中列出。

实验中使月	目的4个	UCI数据	·集的基;	本信息
-------	------	-------	-------	-----

Tab. 1 Fundamental information of				
4 UCI datasets used in the experiment				
***********	代号	属性数(样本	实例数	
如16年石小		量的维数)		
abalone	Set 1	9	4 177	
Banknote Authentication	Set 2	5	1 372	
wine equality-white	Set 3	12	4 898	
Wireless Indoor Localization	Set 4	8	2 000	

在本次实验中,对于每个数据集,将样本最多的那一类作为正常类,将剩余所有类作为故障类,并且将标签分别改为+1和-1。在正常类中随机抽取70%的样本作为训练集的一部分,在故障类中分别随机抽取占整个训练集5%和10%的样本作为训练集另一部分,并且将其标签人为改为+1。测试集的样本构成则为剩余的正常样本和故障样本。

故障样本占整个训练集5%和10%时所得实验结果列于 表2。其中,每组实验的最优结果用黑体标出。

#### 3.1.2 结果分析

由表2可以看出,在Set1和Set3两个数据集上,故障样本占比10%且使用AUC指标衡量时,LD-WOCSVM算法的性能最优,其余情况下,均是FCLD-WOCSVM算法具有最优的性能。以在Set4数据集上,故障样本占比10%时为例,与A-WOCSVM算法相比,FCLD-WOCSVM算法的G-mean值提高了12.1%。在Set3数据集上,故障样本占比5%时,与A-WOCSVM算法相比,LD-WOCSVM算法的AUC值降低了0.01%,但在使用G-mean值衡量时,与A-WOCSVM算法相比,LD-WOCSVM算法的值增加了0.26%。在Set1数据集上,故障样本占比10%时,与A-WOCSVM算法的值增加了0.26%。在Set1数据集上,故障样本占比10%时,与A-WOCSVM算法的G-mean值降低了0.7%,但在使用AUC值衡量时,与A-WOCSVM算法相比,LD-WOCSVM算法的值提高了0.5%。且在其余情况下,LD-WOCSVM算法的性能均优于A-WOCSVM算法的性能。

## 黄功等:基于局部密度的加权一类支持向量机算法及其在涡轴发动机故障检测中的应用

		故障样本数在训练集中的占比				
数据集	算法	5%		10	)%	
		AUC	G-mean	AUC	G-mean	
Set 1	FCLD-WOCSVM	$0.\ 659\ 8\pm 0.\ 010\ 6$	$0.\ 554\ 2\pm 0.\ 012\ 5$	0.6456±0.0119	$0.\ 524\ 7\pm 0.\ 013\ 8$	
	LD-WOCSVM	$0.6591 \pm 0.0089$	$0.5527 \pm 0.0143$	$0.\ 646\ 2\pm 0.\ 011\ 4$	$0.4215 \pm 0.0143$	
	A-WOCSVM	$0.6582 \pm 0.0078$	$0.5519 \pm 0.0161$	$0.6428 \pm 0.0089$	$0.4245 \pm 0.0117$	
	WOCSVM	$0.6578 \pm 0.0082$	$0.5508 \pm 0.0116$	$0.6446 \pm 0.0095$	$0.4213 \pm 0.0115$	
	OCSVM	$0.6584 \pm 0.0080$	$0.5475 \pm 0.0133$	$0.6458 \pm 0.0092$	$0.4220 \pm 0.0123$	
	WSVDD	$0.\ 658\ 6\pm 0.\ 007\ 2$	$0.\ 552\ 4\pm 0.\ 020\ 7$	0. $6451 \pm 0.0084$	$0.4217 \pm 0.0128$	
	LOF	$0.6307 \pm 0.0071$	$0.\ 551\ 2\pm 0.\ 013\ 2$	0. 643 3 ± 0. 007 6	$0.4133 \pm 0.0121$	
	IFOREST	$0.6415 \pm 0.0058$	0. 541 1 ± 0. 013 3	0. $6425 \pm 0.0083$	$0.4207 \pm 0.0153$	
	FCLD-WOCSVM	$0.\ 996\ 3\pm 0.\ 002\ 3$	$0.\ 887\ 4\pm 0.\ 024\ 5$	$0.\ 953\ 4\pm 0.\ 012\ 5$	$0.\ 813\ 0\pm 0.\ 015\ 8$	
	LD-WOCSVM	0. 995 8 ± 0. 003 3	0. $8660 \pm 0.0202$	$0.9465 \pm 0.0142$	0. 808 9 $\pm$ 0. 017 4	
	A-WOCSVM	$0.9958 \pm 0.0040$	0. $863.6 \pm 0.030.3$	$0.9453 \pm 0.0172$	0. 807 8 $\pm$ 0. 016 8	
8.1.2	WOCSVM	0. 994 0 ± 0. 003 2	0. $8639 \pm 0.0277$	$0.9473 \pm 0.0127$	0. 808 3 $\pm$ 0. 018 6	
Set 2	OCSVM	0. 996 1 ± 0. 003 7	0. 868 $5 \pm 0.0261$	0. 946 9 ± 0. 013 1	$0.8115 \pm 0.0135$	
	WSVDD	$0.9959 \pm 0.0026$	0. 868 $2 \pm 0.0291$	0.9470±0.0113	0. $8079 \pm 0.0142$	
	LOF	0. 995 1 ± 0. 002 6	0.8671±0.0259	0. 941 1 ± 0. 013 2	0. $8036 \pm 0.0143$	
	IFOREST	0. 992 1 ± 0. 004 8	$0.\ 853\ 2\pm 0.\ 021\ 8$	$0.\ 937\ 7\pm 0.\ 012\ 8$	$0.\ 791\ 3\pm 0.\ 013\ 2$	
	FCLD-WOCSVM	$0.\ 973\ 4\pm 0.\ 002\ 2$	$0.\ 537\ 9\pm 0.\ 007\ 3$	0. $5642 \pm 0.0053$	$0.\ 539\ 9\pm 0.\ 006\ 2$	
	LD-WOCSVM	0. $8030 \pm 0.0056$	$0.\ 536\ 5\pm 0.\ 007\ 5$	$0.\ 565\ 2\pm 0.\ 005\ 2$	$0.5390 \pm 0.0089$	
	A-WOCSVM	0. $8039 \pm 0.0057$	0. 535 1 $\pm$ 0. 006 7	$0.\ 562\ 9\pm 0.\ 005\ 0$	$0.5383 \pm 0.0069$	
Set 2	WOCSVM	0. $8022 \pm 0.0074$	0. 537 3 $\pm$ 0. 008 4	$0.\ 563\ 7\pm 0.\ 005\ 6$	$0.5379 \pm 0.0077$	
Set 5	OCSVM	$0.\ 802\ 5\pm 0.\ 005\ 3$	0. 536 8 $\pm$ 0. 008 4	0. 563 9 $\pm$ 0. 004 6	$0.5381 \pm 0.0085$	
	WSVDD	0. $8021 \pm 0.0036$	0. 537 1 ± 0. 007 9	0. $5635 \pm 0.0051$	$0.5376 \pm 0.0065$	
	LOF	0. $8015 \pm 0.0043$	$0.5305 \pm 0.0062$	0. $5632 \pm 0.0048$	$0.\ 530\ 4\pm 0.\ 007\ 3$	
	IFOREST	0. $8007 \pm 0.0052$	$0.5258 \pm 0.0075$	0. $5593 \pm 0.0037$	$0.\ 531\ 1\pm 0.\ 005\ 9$	
	FCLD-WOCSVM	$0.\ 989\ 2\pm 0.\ 003\ 7$	$0.9211 \pm 0.0162$	$0.\ 975\ 7\pm 0.\ 006\ 5$	$0.\ 907\ 6\pm 0.\ 013\ 4$	
	LD-WOCSVM	0.9846±0.0050	$0.9104 \pm 0.0181$	$0.\ 973\ 4\pm 0.\ 007\ 1$	$0.8100 \pm 0.0255$	
	A-WOCSVM	0. 983 4 ± 0. 004 3	$0.9103 \pm 0.0187$	$0.9715 \pm 0.0066$	0. $8097 \pm 0.0212$	
Set 4	WOCSVM	$0.9836 \pm 0.0049$	$0.9144 \pm 0.0168$	$0.9717 \pm 0.0066$	0. $8086 \pm 0.0327$	
Jet 4	OCSVM	$0.9834 \pm 0.0040$	$0.\ 908\ 3\pm 0.\ 017\ 2$	$0.9739 \pm 0.0058$	0.8164±0.0229	
	WSVDD	$0.\ 984\ 2\pm 0.\ 003\ 1$	$0.\ 915\ 3\pm 0.\ 010\ 8$	$0.9741 \pm 0.0048$	$0.\ 813\ 6\pm 0.\ 012\ 4$	
	LOF	$0.9836 \pm 0.0041$	$0.9033 \pm 0.0205$	$0.9733 \pm 0.0082$	$0.\ 808\ 6\pm 0.\ 021\ 0$	
	IFOREST	$0.9713 \pm 0.0038$	$0.9014 \pm 0.0201$	$0.9665 \pm 0.0073$	$0.8011 \pm 0.0155$	

#### 表 2 故障样本数占整个训练集 5% 或 10% 时,在 4 个 UCI 数据集上的实验结果 b. 2 Experimental results on 4 UCI datasets when fault samples account for 5% or 10% of the total training

总而言之,在本文选取的4个数据集上,与A-WOCSVM、WOCSVM、WSVDD、标准OCSVM、LOF、IFOREST等算法相比,算法FCLD-WOCSVM和算法LD-WOCSVM的性能更优。这说明了本文提出的FCLD-WOCSVM算法和LD-WOCSVM算法在解决故障检测问题上的科学性和有效性。

## 3.2 涡轴发动机故障检测

#### 3.2.1 背景知识

涡轴发动机故障主要发生在压气机、燃气涡轮和动力涡

轮三个气路部件处,因此对这三个部件进行故障检测尤为重要。本文以T700发动机为研究对象,在部件级数学模型的基础上以性能参数蜕化一定的量来模拟相应的故障。T700发动机模型简图如图1所示,分别用代号3、42、5和9来表示压气机出口截面、燃气涡轮出口截面、功率涡轮出口截面以及尾喷管出口截面。在图1中,GT(Gas Turbine)表示燃气涡轮,PT (Power Turbine)表示动力涡轮。



Fig. 1 Principle of T700 turboshaft engine with state number

在参照文献[35]的基础上,本文选择5个性能参数:压气 机流量 $W_c$ 、燃气涡轮流量 $W_{rc}$ 、压气机效率 $\eta_c$ 、燃气涡轮效率  $\eta_{rc}$ 、动力涡轮效率 $\eta_{rp}$ 。经过综合比较各参数的测量难易程 度,本文选择11个传感器测量参数:压气机出口总温*T*<sub>3</sub>、压气机出口总压*P*<sub>3</sub>、飞行高度*H*、前飞速度*v*<sub>x</sub>、燃油流量*W*<sub>f</sub>、燃气涡轮船口总压*P*<sub>42</sub>、燃气涡轮出口总压



P<sub>42</sub>、动力涡轮输出转速PNP、动力涡轮出口总温T<sub>5</sub>、动力涡轮 出口总压P<sub>5</sub>。本文在全飞行包线范围内,稳定工况下采集数 据。限于篇幅,只设置单一故障模式。各个性能参数蜕化量 及对应样本数列于表3。

表 3 性能参数的主要信息 Tab 3 Main information about performance parameters

Tuble Main mornation about performance parameters					
性能参数	蜕化因子/%	样本数			
正常	—	3 300			
$oldsymbol{\eta}_{C}$	1.0	3 300			
W <sub>c</sub>	1.0	3 3 1 1			
$oldsymbol{\eta}_{TG}$	2.0	3 300			
$W_{TG}$	1.5	3 300			
${m \eta}_{\scriptscriptstyle TP}$	2.0	3 300			

使用 LD-WOCSVM / FCLD-WOCSVM 算法进行涡轴发动 机故障检测时,其流程可简述为:

1) 获取涡轴发动机工作状态的数据集,划分数据集为训 练集和测试集。

2) 在训练集中,利用5 折交叉验证法求得核函数参数 σ 和正则化参数 ν。

3) 从训练集中随机抽取部分数据用于训练模型。利用 式(11)/(16)求得权重,将权重代入式(4)并利用SMO算法求 解,再利用式(5)求得决策函数。

4) 将测试集中样本代入决策函数,即可得测试结果。

### 3.2.2 参数设置

在本次实验中,根据采集得到的5种故障样本,分别设计 5组实验。由表4中性能参数所列顺序,将对应的5组实验依 次命名为Case1、Case2、Case3、Case4和Case5。将正常样本 作为正常类,并将其标签改为+1;将5种故障样本作为故障 类,并将其标签改为-1。对每组仿真实验,从正常类中随机抽 取40%的样本作为训练集的一部分,从对应的故障类中随机 抽取占比整个训练集1%的样本作为训练集另一部分,并将 该故障样本的标签人为改成+1,即:训练集中正常样本数 为1320,故障样本数为13。测试集的样本构成则为剩余的正 常样本和故障样本。为了检验算法在故障样本占比较高时的 分类性能,本文还设计一组故障样本占比5%时的仿真实验, 此时训练集中正常样本数为1320,故障样本数为69。

故障样本占整个训练集 1% 和 5% 时的实验结果列于表 4,其中,每组实验的最优结果用黑体标出。在这一部分,选取 了 4 个对比算法:文献[27]中的 A-WOCSVM 算法、文献[31] 中的 WOCSVM 算法、文献[32]中的 WSVDD 算法以及标准 OCSVM算法。

#### 表4 故障样本数占整个训练集1%或5%时,在涡轴发动机数据集上的实验结果

Tab. 4 Experimental results on turbo shaft engine datasets when fault samples account for 1% or 5% of the total training set

		故障样本数在训练集中的占比			
数据集	算法 –	1%		59	%
		AUC	G-mean	AUC	G-mean
Case1	FCLD-WOCSVM	0.9883±0.0089	0. 799 8 ± 0. 021 2	$0.\ 934\ 8\pm 0.\ 007\ 5$	0.6785±0.0122
	LD-WOCSVM	0.9810±0.0113	0. 824 1 ± 0. 051 3	$0.9336 \pm 0.0089$	$0.\ 691\ 3\pm 0.\ 023\ 2$
	A-WOCSVM	$0.9795 \pm 0.0133$	$0.7962 \pm 0.0244$	$0.9322 \pm 0.0081$	$0.6770 \pm 0.0164$
	WOCSVM	$0.9776 \pm 0.0140$	$0.7930 \pm 0.0229$	$0.9332 \pm 0.0087$	$0.6752 \pm 0.0114$
	OCSVM	$0.9748 \pm 0.0036$	0. $8002 \pm 0.0274$	$0.9339 \pm 0.0073$	$0.6763 \pm 0.0167$
	WSVDD	$0.4761 \pm 0.0012$	$0.4251 \pm 0.0325$	0. $4397 \pm 0.0053$	$0.\ 407\ 6\pm 0.\ 018\ 2$
	FCLD-WOCSVM	$0.\ 979\ 7\pm 0.\ 003\ 8$	$0.\ 838\ 3\pm 0.\ 034\ 1$	$0.9351 \pm 0.0063$	0. 673 5 ± 0. 023 9
	LD-WOCSVM	$0.\ 975\ 5\pm 0.\ 014\ 4$	0. $8313 \pm 0.0359$	$0.\ 939\ 6\pm 0.\ 008\ 1$	$0.\ 684\ 9\pm 0.\ 010\ 3$
C 2	A-WOCSVM	$0.9794 \pm 0.0118$	$0.\ 837\ 4\pm 0.\ 034\ 9$	$0.9286 \pm 0.0094$	$0.6699 \pm 0.0301$
Case 2	WOCSVM	$0.9513 \pm 0.0264$	$0.\ 822\ 4\pm 0.\ 037\ 3$	0. 929 1 ± 0. 009 5	$0.6645 \pm 0.0230$
	OCSVM	$0.9473 \pm 0.0286$	$0.\ 830\ 1\pm 0.\ 037\ 2$	$0.9299 \pm 0.0113$	$0.6623 \pm 0.0255$
	WSVDD	$0.4319 \pm 0.0182$	$0.4122 \pm 0.0265$	$0.4105 \pm 0.0012$	$0.4022 \pm 0.0206$
	FCLD-WOCSVM	$0.9952 \pm 0.0016$	$0.\ 880\ 3\pm 0.\ 035\ 7$	$0.\ 996\ 4\pm 0.\ 002\ 1$	$0.\ 773\ 6\pm 0.\ 018\ 2$
Case 3	LD-WOCSVM	$0.\ 996\ 5\pm 0.\ 001\ 4$	$0.\ 866\ 5\pm 0.\ 034\ 9$	$0.9826 \pm 0.0045$	$0.7708 \pm 0.0176$
	A-WOCSVM	$0.9927 \pm 0.0016$	0.8701±0.0319	$0.9810 \pm 0.0053$	$0.\ 768\ 2\pm 0.\ 015\ 9$
	WOCSVM	$0.9944 \pm 0.0018$	$0.\ 879\ 7\pm 0.\ 039\ 0$	$0.9820 \pm 0.0050$	$0.7723 \pm 0.0201$
	OCSVM	$0.9927 \pm 0.0023$	0. 873 1 $\pm$ 0. 036 7	$0.9806 \pm 0.0057$	$0.7690 \pm 0.0179$
	WSVDD	0. 463 1 ± 0. 001 8	0. 406 1 $\pm$ 0. 017 8	0. 443 7 $\pm$ 0. 002 5	$0.\ 387\ 2\pm 0.\ 016\ 5$
	FCLD-WOCSVM	$0.9917 \pm 0.0029$	$0.\ 910\ 6\pm 0.\ 023\ 7$	$0.\ 946\ 7\pm 0.\ 007\ 5$	$0.\ 782\ 9\pm 0.\ 018\ 8$
	LD-WOCSVM	$0.\ 992\ 1\pm 0.\ 002\ 5$	$0.9054 \pm 0.0175$	$0.9434 \pm 0.0083$	$0.7657 \pm 0.0223$
Case 4	A-WOCSVM	$0.9906 \pm 0.0030$	$0.9076 \pm 0.0230$	$0.9417 \pm 0.0082$	$0.\ 777\ 4\pm 0.\ 018\ 3$
	WOCSVM	$0.9867 \pm 0.0063$	$0.\ 909\ 1\pm 0.\ 022\ 7$	$0.9444 \pm 0.0086$	$0.\ 777\ 8\pm 0.\ 015\ 2$
	OCSVM	$0.9842 \pm 0.0085$	$0.\ 9106 \pm 0.\ 0219$	0. 944 1 $\pm$ 0. 006 3	$0.\ 780\ 5\pm0.\ 019\ 0$
	WSVDD	$0.4823 \pm 0.0040$	$0.\ 432\ 2\pm 0.\ 018\ 3$	0. $4527 \pm 0.0052$	$0.4102 \pm 0.0115$
	FCLD-WOCSVM	$0.9616 \pm 0.0169$	$0.\ 829\ 8\pm 0.\ 040\ 8$	0. $8809 \pm 0.0084$	0. 669 $3 \pm 0.0141$
	LD-WOCSVM	$0.\ 965\ 2\pm 0.\ 011\ 1$	$0.8486\pm0.0391$	0. $8800 \pm 0.0111$	$0.\ 670\ 6\pm 0.\ 020\ 5$
Casa5	A-WOCSVM	$0.\ 955\ 4\pm 0.\ 015\ 7$	$0.\ 823\ 1\pm 0.\ 042\ 0$	0. $8774 \pm 0.0093$	$0.\ 667\ 3\pm 0.\ 016\ 3$
CaseJ	WOCSVM	$0.\ 953\ 4\pm 0.\ 019\ 0$	0. $8257 \pm 0.0413$	$0.\ 886\ 8\pm 0.\ 020\ 5$	$0.\ 665\ 8\pm 0.\ 012\ 3$
	OCSVM	$0.9431 \pm 0.0133$	$0.\ 825\ 6\pm 0.\ 038\ 6$	$0.\ 878\ 5\pm 0.\ 011\ 6$	0. 665 3 ± 0. 011 3
	WSVDD	$0.4479 \pm 0.0201$	$0.\ 433\ 5\pm 0.\ 032\ 2$	$0.\ 424\ 6\pm 0.\ 013\ 1$	$0.\ 407\ 3\pm 0.\ 014\ 2$



923

#### 3.2.3 结果分析

由表4可以看出,对比算法WSVDD的性能最差。无论是用AUC还是G-mean来度量,所得结果均小于0.5,说明该算法不适合用来解决涡轴发动机故障检测问题。

由表4还可以看出,在Case1数据集中:使用AUC值作指 标时FCLD-WOCSVM算法的性能最优,使用G-mean值作指标 时LD-WOCSVM算法性能最优,且本文所提出的两种算法的 性能均优于对比算法 A-WOCSVM。在 Case 2 数据集中:故障 样本占比1%时FCLD-WOCSVM算法的性能最优, A-WOCSVM 算法性能次优;故障样本占比5%时LD-WOCSVM 算法的性能最优,FCLD-WOCSVM算法的性能次优。在Case 3数据集上:故障样本占比1%且使用AUC指标衡量时,LD-WOCSVM算法性能最优,其余情况下均是FCLD-WOCSVM算 法的性能最优。在Case 4数据集中:故障样本占比5%时 FCLD-WOCSVM 算法性能最优;占比1%时 FCLD-WOCSVM 算法的性能次优。在Case 5数据集中:故障样本占比 5% 时 WOCSVM算法性能最优;其余情况下,算法LD-WOCSVM性 能最优。从统计角度看,5个数据集,2种指标和2种故障样本 比例一共20种情况,其中:算法FCLD-WOCSVM性能最优的 次数为9,次优的次数为10:算法LD-WOCSVM性能最优的次 数为9,次优的次数为2;算法A-WOCSVM性能最优的次数为 0,次优的次数为2;算法WOCSVM性能最优的次数为1,次优 的次数为3:算法 OCSVM 性能最优的次数为1,次优的次 数为3。

从以上分析可以看出,在发动机故障检测方面,算法。 FCLD-WOCSVM性能最优,算法LD-WOCSVM性能次优,而对 比算法A-WOCSVM和WSVDD的性能最差。这进一步说明不 本文所提算法FCLD-WOCSVM和算法LD-WOCSVM和以作为 涡轴发动机故障检测的两种候选算法。

### 4 结语

本文通过在算法 A-WOCSVM 的基础上提出两种改进算法:FCLD-WOCSVM和LD-WOCSVM,并将其运用在涡轴发动机故障检测方面,可得以下结论:

1)使用马氏距离来选取每个样本对应的k近邻样本,再 通过计算样本与局部密度中心而不是全局密度中心的距离提 出的LD-WOCSVM算法,能够使训练的分类器更精确地描述 数据样本边界,分类效果更好。

2) 通过对一种快速聚类分析算法进行改进,使该算法能 更准确地识别故障数据,进而计算出更合适的权重而提出的 FCLD-WOCSVM算法,能够最大限度地降低故障数据对分类 边界的影响,使得分类精度得到进一步的提高。

3) 将算法 FCLD-WOCSVM 和 LD-WOCSVM 运用于涡轴 发动机故障检测中,仿真实验表明,两种算法能够构建精度较 高的分类器,可以作为故障检测的候选算法。

参考文献 (References)

- GARG S. Controls and health management technologies for intelligent aerospace propulsion systems [C]// Proceedings of the 42nd AIAA Aerospace Sciences Meeting and Exhibit. Reston, VA: AIAA, 2004: No. 949.
- [2] 邓明,金业壮. 航空发动机故障诊断[M]. 北京:北京航空航天 大学出版社, 2012: 102-105. (DENG M, JIN Y Z. Aircraft Engine Fault Diagnosis [M]. Beijing: Beihang University Press,

2012:102-105.)

- [3] LU F, HUANG J, JI C, et al. Gas path on-line fault diagnostics using a nonlinear integrated model for gas turbine engines[J]. International Journal of Turbo and Jet-Engines, 2014, 31(3):261-275.
- [4] KOBAYASHI T, SIMON D L. Application of a bank of Kalman filters for aircraft engine fault diagnostics [R]. Washington, DC: National Aeronautics and Space Administration, 2003.
- [5] HE C, ZHANG X, JIA B. UIO based robust fault diagnosis approach for aero-engine fiber-optic sensor [C]// Proceedings of the 2013 IEEE International Conference on Automation Science and Engineering. Piscataway: IEEE, 2013: 550-553.
- [6] 殷锴,钟诗胜,那媛,等. 基于 BP 神经网络的航空发动机故障检测技术研究[J]. 航空发动机,2017,43(1):53-57.(YIN K,ZHONG S S, NA Y, et al. Research on aeroengine fault detection technology based on BP neural network[J]. Aeroengine, 2017,43 (1):53-57.)
- [7] QU H, HUANG Y. The research on aero-engine gas path fault diagnosis by genetic algorithm-BP neural network [J]. Hydraulics Engineering, 2015, 43(18):31-36.
- [8] 王小雷,林学森,杨欣毅. ELM 在航空发动机气路部件故障诊断的应用研究[J]. 系统仿真技术, 2016, 12(2):106-110. (WANG X L, LIN X S, YANG X Y. Research on ELM application in aeroengine gas path components fault diagnosis[J]. System Simulation Technology, 2016, 12(2):106-110.)
- [9] 廖洪一基于极限学习机的航空发动机故障诊断研究[D].广 汉:中国民用航空飞行学院, 2016:1-75. (LIAO H Y. Research on aero-engine fault diagnosis based on extreme learning machine [D]. Guanghan: Civil Aviation Flight University of China, 2016: 1-75.)
- [10] ZHAO Y, HUANG G, HU Q, et al. Soft extreme learning machine for fault detection of aircraft engine [J]. Aerospace Science and Technology, 2019, 91:70-81.
- [11] CORTES C, VAPNIK V. Support-vector networks [J]. Machine Learning, 1995, 20(3): 273-297.
- [12] 刘俊,李鹏飞.基于支持向量机的连续超声图像集分割算法
  [J]. 计算机应用, 2017, 37(7):2089-2094, 2113. (LIU J, LI P F. Continuous ultrasound image set segmentation method based on support vector machine[J]. Journal of Computer Applications, 2017, 37(7): 2089-2094, 2113.)
- [13] 陈云凤,云挺,周宇,等. 基于 PSO 优化 SVM 的纹理图像分割
  [J]. 计算机应用与软件, 2014, 31(4):214-218. (CHEN Y F, YUN T, ZHOU Y, et al. Texture image segmentation based on PSO optimizing SVM [J]. Computer Applications and Software, 2014, 31(4):214-218.)
- [14] 杨怀义.支持向量机的多目标图像分割算法仿真研究[J]. 计算机仿真, 2011, 28(12):223-226. (YANG H Y. Segment objects from natural images using SVM[J]. Computer Simulation, 2011, 28(12):223-226.)
- [15] 于真.基于支持向量机的人脸识别技术研究[J]. 计算机仿真,
   2011, 28(12): 296-299. (YU Z. Face recognition method based on support vector machine [J]. Computer Simulation, 2011, 28 (12): 296-299.)
- [16] 舒双宝,罗家融,徐从东,等. 一种基于支持向量机的人脸识别 新方法[J]. 计算机仿真, 2011, 28(2):280-283. (SHU S B, LUO J R, XU C D, et al. A new method of face recognition base on support vector machine [J]. Computer Simulation, 2011, 28 (2):280-283.)

- [17] 吴青,梁勃,王婉,等. 多分类光滑支持向量机的人脸识别方法
  [J]. 计算机应用, 2015, 35(S1):122-126. (WU Q, LIANG B, WANG W, et al. Face recognition method based on multi-class classification of smooth support vector machine [J]. Journal of Computer Applications, 2015, 35(S1):122-126.)
- [18] 李妍,刘茂福,姬东鸿.基于支持向量机的中文文本蕴涵识别研究[J]. 计算机应用与软件, 2014, 31(4):51-55. (LIY, LIUMF, JIDH. On SVM-based Chinese textual entailment recognition [J]. Computer Applications and Software, 2014, 31(4): 51-55.)
- [19] 郭超磊,陈军华.基于SA-SVM的中文文本分类研究[J].计算 机应用与软件,2019,36(3):277-281.(GUO C L, CHEN J H. Chinese text categorization based on SA-SVM[J]. Computer Applications and Software, 2019, 36(3):277-281.)
- [20] HENG H, ZHANG J, XIN C. Research on aircraft engine fault detection based on support vector machines [C]// Proceedings of the 2nd International Conference on Consumer Electronics, Communications and Networks. Piscataway: IEEE, 2012: 496-499.
- [21] 庄夏,戴敏,何元清.基于改进粒子群优化LSSVM的飞机发动 机故障诊断[J].实验技术与管理,2013,30(2):54-57. (ZHUANG X, DAI M, HE Y Q. Fault diagnosis for aero-engine based on improved particle swarm algorithm optimizing support vector machine [J]. Experimental Technology and Management, 2013,30(2):54-57.)
- [22] XI P, ZHAO Y, WANG P, et al. Least squares support vector machine for class imbalance learning and their applications to fault detection of aircraft engine [J]. Aerospace Science and Technology, 2019, 84:56-74.
- [23] MAHADEVAN S, SHAH S L. Fault detection and diagnosis in process data using one-class support vector machines [J]. Journal of Process Control, 2009, 19(10):1627-1639.
- [24] YIN S, ZHU X, JING C. Fault detection based on a robust one class support vector machine [J]. Neurocomputing, 2014, 145: 263-268.
- [25] 李琳,尚文利,姚俊,等.单类支持向量机在工业控制系统入侵 检测中的应用研究综述[J]. 计算机应用研究, 2016, 33(1):7-11. (LI L, SHANG W L, YAO J, et al. Overview of one-class support vector machine in intrusion detection of industrial control system [J]. Application Research of Computers, 2016, 33 (1): 7-11.)
- [26] 刘万军,秦济韬,曲海成.基于改进单类支持向量机的工业控制网络入侵检测方法[J].计算机应用,2018,38(5):1360-1365,1371. (LIU W J, QIN J T, QU H C. Intrusion detection algorithm of industrial control network based on improved one-class

support vector machine [J]. Journal of Computer Applications, 2018, 38(5): 1360-1365, 1371.)

- [27] YANG J, DENG T, SUI R. An adaptive weighted one-class SVM for robust outlier detection [C]// Proceedings of the 2015 Chinese Intelligent Systems Conference. Berlin: Springer, 2015: 475-484.
- [28] SCHÖLKOPF B, WILLIAMSON R, SMOLA A, et al. Support vector method for novelty detection [C]// Proceedings of the 12th International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 1999:582-588.
- [29] 常剑荻. 局部加权单类支持向量机研究[D]. 保定:河北大学, 2017: 1-44. (CHANG J D. Research on locality weighting oneclass support vector machines [D]. Baodin: Hebei University, 2017:1-44.)
- [30] RODRIGUEZ A, LAIO A. Clustering by fast search and find of density peaks [J]. Science, 2014, 344(6191):1492-1496.
- [31] ZHU F, YANG J, GAO C, et al. A weighted one-class support vector machine[J]. Neurocomputing, 2016, 189:1-10.
- [32] HOU T, LIU Y, WANG K, et al. A new weighted SVDD algorithm for outlier detection [C]// Proceedings of the 2016 Chinese Control and Decision Conference. Piscataway: IEEE, 2016:5456-5461.
- [33] CHEN Z, XU K, WEI J, et al. Voltage fault detection for lithiumion battery pack using local outlier factor [J]. Measurement, 2019, 146:544-556.
- [34] LIUT F TUNG K M, ZHOU Z. Isolation forest [C]// Proceedings of the 8th IEEE International Conference on Data Mining. Piscat-
- [33] 李本威,胡国才,高国胜,等. 涡轴发动机监视参数选择与诊断 方法研究[J]. 航空动力学报,2002,17(2):160-164.(LIBW, HUGC,GAOGS, et al. Study of selecting parameters and fault diagnosis for turboshaft[J]. Journal of Aerospace Power, 2002, 17 (2):160-164.)

This work is partially supported by the Fundamental Research Funds for the Central Universities (NS2017013).

HUANG Gong, born in 1994, M. S. candidate. His research interests include aero-engine modeling and control, aero-engine fault detection, machine learning algorithm.

ZHAO Yongping, born in 1982, Ph. D., research fellow. His research interests include aero-engine modeling and control, artificial intelligence.

XIE Yunlong, born in 1996, M. S. candidate. His research interests include aero-engine modeling and control, machine learning algorithm.

第40卷