Research Article

Fault Detection for Turbine Engine Disk Based on One-Class Large Vector-Angular Region and Margin

Jiusheng Chen ¹, Ye Wei,¹ Haibing Wang,² and Xiaoyu Zhang ¹

¹College of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China

²China Southern Airlines Limited, Beijing 101138, China

Correspondence should be addressed to Xiaoyu Zhang; xy_zhang@cauc.edu.cn

Received 26 August 2019; Revised 16 December 2019; Accepted 9 January 2020; Published 1 February 2020

Academic Editor: Elena Zaitseva

Copyright © 2020 Jiusheng Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Fault detection is an important technique to detect divergence based on unknown abnormalities, which involves establishing a computational model exclusively originated from the key features of the normal samples. The multimodality of process data distribution of engine turbine disk is inevitably affected by incorporation of ambient disturbance; the mean and covariance would vary significantly, resulting in decayed detecting accuracy. By adopting a strategy to maximize vector-angular mean and minimize vector-angular variance simultaneously in the feature space, a one-class large vector-angular region and margin (one-class LARM) framework is systematically conducted for fault detection of turbine engine disk which will enhance the robustness of the dynamic multimode process monitoring. Simulation based on the single mode and multimode of turbine engine disk is thoroughly performed and compared that the results of which solidly validated the favorable efficiency of the proposed method.

1. Introduction

Fault detection is an active research topic which learns a model from normal data and identifies anomalous behavior in data. In real world, fault detection has widely practical applications, for example, machine fault detection [1–4], industrial process monitoring [5], and aviation safety [6]. Anomaly detection could be generally categorized into classification problem. In the classification-based anomaly detection methods, the detection model is learned from the labeled training data which can predict if the test samples are normal or abnormal. In real fault detection applications, the normal patterns can be easily possessed while abnormal ones are difficult to obtain. In order to effectively solve this problem, one-class classification methods have been proposed which can effectively detect the fault signal under the condition of the normal sample.

Recently, kernel-based methods have been widely adopted in the field of fault detection, such as kernel principal component analysis (KPCA) [7], one-class support vector machine (OCSVM) [8], and support vector data description (SVDD) [9]. The KPCA algorithm is introduced

by Schölkopf et al. [7], which is an extension of principal component analysis (PCA) using techniques of kernel methods. Using a kernel function, the original linear operations of PCA are performed in a reproduced kernel Hilbert space. Choi et al. [10] introduce the KPCA algorithm for fault detection and identification of nonlinear processes, in which the unified fault detection index is calculated for fault detection and a reconstruction-based identification index is developed for fault identification. In addition, the monitoring method is successfully applied to simulated continuous stirred tank reactor (CSTR). The fault detection and identification results confirm that the KPCA-based process monitoring can successfully be applied to a CSTR process under four different fault scenarios. The computational complexity and spatial complexity are $O(n^3)$ and $O(n^2)$, respectively. In practice, the dimension of a kernel matrix usually increases with the size of dataset and consequently leads to higher computation complexity and storage, which becomes the disaster of using the KPCA-based fault detection method. Günter et al. [11] introduce a gain adaptation method to increase the computation efficiency of KPCA algorithm, in which the kernel algorithm is

accelerated by improving the convergence of the Kernel Hebbian Algorithm (KHA) for iterative kernel PCA. The conventional KPCA is intrinsically sensitive to abnormal data because the objective function is expressed by L2 norm in the training procedure. Xiao et al. [12] introduce the L1 norm into conventional KPCA, which is applied to fault detection. The experimental results confirmed that the L1 normal-based KPCA shows good robustness to outliers.

OCSVM is proposed by Schölkopf et al. [8] for tackling fault detection problem, which is an unsupervised learning approach. In order to enhance generalization ability, OCSVM calculates a hyperplane correlated to a kernel function constructed by the maximum-margin algorithm, in which the optimization hyperplane is adopted to separate the normal and abnormal samples in the feature space. It is of high complexity to solve OCSVM-related problems by conventional methods using large samples. Therefore, a series of training algorithms are proposed to reduce the computation time for resolving the benchmark of OCSVM algorithm. In references [13, 14], reduced support vector machine (RSVM) is proposed to tackle with large datasets using SVM coupled with nonlinear kernels. In RSVM, the procedure of training adopted active training set, which is a subset of the complete training set. Similarly, ν -Anomica for anomaly detection is proposed by Das et al. [15]. The accuracy of ν -Anomica is comparable to that of OCSVM while the training and test time have improved ranging from 5 to 20 times.

The one-class SVDD algorithm is introduced by Tax and Duin [9], which is a single classification method based on minimum bounding sphere theory, can handle the fault signal classification problem only under the condition of the normal samples. The optimization procedure of one-class SVDD established a minimal volume hypersphere by the normal data in the feature space to realize fault detection. Anomalous behavior in the datasets can be utilized by the minimal hypersphere of one-class SVDD. In order to improve generalization ability, Wu and Ye [16] propose a small sphere and large margin (SSLM) algorithm for anomaly detection with few outlier training data. The SSLM is aimed at constructing a minimal volume hypersphere enclosing most of the normal training examples in the feature space; meanwhile, the margin from any outlier to this hypersphere is as large as possible. In real application, fault detection problems are always required to process continuous massive data streams. Online one-class SVDD [17] and incremental one-class SVDD [18] are proposed for data stream anomaly detection.

In this paper, we propose a novel algorithm for learning a one-class large vector-angular region and margin (one-class LARM) for fault detection of turbine engine disk. The major improvement of presented work lies in the following two aspects. Firstly, one-class LARM is a flexible extension of two-class LARM algorithm. The one-class LARM is an unsupervised classification method which means only normal training samples are needed to build the detection model. This is of great significance because plenty of normal samples can be acquired easily, while the fault samples are difficult to obtain or even not available. Therefore, the one-class LARM can be better applied into practice. Secondly, the

multimodality of process data distribution somehow inevitably encounters disturbance that the data's mean and covariance of multimodality will change significantly, leading to the reduction of estimation accuracy. The proposed method is adjusted to maximize the vector-angular mean and minimize the vector-angular variance simultaneously, which can achieve a better generalization performance in the process working with multimode.

The article is organized as follows. Section 1 introduces the kernel-based fault detection technique. Section 2 summarizes a review of the two-class LARM approach. Section 3 discusses the proposed one-class LARM based fault detection method. Section 4 elucidated the detailed experimental design and compares the results on dataset. Section 5 discusses the detection performance of different comparative methods. Section 6 concludes the whole work and gives future prospects.

2. Two-Class LARM

We proposed two-class LARM [19] approach and applied successfully the anomalous detection problem based on imbalanced data. The two-class LARM algorithm is described as follows.

It is assumed that $S = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)\}$ is the training set with n examples, $\mathbf{x}_i \in \mathcal{R}^d$, $1 \leq i \leq n$ is the input data, and $y_i \in \{-1, +1\}$, $1 \leq i \leq n$ is the class label. $\phi(\cdot)$ is a feature mapping function which can map input data examples from a given input space $\mathcal{X} \subset \mathcal{R}^d$ to a high-dimensional feature space $\mathcal{F}$. The feature mapping function can be calculated by a given kernel function $\kappa(\cdot, \cdot)$, i.e., $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$, $\forall \mathbf{x}_i, \mathbf{x}_j \in \mathcal{X}$, $\forall \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \in \mathcal{F}$. The vector-angular [20] function expresses the length of the perpendicular projection of the training pattern $\phi(\mathbf{x}_i) \in \mathcal{F}$ onto a vector $\mathbf{v} \in \mathcal{F}$, which can be calculated via $\langle \mathbf{v}, \phi(\mathbf{x}_i) \rangle = \phi(\mathbf{x}_i)^T \mathbf{v}$.

The primal two-class LARM can be formulated as follows:

$$\begin{aligned}
 \min_{\omega, \rho, \mathbf{v}, \boldsymbol{\xi}} \quad & -\omega^2 - \rho^2 + \frac{2\lambda}{n^2} (\mathbf{n}\mathbf{v}^T \mathbf{X} \mathbf{X}^T \mathbf{v} - \mathbf{v}^T \mathbf{X} \mathbf{y} \mathbf{y}^T \mathbf{X}^T \mathbf{v}) \\
 & - \frac{\lambda}{n} (\mathbf{X} \mathbf{y})^T \mathbf{v} + \frac{1}{\nu_1 m_1} \sum_{i=1}^{m_1} \xi_i + \frac{1}{\nu_2 m_2} \sum_{j=m_1+1}^n \xi_j, \\
 \text{s.t.} \quad & \mathbf{v}^T \phi(\mathbf{x}_i) \geq \omega^2 - \xi_i, \quad i = 1, \dots, m_1 \\
 & \mathbf{v}^T \phi(\mathbf{x}_j) \leq \omega^2 - \rho^2 + \xi_j, \quad j = m_1 + 1, \dots, n \\
 & \xi_i \geq 0, i = 1, \dots, n,
 \end{aligned} \tag{1}$$

where $\mathbf{X} = [\phi(\mathbf{x}_1), \phi(\mathbf{x}_2), \dots, \phi(\mathbf{x}_n)]$, $\mathbf{y} = [y_1, y_2, \dots, y_n]^T$, m_1 is the number of normal training patterns, $m_2 = n - m_1$ is the number of abnormal training patterns, ω^2 ($\omega > 0$) is the width of vector-angular region, ρ^2 ($\rho > 0$) is the vector-angular margin, $\boldsymbol{\xi} = [\xi_1, \dots, \xi_n]^T$ is the vector of slack variables, and ν , ν_1 , ν_2 , and λ are four positive constants.

For training the two-class LARM based on the fault detection model, both normal samples and abnormal samples are required. However, in the fault detection of turbine engine disk, we usually can obtain abundant normal examples under normal working conditions, while the abnormal data are difficult or expensive to obtain. Unsupervised technique is a fault detection method that only requires normal samples, which aims to build a description model of normal samples. Therefore, one-class LARM is proposed to construct the fault detection model, which is an unsupervised learning method and does not require abnormal samples to detect the fault of turbine engine disk.

3. One-Class LARM-Based Fault Detection for Turbine Engine Disk

One-class LARM utilizes the strategy to maximize the vector-angular mean and minimize vector-angular variance simultaneously. According to the definition in ref. [21], the vector-angular mean $\bar{\gamma}$ and variance $\hat{\gamma}$ between training patterns $\phi(\mathbf{x}_i), i = 1, \dots, \ell$ and vector $\mathbf{v}$ can be expressed as

$$\bar{\gamma} = \frac{1}{\ell} \sum_{i=1}^{\ell} \mathbf{v}^T \phi(\mathbf{x}_i) = \frac{1}{\ell} (\mathbf{X}\mathbf{e})^T \mathbf{v}, \quad (2)$$

$$\begin{aligned} \hat{\gamma} &= \frac{1}{\ell^2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} (\mathbf{v}^T \phi(\mathbf{x}_i) - \mathbf{v}^T \phi(\mathbf{x}_j))^2 \\ &= \frac{2}{\ell^2} (\ell \mathbf{v}^T \mathbf{X} \mathbf{X}^T \mathbf{v} - \mathbf{v}^T \mathbf{X} \mathbf{e} \mathbf{e}^T \mathbf{X}^T \mathbf{v}), \end{aligned} \quad (3)$$

where $\mathbf{X} = [\phi(\mathbf{x}_1), \phi(\mathbf{x}_2), \dots, \phi(\mathbf{x}_\ell)]$, $\mathbf{e} = [1, \dots, 1]^T$, and ℓ is the number of training patterns. In this case, maximizing the vector-angular mean and minimizing the vector-angular variance can be calculated as follows:

$$\min_{\rho, \mathbf{v}, \xi} -\rho + \frac{2\lambda}{\ell^2} (\ell \mathbf{v}^T \mathbf{X} \mathbf{X}^T \mathbf{v} - \mathbf{v}^T \mathbf{X} \mathbf{e} \mathbf{e}^T \mathbf{X}^T \mathbf{v}) - \frac{\lambda}{\ell} (\mathbf{X}\mathbf{e})^T \mathbf{v} + \frac{1}{\nu \ell} \sum_{i=1}^{\ell} \xi_i,$$

$$\text{s.t.} \quad \mathbf{v}^T \phi(\mathbf{x}_i) \geq \rho - \xi_i, \quad i = 1, \dots, \ell$$

$$\xi_i \geq 0, \quad i = 1, \dots, \ell,$$

(4)

where $\mathbf{v}$ is the optimal vector, ρ is the vector-angular margin, $\xi = [\xi_1, \dots, \xi_\ell]^T$ is the vector of slack variables, and ν and λ are the two positive constants. Unlike two-class LARM, the one-class LARM is an unsupervised classification method, and only the normal training matrix $\mathbf{X} = [\phi(\mathbf{x}_1), \dots, \phi(\mathbf{x}_\ell)]$ is needed to build the fault detection model. For the turbine engine disk, the majority of data are obtained under normal working condition, while fault signals under abnormal working condition are difficult and expensive to obtain.

According to reference [21], the optimal vector $\mathbf{v}^*$ for the above problem listed in (4) could be rewritten as follows:

$$\mathbf{v}^* = \sum_{i=1}^{\ell} \alpha_i \phi(\mathbf{x}_i) = \mathbf{X}\mathbf{a}. \quad (5)$$

Hence, $\mathbf{X}^T \mathbf{v} = \mathbf{X}^T \mathbf{X} \mathbf{a} = \mathbf{K} \mathbf{a}$ can be obtained, where $\mathbf{K} = \mathbf{X}^T \mathbf{X}$ is the kernel matrix. Problem (4) can be rewritten as follows:

$$\begin{aligned} \min_{\rho, \mathbf{v}, \xi} \quad & -\rho + \frac{1}{2} \mathbf{a}^T \mathbf{Q} \mathbf{a} + \mathbf{q}^T \mathbf{a} + \frac{1}{\nu \ell} \sum_{i=1}^{\ell} \xi_i, \\ \text{s.t.} \quad & \mathbf{a}^T \mathbf{K}_{:,i} \geq \rho - \xi_i, \quad i = 1, \dots, \ell \\ & \xi_i \geq 0, \quad i = 1, \dots, \ell, \end{aligned} \quad (6)$$

where $\mathbf{Q} = 4\lambda(\ell \mathbf{K}^T \mathbf{K} - (\mathbf{K}\mathbf{e})(\mathbf{K}\mathbf{e})^T)/\ell^2$, $\mathbf{q} = -\lambda(\mathbf{K}\mathbf{e})/\ell$, and $\mathbf{K}_{:,i}$ is the i -th column of $\mathbf{K}$.

To investigate the solving method with constraints described in (6), the Lagrangian function is constructed as follows:

$$\begin{aligned} L(\rho, \mathbf{a}, \xi, \beta, \eta) &= -\rho + \frac{1}{2} \mathbf{a}^T \mathbf{Q} \mathbf{a} + \mathbf{q}^T \mathbf{a} + \frac{1}{\nu \ell} \sum_{i=1}^{\ell} \xi_i - \sum_{i=1}^{\ell} \beta_i (\mathbf{a}^T \mathbf{K}_{:,i} - \rho + \xi_i) \\ &\quad - \sum_{i=1}^{\ell} \eta_i \xi_i, \end{aligned} \quad (7)$$

where $\beta = [\beta_1, \dots, \beta_\ell]^T$ and $\eta = [\eta_1, \dots, \eta_\ell]^T$ are Lagrange multipliers. The following equations can be deduced and obtained by substituting the partial derivatives of $L(\rho, \mathbf{a}, \xi, \beta, \eta)$ with respect to the primal variables to zero:

$$\frac{\partial L}{\partial \rho} = 0 \implies -1 + \sum_{i=1}^{\ell} \beta_i = 0,$$

$$\frac{\partial L}{\partial \xi_i} = 0 \implies \frac{1}{\nu \ell} - \beta_i - \eta_i = 0, \quad i = 1, \dots, \ell, \quad (8)$$

$$\frac{\partial L}{\partial \mathbf{a}} = 0 \implies \mathbf{Q} \mathbf{a} + \mathbf{q} - \sum_{i=1}^{\ell} \beta_i \mathbf{K}_{:,i} = 0.$$

Substituting (8) into (7), the dual form can be obtained as follows:

$$\begin{aligned} \max_{\beta} \quad & -\frac{1}{2} \beta^T \mathbf{H} \beta + \mathbf{p}^T \beta, \\ \text{s.t.} \quad & 0 \leq \beta_i \leq \frac{1}{\nu \ell}, \quad i = 1, \dots, \ell, \\ & \sum_{i=1}^{\ell} \beta_i = 1, \end{aligned} \quad (9)$$

where $\mathbf{H} = \mathbf{K} \mathbf{Q}^{-1} \mathbf{K}$, $\mathbf{p} = -\lambda(\mathbf{H}\mathbf{e})/\ell$, and $\mathbf{Q}^{-1}$ is the inverse matrix of $\mathbf{Q}$. By solving the dual problem (9), the decision function is

$$f(\mathbf{z}) = \text{sgn}(\mathbf{v}^T \phi(\mathbf{z}) - \rho) = \text{sgn}\left(\sum_{i=1}^{\ell} \alpha_i^* \kappa(\mathbf{x}_i, \mathbf{z}) - \rho\right), \quad (10)$$

where $\alpha^* = \mathbf{Q}^{-1}(\sum_{i=1}^{\ell} \beta_i^* \mathbf{K}_{:,i} - \mathbf{q}) = \mathbf{Q}^{-1}(\mathbf{K}\beta^* - \mathbf{q})$, β^* is the optimal vector of the dual problem (9), $\rho = (1/m)\sum_{\mathbf{x}_i \in S} \mathbf{v}^T \phi(\mathbf{x}_i) = (1/m)\sum_{\mathbf{x}_i \in S} \sum_{k=1}^{\ell} \alpha_k^* \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_k) = (1/m)\sum_{\mathbf{x}_i \in S} \sum_{k=1}^{\ell} \alpha_k^* \kappa(\mathbf{x}_i, \mathbf{x}_k)$, S is the set of support vectors, and m denotes the number of support vectors.

4. Experimental Results

The performance of the proposed fault detection algorithm is evaluated on Disk Defect Data [22, 23], which were collected by the National Aeronautics and Space Administration (NASA) Glenn Research Center's Rotor Dynamics Laboratory. In the experimental testing system, the notch was set on the rim region of the disk. Three different spin speeds (3000 revolutions per minute (RPM), 4000 RPM, and 5000 RPM) and three states (normal, notch, and large notch) were recorded in the Disk Defect Data.

All the testing simulations were conducted on a computer with a Xeon(R) E5-1630 v4 at 3.70 GHz CPU and a 16 GB main memory. The accuracy and recall rates are used to evaluate the performance of the results. The accuracy and recall rates are expressed as follows:

$$\begin{aligned} \text{ACC} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}, \\ \text{REC} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \end{aligned} \quad (11)$$

where TP is the number of true positive samples, TN is the number of true negative samples, FP is the number of false positive samples, and FN is the number of false negative samples.

4.1. Parameter Selection and Impact Analysis. For all the experiments, the Radial Basis Function (RBF) is selected as the kernel function:

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right), \quad 0 < \gamma < +\infty, \quad (12)$$

where γ is the kernel parameter of the RBF. The parameter γ of the RBF kernel is selected in the set of $\{\alpha_0/32, \alpha_0/16, \alpha_0/8, \alpha_0/4, \alpha_0/2, \alpha_0\}$ by 5-fold cross validation, where α_0 is the average norm of the training instances. For one-class LARM, ν and λ are two regularization parameters which would affect the detection performance. Therefore, the following experiments were conducted to understand how the regularization parameters ν and λ influence the performance.

The first simulation is used to measure the effects of ν on the detection performance, in which λ is selected in the set of $\{0.01k, 0.1k\}$, $k = 1, 3, 5, 7, 9$ by 5-fold cross validation. In this experiment, the training and testing patterns are randomly selected from Disk Defect Data on each RPM, the

number of training patterns is set to 45, 60, 75, 90, and 105 (15, 20, 25, 30, and 35 normal samples independently on each RPM), and the number testing patterns is set to 6000 (1000 normal samples and 1000 abnormal samples independently on each RPM). All the experiments are repeated 10 times, and the average accuracy and recall rates are displayed in Figure 1.

The second simulation is used to measure the influences of the parameter λ on the detection performance, in which the parameter ν is selected in the set of $\{0.1k\}$, $k = 1, 2, \dots, 9$ by 5-fold cross validation. The settings of training and testing samples are the same as previous experiment. All the experiments were repeated 10 times. The average values are displayed in Figure 2.

From Figures 1 and 2, when the number of training patterns is 60, 75, and 90, the values of ν and λ have a faint impact on accuracy and recall rates. However, the number of training patterns is 45 and 105; both the accuracy rates and recall rates show more fluctuation according to ν or λ . That is because a small number of training samples may cause underfitting result and many training samples may cause overfitting result. In this paper, the hyperparameters ν and λ of one-class LARM are selected by cross-validation technique. The hyperparameter ν is selected in the set of $\{0.1, 0.3, 0.6, 0.9\}$, while the hyperparameter λ is selected in the set of $\{0.01, 0.05, 0.1, 0.5, 0.9\}$ by 5-fold cross validation. The selected range of regularized ν and λ has demonstrated to be suitable for further simulations.

4.2. Performance Evaluation of One-Class LARM. In this section, the one-class LARM is compared with OCSVM and one-class SVDD for notch detection in Disk Defect Data. We denote one-class LARM implemented by the LIBSVM package [24, 25]. For comparison, one-class SVDD and OCSVM also used LIBSVM package in the simulation.

ν of OCSVM is selected in the set of $\{0.01k, 0.1k\}$, $k = 1, 2, \dots, 9$. C in one-class SVDD is selected in the set of $\{1/\ell + k\delta\}$, $k = 0, 1, 2, \dots, 17$, where $\delta = (\ell - 1)/17\ell$ and ℓ is the number of training samples.

4.2.1. Case 1: Notch Detection at Single Mode. In case 1, the fault of turbine engine disk is detected at single mode process in which the training samples are composed of the normal samples with a rotating speed of 3000 RPM. The process is run normally for 3000 samples followed by switching to a notch model for 1000 samples with a rotating speed of 3000 RPM, 4000 RPM, and 5000 RPM, respectively. The experiment is repeated 20 times in which all the data points of the training and testing patterns are randomly selected from the Disk Defect Data. Detection results calculated by three algorithms are statistically depicted in Figure 3.

From Figure 3(a), we can see that one-class LARM has obvious advantages in the accuracy rate when the number of training samples is larger than 30. The recall rate of one-class LARM (see Figure 3(b)) is obviously better than of OCSVM and one-class SVDD. However, the training speed of one-class LARM, as shown in Figure 3(c), is slower compared to OCSVM and one-class SVDD. This is attributed to solving

FIGURE 1: Influence of ν on (a) the accuracy rate and (b) the recall rate.

the quadratic programming (QP) problem of one-class LARM which is required to calculate the inverse matrix of $\mathbf{Q}$. With the increase in the amount of training samples, calculation of the matrix inversion becomes considerably time-consuming. The results in Figure 3(d) indicate that there is minor difference between the testing times of OCSVM, one-class SVDD, and one-class LARM. When the number of training and testing samples is 105 and 6000, the average testing time of one-class LARM is approximately 0.0365 seconds.

4.2.2. Case 2: Notch Detection Process at Multimode. The turbine engine disk operates in a multimode process with

three rotating speeds, 3000 RPM, 4000 RPM, and 5000 RPM, respectively. The samples are selected in 6 phrases of selection strategy as follows: in phrase 1, 1000 samples at speed of 3000 RPM; in phrase 2, 1000 samples at 4000 RPM; in phrase 3, another 1000 samples at 5000 RPM. After normal working modes, the notch model occurs on the turbine engine disk and is switched to phrase 4. In this phrase, the process is working at the speed of 5000 RPM, 1000 samples are picked, and 1000 samples are recorded at 4000 RPM of phrase 5, followed by 1000 samples at 3000 RPM. The training samples consist of the same number of normal samples at 3000 RPM, 4000 RPM, and 5000 RPM. The settings of training and testing samples are the same as those in case 1. All the results are listed in Figure 4.

FIGURE 2: Influence of λ (a) on the accuracy rate and (b) the recall rate.

The results in Figure 4 show that one-class LARM has obvious advantages in accuracy rate and recall rate. It is notable to find that the recall rate nearly remains 98%, which is relatively higher than the OCSVM and one-class SVDD. Similar analysis could also be made as those in case 1; Figure 4(c) shows that the training time increases linearly of OCSVM and one-class SVDD while one-class LARM exhibits a nonlinear trend, particularly when the sample exceeds 75. At the same time, only minor difference is observed in the testing time in Figure 4(d). It shows that the one-class

LARM can detect testing examples faster, and specifically, 6,000 samples are detected in merely 0.03 seconds.

5. Discussion

In OCSVM, an optimal hyperplane expressed by the normal vector $\mathbf{w}$ and the offset ρ is constructed by maximizing the margin between the origin and the hyperplane in the feature space. The OCSVM can be solved by the following primal optimization problem [8]:

FIGURE 3: Comparing results of OCSVM, one-class SVDD, and one-class LARM at single mode: (a) comparison of the accuracy rate, (b) comparison of the recall rate, (c) comparison of the average training time, and (d) comparison of the average testing time.

$$\begin{aligned}
 \min_{\mathbf{w}, \rho, \xi} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{\nu \ell} \sum_{i=1}^{\ell} \xi_i - \rho, \\
 \text{s.t.} \quad & \mathbf{w}^T \phi(\mathbf{x}_i) \geq \rho - \xi_i, \quad i = 1, \dots, \ell \\
 & \xi_i \geq 0, \quad i = 1, \dots, \ell,
 \end{aligned} \tag{13}$$

where ξ is the nonzero slack variable and ν is a positive constant.

In one-class SVDD, a minimal enclosing ball by the radius R and the center $\mathbf{c}$ is constructed to enclose most of

the training samples in the feature space. The one-class SVDD can be solved by the following primal optimization problem [9]:

$$\begin{aligned}
 \min_{R, \mathbf{c}, \xi} \quad & R^2 + C \sum_{i=1}^{\ell} \xi_i, \\
 \text{s.t.} \quad & \|\phi(\mathbf{x}_i) - \mathbf{c}\|^2 \leq R + \xi_i, \quad i = 1, \dots, \ell \\
 & \xi_i \geq 0, \quad i = 1, \dots, \ell,
 \end{aligned} \tag{14}$$

where C is a positive constant and ξ is the nonzero slack variable.

FIGURE 4: Comparing results of OCSVM, one-class SVDD, and one-class LARM at multimode process: (a) comparison of the accuracy rate, (b) comparison of the recall rate, (c) comparison of the average training time, and (d) comparison of the average testing time.

Including one-class SVDD and OCSVM, several kernel-based methods have been proposed for solving the fault detection problem when the abnormal data are unavailable for model training. Like OCSVM, one-class LARM also aims at building the separating hyperplane by maximizing the margin between the hyperplane and the origin. However, one-class LARM utilizes the strategy to find an optimal vector $\mathbf{v}$ in the feature space just by using positive labeled training data set, which attempts to maximize the vector-angular mean and minimize the vector-angular variance simultaneously. Figure 5 illustrates the principle of one-class LARM.

From Figure 5, we can see that the one-class LARM finds an optimal vector $\mathbf{v}$ in the feature space, and the training

samples project onto the optimal vector $\mathbf{v}$ joining their large mean and small variance. The projection onto the optimal vector $\mathbf{v}$ can achieve the effect that the distribution between the origin and the vector-angular mean could be separated as much as possible, and the projection of samples within classes could be as dense as possible. Therefore, the one-class LARM-based fault detection model can achieve a favorable generalization performance at single and multimode processes of turbine engine disk.

To make it easier to compare the detection performance of the three methods, all involved parameters at different modes are statistically calculated. The results of all working modes are combined and drawn in Figure 6.

FIGURE 5: Geometric interpretation of optimal hyperplane of one-class LARM.

FIGURE 6: Continued.

FIGURE 6: Comparing results of OCSVM, one-class SVDD, and one-class LARM at different mode processes: (a) comparison of the average accuracy rate; (b) comparison of the average recall rate; (c) comparison of the average training time; and (d) comparison of the average testing time.

As shown in Figures 6(a) and 6(b), for all examined process modes, the detection accuracy and recall rates indicate that one-class LARM compares favorably to both one-class SVDD and OCSVM. The one-class LARM achieves the best detection performance among all the methods, especially for the turbine engine disk working at a multimode process.

From Figure 6(c), we can see that the training time of OCSVM and one-class SVDD performs better compared to one-class LARM. That is also because QP problem solvers of one-class LARM are required to calculate the matrix inversion. With the increase of training samples, solving the matrix inversion becomes extremely time-consuming. It is the same reason that the training time of one-class LARM is a major issue which will further restrict the algorithm's applicability for medium- and large-sized datasets. When the number of training sample is relatively small, the training time of three methods exhibits different. But the one-class LARM has obvious advantages in detection performance in terms of accuracy and recall rates.

It is observed from Figure 6(d) that the testing time is slightly different. The one-class LARM can detect test examples faster, using 0.0365 s to detect 6,000 samples in single mode process and to detect 6,000 samples in multimode process within 0.03 s.

6. Conclusions

In the presented work, one-class LARM algorithm is proposed for novelty detection of turbine engine disk, of which the detection model could be built merely by normal training samples. Considering the circumstances that fault samples are difficult to obtain or available, the proposed method can be better applied to solve practical issues. In addition, the method searches the optimization vector in the feature space

via maximizing the vector-angular mean and minimizing the variance simultaneously, which can achieve better detection performance particularly in the multimode process. Comprehensive comparisons of simulation results have demonstrated the effectiveness of proposed algorithm.

In order to take full advantage of the proposed method and examine the robustness, future work will be extending one-class LARM coupled with incremental learning to test on medium- and large-size datasets.

Data Availability

The Disk Defect data used to support the findings of this study have been deposited in the DASHlink repository (<https://c3.nasa.gov/dashlink/resources/314/>), which is collected by the National Aeronautics and Space Administration (NASA) Glenn Research Center's Rotor Dynamics Laboratory.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under grant nos. 61603395 and U1833108.

References

- [1] G. F. Bin, J. J. Gao, X. J. Li, and B. S. Dhillon, "Early fault diagnosis of rotating machinery based on wavelet packets-Empirical mode decomposition feature extraction and neural network," *Mechanical Systems and Signal Processing*, vol. 27, pp. 696–711, 2012.

- [2] L. Siliang, H. Qingbo, and W. Jun, "A review of stochastic resonance in rotating machine fault detection," *Mechanical Systems and Signal Processing*, vol. 116, pp. 230–260, 2019.
- [3] R. Guo and Q. Gan, "Prognostics for a leaking hydraulic actuator based on the F-distribution particle filter," *IEEE Access*, vol. 5, no. 5, pp. 22409–22420, 2017.
- [4] R. Guo and Na Zhang, "A process monitoring scheme for uneven duration batch process based on sequential moving principal component analysis," *IEEE Transactions on Control Systems Technology*, 2018.
- [5] S. Yin, S. X. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for industrial process monitoring," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 11, pp. 6418–6428, 2014.
- [6] E. Smart, D. Brown, and J. Denman, "A two-phase method of detecting abnormalities in aircraft flight data and ranking their impact on individual flights," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1253–1265, 2012.
- [7] B. Schölkopf, A. Smola, and K. R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 2008.
- [8] B. Schölkopf, R. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt, "Support vector method for novelty detection," in *Advances in Neural Information Processing Systems-NIPS 1999*, pp. 582–588, MIT Press, Cambridge, MA, USA, 1999.
- [9] D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Machine Learning*, vol. 54, no. 1, pp. 45–66, 2004.
- [10] S. W. Choi, C. Lee, J.-M. Lee, J. H. Park, and I.-B. Lee, "Fault detection and identification of nonlinear processes based on kernel PCA," *Chemometrics and Intelligent Laboratory Systems*, vol. 75, no. 1, pp. 55–67, 2005.
- [11] S. Günter, N. Schraudolph, and S. Vishwanathan, "Fast iterative kernel principal component analysis," *Journal of Machine Learning Research*, vol. 8, pp. 1893–1918, 2007.
- [12] Y. Xiao, H. Wang, W. Xu, and J. Zhou, "L1 norm based KPCA for novelty detection," *Pattern Recognition*, vol. 46, no. 1, pp. 389–396, 2013.
- [13] C. C. Chang and Y. J. Lee, "Generating the reduced set by systematic sampling," in *Intelligent Data Engineering and Automated Learning—IDEAL 2004*, pp. 720–725, Springer, Berlin, Germany, 2004.
- [14] K. M. Lin and C. J. Lin, "A study on reduced support vector machines," *IEEE Transactions on Neural Networks*, vol. 14, no. 6, pp. 1449–1459, 2003.
- [15] S. Das, K. Bhaduri, N. C. Oza et al., " ν -Anomica: a fast support vector based novelty detection technique," in *Proceedings of the 2009 Ninth IEEE International Conference on Data Mining*, pp. 101–109, Miami, FL, USA, December 2009.
- [16] M. Wu and J. Ye, "A small sphere and large margin approach for novelty detection using training data with outliers," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 31, no. 11, pp. 2088–2092, 2009.
- [17] Y. Zhang, N. Meratnia, and P. Havinga, "Adaptive and online one-class support vector machine-based outlier detection techniques for wireless sensor networks," in *Proceedings of the International Conference on Advanced Information Networking and Applications Workshops*, pp. 990–995, IEEE, Bradford, UK, May 2009.
- [18] B. Krawczyk and M. Woźniak, "Incremental weighted one-class classifier for mining stationary data streams," *Journal of Computational Science*, vol. 9, pp. 19–25, 2015.
- [19] J. Chen, K. Zhang, and K. Guo, "A core set based large vector-angular region and margin approach for novelty detection," *Mathematical Problems in Engineering*, vol. 2016, Article ID 1658758, 12 pages, 2016.
- [20] W. Hu, F. L. Chung, and S. Wang, "The maximum vector-angular margin classifier and its fast training on large datasets using a core vector machine," *Neural Networks*, vol. 27, no. 3, pp. 60–73, 2011.
- [21] T. Zhang and Z. H. Zhou, "Large margin distribution machine," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 313–322, New York, NY, USA, August 2014.
- [22] A. Abdul-Aziz, M. R. Woike, N. C. Oza et al., "Propulsion health monitoring of a turbine engine disk using spin test data," in *Proceedings of the SPIE Smart Structures and Materials & Nondestructive Evaluation and Health Monitoring*, San Diego, CA, USA, March 2010.
- [23] A. Abdul-Aziz, M. R. Woike, N. C. Oza, B. L. Matthews, and J. D. Lekki, "Rotor health monitoring combining spin tests and data-driven anomaly detection methods," *Structural Health Monitoring: An International Journal*, vol. 11, no. 1, pp. 3–12, 2011.
- [24] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, p. 27, 2011.
- [25] L. Wang, H. Froehlich, K. Rieck et al., "LIBSVM for SVDD and finding the smallest sphere containing all data," 2004, https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/#libsvm_for_svdd_and_finding_the_smallest_sphere_containing_all_data.