



ARTICLE

A Fast Small-Sample Modeling Method for Precision Inertial Systems Fault Prediction and Quantitative Anomaly Measurement

Hongqiao Wang^{1,*} and Yanning Cai²

¹Xi'an Research Institute of High-Tech, Xi'an, 710025, China

²Northwest University of Politics and Law, Xi'an, 710122, China

*Corresponding Author: Hongqiao Wang, Email: ep.hqwang@gmail.com

Received: 21 June 2021 Accepted: 29 July 2021

ABSTRACT

Inertial system platforms are a kind of important precision devices, which have the characteristics of difficult acquisition for state data and small sample scale. Focusing on the model optimization for data-driven fault state prediction and quantitative degree measurement, a fast small-sample supersphere one-class SVM modeling method using support vectors pre-selection is systematically studied in this paper. By theorem-proving the irrelevance between the model's learning result and the non-support vectors (NSVs), the distribution characters of the support vectors are analyzed. On this basis, a modeling method with selected samples having specific geometry character from the training sets is also proposed. The method can remarkably eliminate the NSVs and improve the algorithm's efficiency. The experimental results testify that the scale of training samples and the modeling time consumption both give a sharply decrease using the support vectors pre-selection method. The experimental results on inertial devices also show good fault prediction capability and effectiveness of quantitative anomaly measurement.

KEYWORDS

Fault prediction; anomaly measurement; precision inertial devices; support vector pre-selection

1 Introduction

As the emerging growth of demand for systems' health monitoring and reliability estimation, the fault diagnosis and prediction problems have become a research focus at present [1–4]. To realize the full-process health management of system, the data collection, real-time detection, state measurement, modeling, prediction and anomaly evaluation have become more and more important, especially for some precision devices and complex systems requiring high reliability [5,6]. Inertial system platforms are a kind of important precision devices used in many industrial fields, such as navigation, guidance and aerospace, which have the characteristics of difficult acquisition for state data and small sample scale, so the fault trend of the devices should be discovered as early as possible based on the limited data samples to prevent the occurrence of serious damages to the whole system. The fault state of system is actually an abnormal phenomena deviating from the normal state, so the early warning of fault may be realized by novelty recognition before the



fault completely occurs [7–10]. The concrete method based on this idea is as follows: build the system's normal working state space directly using the system's normal features, and recognize the novel samples by detecting whether the deviation from the normal features is existed.

For system fault diagnosis and prediction problems using data-driven methods [11–14], one-class SVM (OCSVM) is a typical classification tool, which can identify the known object class(normal sample) and the unknown object class(novel sample). As a popular method, a lot of improved OCSVM algorithms are studied. For example, FernaNdez-Francos et al. [15] presented a ν -SVM based OCSVM. Amer et al. [16] proposed an enhanced OCSVM for unsupervised method of anomaly detecting. Yin et al. [17] and Xiao et al. [18] studied the robust OCSVM algorithms. Yan et al. [19] introduced the OCSVM algorithm to fault prediction in the online condition. Huang et al. [20] combined the wavelet T-F entropy and OCSVM, and applied the method to mechanical system fault prediction. In addition, the one-class classification and the learning algorithms are suitable for data-driven modeling of black-box problems, especially for some pattern recognition applications [21–23]. In the past few years, the OCSVM methods are also used for faults detection, identification and diagnosis [24,25]. To detect the novelty sample effectively, Yi et al. [26] presented a supervised novelty detection SVM, which is extended from OCSVM and has a fast learning speed. Miao et al. [27] presented a distributed model and its online learning method with OCSVM for anomaly finding. Moreover, although most of the OCSVMs are applied for the classification fields, Bhand et al. [28] extended the model to the regression modeling and automated design process. As another data-driven modeling method, the extreme learning machine model is also introduced to the prediction application field [29]. With the widespread utilization of deep learning, the idea of OCSVM is also applied in some deep neural networks. For example, by means of the combination of OCSVM and deep neural networks [30], the deep OCSVM algorithm is proposed [31,32], the one-class convolutional neural network and its applications are also studied [33]. Moreover, by introducing the idea of the adversarial learning, some one-class classification and the anomaly detection methods are proposed based the adversarial framework [34,35]. Considering the small sample scale and the online monitoring conditions of precision inertial devices fault prediction, the training samples' scale for modeling is too small and the training speed is too fast for all of the deep neural network models. As a result, the deep learning methods are not the ideal solutions in this application situation.

The OCSVM model has good sparseness, namely, the specific form of the model is only decided by part of the samples in the training set, namely the support vectors (SVs). Even so, the modeling efficiency is still greatly decreased as the calculation of a large number of NSVs in the model optimization process. To enhance the modeling efficiency of the fault diagnosis and prediction applications, a synthetical method using SV pre-selection is proposed aiming at the precision inertial device fault prediction in the small sample condition, which realizes data-driven fault diagnosis, state prediction, anomaly detection and quantitative measurement. Considering an OCSVM model, the distribution characteristics of the support vectors in the high dimensional space from are analyzed, respectively. On this basis, a modeling method with the samples owning some geometry characters selected from the training set is also executed. The fault prediction efficiency can be greatly improved using the pre-selection of the non-support-vector samples in the condition of having no effect on the fault prediction capability of OCSVM.

In the remainder of this paper, the sections are organized as follows: in [Section 2](#), the OCSVM method in fault prediction field and the supersphere models for OCSVM are introduced. The support vector pre-selection method for the supersphere model is studied detailedly in [Section 3](#), the number calculation of pre-selected samples and the algorithm's complexity are also

analysed. In [Section 4](#), the fault prediction and quantitative anomaly measurement using SV pre-selection OCSVM is studied. Several experiments and result analysis are carried out in [Section 5](#). Finally, we conclude in [Section 6](#).

2 Supersphere Models for OCSVM

2.1 OCSVM Method in Fault Prediction Field

Suppose the system's state can be represented with a d -dimensional variable $\mathbf{x} = \{x_1, x_2, \dots, x_d\} \in \mathbb{R}^d$. Let the sample set collected from the normal working state is $\mathbf{D} = \{\mathbf{x}_i\}$, $i = 1, 2, \dots, l$. With the generation of new sample \mathbf{x}' , it is necessary to judge if the sample indicate the existence of the fault trend. The sample has the characteristic of non-output and unlabeled in the above mentioned problem. The key point of the problem lies in how to build the fault prediction model based on the sample set \mathbf{D} . In this fault prediction condition, the samples representing the system normal working state are seemed as samples in the target class, and the target class samples commonly have the similar distribution in the feature space. Correspondingly, the samples representing the system anomaly working state belong to the non-target class, which have a scattering distribution, so the one-class target information can be regard as the main information source. Unlike the general supervised classification with two values, this classification method can be seemed as the semi-supervised classification. Through the above analysis, it can be concluded that the OCSVM model can be used for fault prediction problems.

In the research field of SVM, the OCSVM is first introduced from Schölkopf in 1999, which is originally applied to probability density estimation of functions. The main thought of the method is: suppose a dataset in the given input space \mathbb{R}^n is given, and it satisfies a distribution with probability P . We hope to find a simple data sub-set S in the input space and approximate the distribution, subject to the probability, that the sample with the P distribution falling into the outside of S , satisfies a certain given probability ν ($0 < \nu < 1$). The OCSVM is an extension of the standard SVM, which has two realization approaches. We will introduce the main thought of the methods in the following subsections.

2.2 Supersphere Model of OCSVM

The form of training samples on the supersphere model of OCSVM is $\mathbf{D} = \{\mathbf{x}_i \in \mathbb{R}^d | i = 1, \dots, l\}$. To build the supersphere model, we firstly should map the samples to a new feature space with a nonlinear mapping function $\varphi(\cdot)$. Secondly, seeking for a minimum sphere with the center \mathbf{a} and the radius R in the new space, which can contain all the samples. For some sample points, some error may exist, we can introduce the ξ_i as the control variables. At the same time, the calculation of inner product in the new space could conveniently be substituted by the kernel functions satisfying the Mercer conditions. Namely, find a kernel function $k(\mathbf{x}_1, \mathbf{x}_2)$ and let $k(\mathbf{x}_1, \mathbf{x}_2) = \varphi(\mathbf{x}_1) \cdot \varphi(\mathbf{x}_2)$, as a result, the optimization problem is

$$\begin{aligned} \min R^2 + \frac{1}{\nu l} \sum_{i=1}^l \xi_i \\ \text{s.t. } [\varphi(\mathbf{x}_i) - \mathbf{a}][\varphi(\mathbf{x}_i) - \mathbf{a}]^T \leq R^2 + \xi_i \\ \xi_i \geq 0, i = 1, 2, \dots, l \end{aligned} \quad (1)$$

In (1), $\nu \in (0, 1)$ plays a compromise role between the radius of the supersphere and the training samples' number containing in the sphere. The slack factors can guarantee that more

training samples are containing in the supersphere as far as possible under the condition of the supersphere is maximally compressed. To calculate the center \mathbf{a} and the radius R , formula (1) can be transformed into its dual form (2), the deduce process is

$$\begin{aligned} \min \quad & \sum_{i,j=1}^l \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i=1}^l \alpha_i k(\mathbf{x}_i, \mathbf{x}_i) \\ \text{s.t.} \quad & \sum_{i=1}^l \alpha_i = 1 \\ & 0 \leq \alpha_i \leq \frac{1}{v \cdot l}, i = 1, 2, \dots, l \end{aligned} \quad (2)$$

Solve the optimization problem (2), it can be found that most of the α_i are 0, for the α_i unequal to 0, the corresponding samples are still called support vectors, and the form of the supersphere is only decided by support vectors. We may analysis the relative location relationship of the sample between the high-dimensional space and the supersphere based on the different values of α_i according to the following three situations:

- 1) If $\alpha_i \in (0, \frac{1}{v \cdot l})$, it means that \mathbf{x}_i satisfies $[\varphi(\mathbf{x}_i) - \mathbf{a}][\varphi(\mathbf{x}_i) - \mathbf{a}]^T = R^2$, namely \mathbf{x}_i is a boundary sample, and locates at the surface of the supersphere as a support vector.
- 2) If $\alpha_i = \frac{1}{v \cdot l}$, it means that \mathbf{x}_i satisfies $[\varphi(\mathbf{x}_i) - \mathbf{a}][\varphi(\mathbf{x}_i) - \mathbf{a}]^T > R^2$, namely \mathbf{x}_i is not a boundary sample, and locates outside of the supersphere as a non-support vector.
- 3) If $\alpha_i = 0$, it means that \mathbf{x}_i satisfies $[\varphi(\mathbf{x}_i) - \mathbf{a}][\varphi(\mathbf{x}_i) - \mathbf{a}]^T < R^2$, namely \mathbf{x}_i locates inside of the supersphere as a non-support vector.

Let I_{SV} is a set of support vectors, and l_{SV} is the support vectors' number, the center of the obtained supersphere is $\mathbf{a} = \sum_{i \in I_{SV}} \alpha_i \varphi(\mathbf{x}_i)$. According to the KKT condition, for the sample \mathbf{x}_i corresponding to $0 < \alpha_i < \frac{1}{v \cdot l}$, we can get

$$R^2 - [k(\mathbf{x}_i, \mathbf{x}_i) - 2 \sum_{j \in I_{SV}} \alpha_j k(\mathbf{x}_j, \mathbf{x}_i) + c^2] = 0 \quad (3)$$

As a result, the radius R can be obtained. For a new sample \mathbf{z} ,

$$\begin{aligned} f(\mathbf{z}) &= [\varphi(\mathbf{z}) - \mathbf{a}][\varphi(\mathbf{z}) - \mathbf{a}]^T - R^2 \\ &= k(\mathbf{z}, \mathbf{z}) - 2 \sum_{i \in I_{SV}} \alpha_i k(\mathbf{z}, \mathbf{x}_i) + \sum_{i \in I_{SV}} \sum_{j \in I_{SV}} \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) - R^2 \end{aligned} \quad (4)$$

If $f(\mathbf{z}) \leq 0$, we can conclude that \mathbf{z} is normal point, if not, \mathbf{z} is a novel point.

3 Sample Set Optimization for the Supersphere Model

The OCSVM model is sparse, namely the decision form of the supersphere is only depend on the support vectors, so the algorithm's complexity may be reduced through simplifying the training set, which is similar to the two-class model of SVM. It also can be found that the probability distribution of the SVs in the high-dimensional mapping space has some characteristics, which provides possibility for the SV pre-selection. Now, we'll discuss how to realize the support vectors pre-selection from the training sample set, namely the sample set optimization, and how to improve the training efficiency of OCSVM.

Definition 1: Suppose a sample set $\mathbf{D} = \{\mathbf{x}_i \in \mathbb{R}^d | i = 1, 2, \dots, l\}$ is known, let the mapping form to the high-dimension feature space is $\varphi(\cdot)$, the central point of the feature space is \mathbf{m} , so \mathbf{m} has a form as

$$\mathbf{m} = \frac{1}{l} \sum_{i=1}^l \varphi(\mathbf{x}_i) \quad (5)$$

Definition 2: Suppose two sample points \mathbf{x} and \mathbf{y} are known, the vectors from the original point to the two sample points are respectively $\vec{\mathbf{x}}$ and $\vec{\mathbf{y}}$. Let the included angle of the two vectors in the high-dimensional space is $\theta(\vec{\mathbf{x}}, \vec{\mathbf{y}})$, its form is

$$\theta(\vec{\mathbf{x}}, \vec{\mathbf{y}}) = \arccos \left(\frac{(\varphi(\mathbf{x}) \cdot \varphi(\mathbf{y}))}{\|\varphi(\mathbf{x})\| \|\varphi(\mathbf{y})\|} \right) \quad (6)$$

Definition 3: Suppose two sample points \mathbf{x}_1 and \mathbf{x}_2 are known, let the distance of these two points is $d(\mathbf{x}_1, \mathbf{x}_2)$, the form is

$$d(\mathbf{x}_1, \mathbf{x}_2) = \|\varphi(\mathbf{x}_1) - \varphi(\mathbf{x}_2)\| = \sqrt{k(\mathbf{x}_1, \mathbf{x}_1) - 2k(\mathbf{x}_1, \mathbf{x}_2) + k(\mathbf{x}_2, \mathbf{x}_2)} \quad (7)$$

Definition 4: If the sample points \mathbf{x}_1 and \mathbf{x}_2 are mapping to the feature space with $\varphi(\cdot)$, the distance of the two points in the feature space satisfies Eq. (8), which implies the Eq. (7).

$$\begin{aligned} d^2(\mathbf{x}_1, \mathbf{x}_2) &= \|\varphi(\mathbf{x}_1) - \varphi(\mathbf{x}_2)\|^2 = \varphi(\mathbf{x}_1) \cdot \varphi(\mathbf{x}_1) - 2\varphi(\mathbf{x}_1) \cdot \varphi(\mathbf{x}_2) + \varphi(\mathbf{x}_2) \cdot \varphi(\mathbf{x}_2) \\ &= k(\mathbf{x}_1, \mathbf{x}_1) - 2k(\mathbf{x}_1, \mathbf{x}_2) + k(\mathbf{x}_2, \mathbf{x}_2) \end{aligned} \quad (8)$$

3.1 Support Vector Pre-Selection Method for the Supersphere Model

Theorem 1: Suppose the training sample set is $\mathbf{D} = \{\mathbf{x}_i \in \mathbb{R}^d | i = 1, 2, \dots, l\}$, according to the Eq. (2), the supersphere model can be constructed, and the optimal solution of the supersphere is $\boldsymbol{\alpha} = \{\alpha_1, \dots, \alpha_l\}$. The samples with the non-zero Lagrange multipliers are selected from \mathbf{D} , with which we can obtain new sets $\tilde{\mathbf{D}} = \{\mathbf{x}_i | \alpha_i \neq 0, i = 1, 2, \dots, l\}$, and $\tilde{\boldsymbol{\alpha}} = \{\alpha_i \neq 0, i = 1, 2, \dots, l\}$. If the kernel function and model parameters are all the same, and suppose the optimal solution of the supersphere based on $\tilde{\mathbf{D}}$ is $\tilde{\boldsymbol{\beta}}$, it can be concluded that $\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\alpha}}$.

Proof: Let the numbers of element in $\tilde{\mathbf{D}}$ and $\tilde{\boldsymbol{\alpha}}$ are both l_{SV} , $I_{SV} = \{i | \alpha_i \neq 0, i = 1, \dots, l\}$, and the optimal solution is $\tilde{\boldsymbol{\beta}} = \{\tilde{\beta}_i | i = I_{SV}(1), \dots, I_{SV}(l_{SV})\}$.

Set the optimization problem based on \mathbf{D} is L_1

$$\begin{aligned} \min L_1(\boldsymbol{\alpha}) &= \sum_{i,j=1}^l \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i=1}^l \alpha_i k(\mathbf{x}_i, \mathbf{x}_i) \\ \text{s.t.} \quad &\sum_{i=1}^l \alpha_i = 1 \\ &0 \leq \alpha_i \leq \frac{1}{v \cdot l}, \quad i = 1, 2, \dots, l \end{aligned} \quad (9)$$

and the optimization problem based on $\tilde{\mathbf{D}}$ is L_2

$$\begin{aligned} \min L_2(\tilde{\boldsymbol{\beta}}) &= \sum_{i,j \in I_{SV}} \tilde{\beta}_i \tilde{\beta}_j k(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i \in I_{SV}} \tilde{\beta}_i k(\mathbf{x}_i, \mathbf{x}_i) \\ \text{s.t.} \quad &\sum_{i \in I_{SV}} \tilde{\beta}_i = 1 \\ &0 \leq \tilde{\beta}_i \leq \frac{1}{v \cdot l}, \quad i = I_{SV}(1), \dots, I_{SV}(l_{SV}) \end{aligned} \quad (10)$$

If $\tilde{\boldsymbol{\beta}}$ is extended to $\boldsymbol{\beta}$ with the relationship $\beta_i = \begin{cases} \tilde{\beta}_i & i \in I_{SV} \\ 0 & i \notin I_{SV} \end{cases}$, $i = 1, \dots, l$, it can be obviously

find that $\sum_{i \in I_{SV}} \tilde{\beta}_i = 1$ and $0 \leq \tilde{\beta}_i \leq \frac{1}{v \cdot l}$, so $\boldsymbol{\beta}$ is the feasible solution of L_1 . As $\boldsymbol{\alpha}$ is the optimal solution of L_1 , it can be concluded that $L_1(\boldsymbol{\alpha}) \leq L_1(\boldsymbol{\beta})$, because

$$\begin{aligned} L_1(\boldsymbol{\beta}) &= \sum_{i,j=1}^l \beta_i \beta_j k(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i=1}^l \beta_i k(\mathbf{x}_i, \mathbf{x}_i) \\ &= \sum_{i,j \in I_{SV}} \tilde{\beta}_i \tilde{\beta}_j k(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i \in I_{SV}} \tilde{\beta}_i k(\mathbf{x}_i, \mathbf{x}_i) = L_2(\tilde{\boldsymbol{\beta}}) \end{aligned} \quad (11)$$

$$\begin{aligned} L_1(\boldsymbol{\alpha}) &= \sum_{i,j=1}^l \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i=1}^l \alpha_i k(\mathbf{x}_i, \mathbf{x}_i) \\ &= \sum_{i,j \in I_{SV}} \tilde{\alpha}_i \tilde{\alpha}_j k(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i \in I_{SV}} \tilde{\alpha}_i k(\mathbf{x}_i, \mathbf{x}_i) = L_2(\tilde{\boldsymbol{\alpha}}) \end{aligned} \quad (12)$$

In summary, $L_2(\tilde{\boldsymbol{\alpha}}) \leq L_2(\tilde{\boldsymbol{\beta}})$. It is obvious that $\tilde{\boldsymbol{\alpha}}$ is the feasible solution of L_2 , and $\tilde{\boldsymbol{\beta}}$ is the optimal solution of L_2 , then we have $L_2(\tilde{\boldsymbol{\alpha}}) \geq L_2(\tilde{\boldsymbol{\beta}})$, as a result, $L_2(\tilde{\boldsymbol{\alpha}}) = L_2(\tilde{\boldsymbol{\beta}})$. Considering the uniqueness of the optimal solution of L_2 , it can be concluded that $\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\alpha}}$.

From **Theorem 1**, it can be concluded that the trained supersphere model won't be damaged if the NSVs are removed. The support vector pre-selection principle of the supersphere model is shown as [Fig. 1](#). In the figure, there are five sample points $\mathbf{x}_1 \sim \mathbf{x}_5$. Based on the construction principle of supersphere, we can know that $\mathbf{x}_1 \sim \mathbf{x}_3$ are non-support-vectors, \mathbf{x}_4 and \mathbf{x}_5 are support vectors. Let the central point of the sample set is \mathbf{m} , we can easily find that the sample points \mathbf{x}_4 and \mathbf{x}_5 , which are far away from \mathbf{m} , are more likely becoming the support vectors.

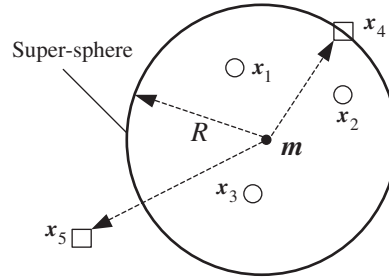


Figure 1: Support vector pre-selection principle of the supersphere model

Through the above analysis, we can conclude the supersphere model’s support vector pre-selection algorithm as the following 3 steps:

Step 1: For a given sample set $D = \{x_i \in \mathbb{R}^d | i = 1, 2, \dots, l\}$, calculate the sphere center

$$m = \frac{1}{l} \sum_{i=1}^l \varphi(x_i) \text{ based on the Definition 1.}$$

Step 2: Considering Definition 3, the square of distance from the sample x to m in the high-dimensional space is

$$\begin{aligned} d^2(x, m) &= \left\| \varphi(x) - \frac{1}{l} \sum_{i=1}^l \varphi(x_i) \right\|^2 \\ &= k(x, x) - \frac{2}{l} \sum_{j=1}^l k(x, x_j) + \frac{1}{l^2} \sum_{i=1}^l \sum_{j=1}^l k(x_i, x_j) \end{aligned} \quad (13)$$

As $\frac{1}{l^2} \sum_{i=1}^l \sum_{j=1}^l k(x_i, x_j)$ is a constant, to reduce the calculation amount, only the former two items are considered.

Step 3: For each sample point x_i in D , calculate s_i according to

$$s_i = k(x_i, x_i) - \frac{2}{l} \sum_{j=1}^l k(x_i, x_j), \quad (14)$$

and update the training set through the selection of the bigger s_i from D .

3.2 Number Calculation of Pre-Selected Samples

In the SV pre-selection process of the supersphere model, the number of pre-selected SVs is an important value to be calculated. Based on (8), it can be find that $\alpha = \{\alpha_1, \dots, \alpha_l\}$ subjects to $\sum_{i=1}^l \alpha_i = 1, 0 \leq \alpha_i \leq \frac{1}{v \cdot l}, v \in (0, 1)$. Considering the SVs, set $\alpha_i > 0, i = 1, 2, \dots, l_{sv}$, which means the SV number is l_{sv} . With regard to the NSVs, set $\alpha_j = 0, j = l_{sv} + 1, l_{sv} + 2, \dots, l$, as a result, it can be easily concluded as (15)

$$\sum_{i=1}^l \alpha_i = \sum_{i=1}^{l_{sv}} \alpha_i + \sum_{j=l_{sv}+1}^l \alpha_i = 1 \quad (15)$$

Then, from the relationship of values between α_i and $\frac{1}{\nu l}$, it can be find that

$$1 = \sum_{i=1}^{l\nu} \alpha_i \leq l\nu \times \frac{1}{\nu l} \quad (16)$$

$$l\nu \geq \nu l$$

From the above-mentioned equations, we can conclude that the total SVs number of the supersphere OCSVM algorithm is νl at least, which also decides the number of pre-selected samples. About how to calculate and obtain the pre-selected samples, the process is: Firstly, rank the set of $s_i(i = 1, 2, \dots, l)$ by the result of (14) using a descending order. Secondly, select νl corresponding samples from the training set based on the former νl elements in the rank set, and set the value with s' for the νl th element. Thirdly, compare and calculate the value of difference between the latter element in the rank set and s' successively. If the calculated value is m , the threshold is Θ , and they satisfy $m < \Theta$, select the corresponding sample as a SV.

From the above-mentioned equations, we can conclude that the total SVs number of the supersphere OCSVM algorithm is νl at least, which also decides the number of pre-selected samples. About how to calculate and obtain the pre-selected samples, the process is: Firstly, rank the set of $s_i(i = 1, 2, \dots, l)$ by the result of (14) using a descending order. Secondly, select νl corresponding samples from the training set based on the former νl elements in the rank set, and set the value with s' for the νl th element. Thirdly, compare and calculate the value of difference between the latter element in the rank set and s' successively. If the calculated value is m , the threshold is Θ , and they satisfy $m < \Theta$, select the corresponding sample as a SV.

3.3 Complexity Analysis of Algorithm

The essence of the supersphere model construction is to solve the quadratic programming problem, the problem scale and the solving efficiency are related to the samples' number in training set. The calculation complexity is the cubic of the summation of samples when the traditional quadratic programming is executed. The algorithm's space complexity is primarily decided by the storage of kernel function matrix. Let the sample number of the supersphere is l , we can find that the space complexity is $O(l^2)$ for the storage of kernel function matrix, and the time complexity for model optimization is $O(l^3)$. There are two steps in the supersphere algorithm based on the support vector pre-selection: 1) Pre-select l' support vectors from the training set with l samples, this operation has a space complexity of $O(2l)$, namely two vectors which can contain l items are required. One vector is used for the storage of the intermediate result on $s_i, i = 1, \dots, l$ in (8) and (14), the other one is used for storage of the values of s_i . The time complexity of this step is $O(l^2)$, which is from the sorting of s_i . 2) Execute the optimization problem on the base of support vector set and build a model. For kernel function matrix, the space complexity is $O(l'^2)$, and the complexity of time consumption for model optimization is $O(l'^3)$ in this step. As the set of support vectors usually is a fraction of the total training sample set, the pre-selection of SVs method for the supersphere model could reduce the space and time complexity.

4 Fault Prediction and Quantitative Anomaly Measurement Using SV Pre-Selection OCSVM

The OCSVM model based on SVs pre-selection has a faster modeling speed, which can improve the efficiency of fault prediction. For the samples with the form of feature vectors, the model can be applied directly. But for the time series, the phase space reconstruction is firstly needed, and then the fault prediction model with OCSVM could be built. The phase space reconstruction is a research method for system's dynamic behavior based on the limited measured

data and the attractor reconstruction. In order to restore the phase space's geometrical structure of dynamic systems from the one-dimensional time series, the one-dimensional time series should be embedded into the m -dimensional space, which has the form as: suppose the observed time series from system is $x(t)$, with $t = 1, 2, \dots, N$. The dimension of the phase space is m , the fixed time delay is called the embedded delay τ , so the sample points in the space of phase are $X(t) = (x(t), x(t - \tau), x(t - 2\tau), \dots, x(t - (m - 1)\tau))^T$, with $t = 1, 2, \dots, N_p$, $N_p = N - (m - 1)\tau$. The time series can be used for novelty detection after the phase space reconstruction, which also provides the theoretical basis for the time series prediction using OCSVM. The main procedure on fault prediction using SVs pre-selection is:

- 1) Suppose the collected sample series in the system's normal state is $x(t)$, $t = 1, 2, \dots, N$. Execute the reconstruction of phase space for the time series, and the sample set can be obtained as $\mathbf{D} = \{\mathbf{x}_i \in R^d | i = 1, \dots, N - (m - 1)\tau\}$.
- 2) Pre-select the support vectors from \mathbf{D} , and build the OCSVM model as the fault prediction model based on the selected sample set. For the supersphere model, calculate the variable's value range representing the relative location relationship between the samples in \mathbf{D} and the supersphere according formula (10). Then the lower bounds of the two ranges are regarded as the respective threshold reference to judge whether the fault trend exist or not.
- 3) Decide whether the time series value $x(t')$ at time t' indicate the fault trend. The method is: firstly, construct the sample being adapt to the model as $\mathbf{x}(t') = (x(t'), x(t' - \tau), x(t' - 2\tau), \dots, x(t' - (m - 1)\tau))^T$ based on $x(t')$. Then $\mathbf{x}(t')$ is substitute to the OCSVM model after the support vector pre-selection, and the variable value representing the relative location relationship between $\mathbf{x}(t')$ and the supersphere can be calculated. Finally, the variable value is used for judging whether the fault trend exist through the comparison to the fault threshold.

To quantitatively measure the anomaly level of one sample, the anomaly indexes (AIs) [6] can be defined as follows.

Set a system's normal state ω_1 can be described by a given sample set $\{v_i, i = 1, \dots, l\}$, if the sample set is known, it means the normal state ω_1 of the system is also known, but the samples in the system's anomaly state or in system's fault state are both unknown. Then set $p(v_i|\omega_1)$ as the probability density of v_i , and let $\rho_1 = \min(p(v_i|\omega_1))$, $\rho_2 = \max(p(v_i|\omega_1))$, $\rho = p(v'|\omega_1)$, where v' is a new sample. As a result, the anomaly level of sample v' can be quantitatively measured by the anomaly indexes $AIs(v')$, which has a form as (17)

$$AIs(v') = \begin{cases} 0 & \rho \geq \rho_2 \\ 0.5 \frac{\rho - \rho_2}{\rho_1 - \rho_2} & \rho_2 + (\rho_1 - \rho_2)/0.5 < \rho < \rho_2 \\ 1 & \rho \leq \rho_2 + (\rho_1 - \rho_2)/0.5 \end{cases} \quad (17)$$

For the samples in ω_1 , it can be found that $0 \leq AI(v_i) \leq 0.5$, $i = 1, \dots, l$. On this basis, suppose a threshold $\theta(0 \leq \theta \leq 1)$ is selected, if $AI(v') \leq \theta$, it means the new sample v' having not fault trend; if $AI(v') > \theta$, it means the new sample v' has a fault trend. In other words, if AIs satisfies the condition $AI(v') \leq \theta$, the sample v' can be regarded as a normal sample; and if AI satisfies the condition $AI(v') > \theta$, it implies that the new sample v' has a fault trend.

5 Experiments and Result Analysis

5.1 A Simulation Experiment and Analysis of Result

For the supersphere model, the training set is constructed from $\theta \in U[0, 2\pi]$, $\rho \in U[6, 10]$, and 1000 samples are random generated, so the coordinates of θ and ρ are dimensionless. The RBF kernel is chosen and the pre-selected SVs are shown in Fig. 2.

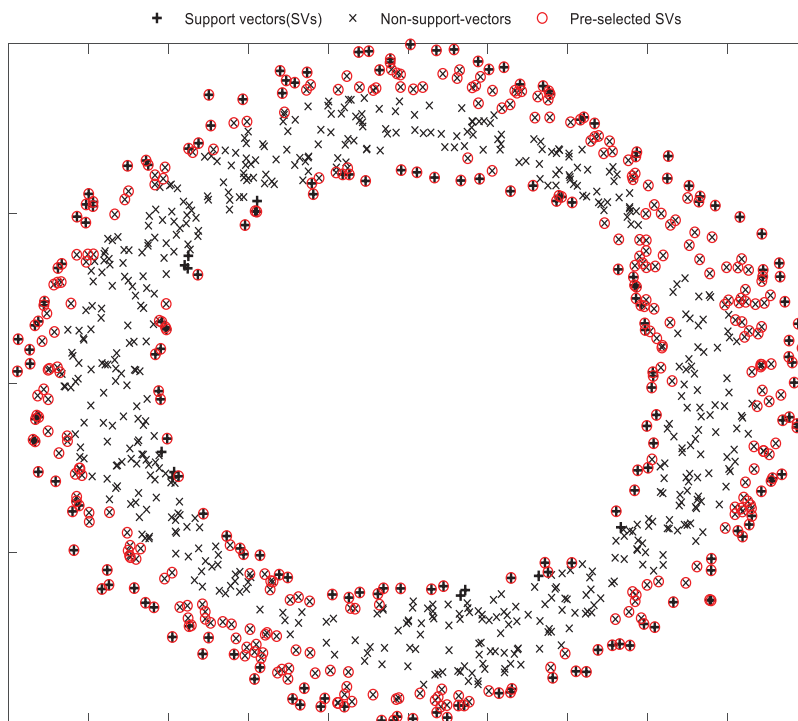


Figure 2: Support vector pre-selection result of the supersphere model

Considering Fig. 2, it can be concluded that the obtained samples' set through SV pre-selection of the supersphere model covers most of the SVs. As the supersphere model's training process is entirely unrelated to NSVs, it's obvious that the SV pre-selection approach may enhance the training efficiency in the condition of guaranteeing the model's novelty detection capacity.

The supersphere algorithm's training results of the simulation samples are shown in Table 1. From data in the table, we also can find the speed of modeling is tremendously enhanced; as a result, the fault prediction and the anomaly measurement efficiency can also be enhanced.

Table 1: Comparison of training result for the supersphere model

Algorithms	Number of training samples	Number of support vectors (Pre-selection)	Number of support vectors	Modeling time (s)
Supersphere	1000	Non-selection	212	15.28
Supersphere (Pre-selection)	1000	450	201	4.35

5.2 Experiments for the Single-Dimensional Drift Data of Gyroscope

When the gyroscope devices are working in the normal state, their drift coefficients satisfy a certain transformation rule. When the fault trend exists, the original transformation rule will inevitably occur the drift, even exceed the normal working range. If we find the drift of transformation rule by the fault prediction model, or the working parameters will soon deviate from the normal working range, the fault trend must be predicted. If we find the transformation rule keeps the original state, and locates in the normal range, the conclusion having no fault trend should be given.

In this experiment, the samples come from the reliability experiments of the new gyroscopes. The experiment parameters include the gravitational acceleration g and the geographical latitude R , where $g = 979.4121 \text{ cm/s}^{-2}$, $R = 34.6006^\circ$.

A data gathering system with special sensor is used for data acquisition and drift data storage. Based on the 91 groups of testing data in the dataset, the supersphere model is trained to detect whether the fault trend exists. Set the dimension of the unlabeled samples is 3, and construct 89 samples, the former 75 samples are used for modeling, and the latter 15 samples are used for fault prediction.

To evaluate the reliability of the experimental results, the experiment with the supersphere model of OCSVM is executed, and the modeling result is described in Fig. 3, which is the support vector pre-selection result. In Fig. 3a, the horizontal ordinates are the sample series numbers, and the vertical ordinate are the corresponding values of Lagrange multipliers, the values decide whether the samples are support vectors. Fig. 3b shows the location relationship between the samples and the supersphere, which is a unitless value of distance. As the pre-selected sample set includes all the SVs of the supersphere model, the results have the same form for the supersphere and the supersphere(SV pre-selection).

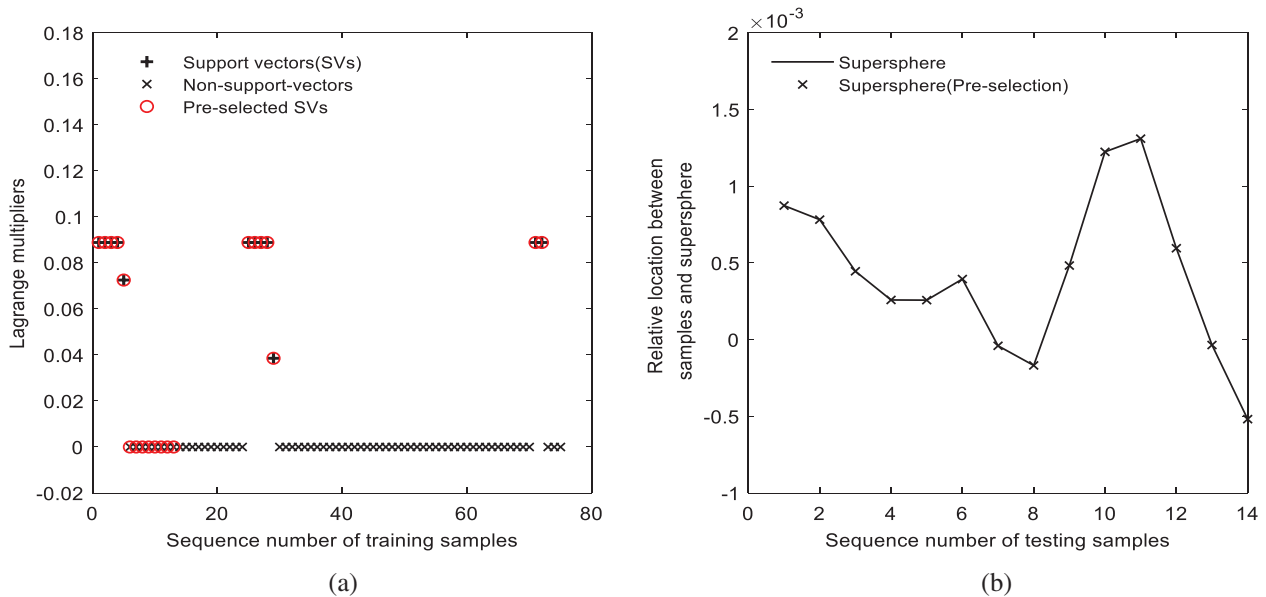


Figure 3: Modelling result of supersphere. (a) Support vector pre-selection (b) Location relationship between the samples and the supersphere

Utilizing the modeling result, the density distribution of the variable representing the relative location of the samples and the supersphere can be estimated as Fig. 4a. On this basis, the sample's anomaly indexes (AIs) are calculated. As shown in Fig. 4b, we still can see that the AIs are all below 0.5, namely there is not any fault trend exists in the gyroscope.

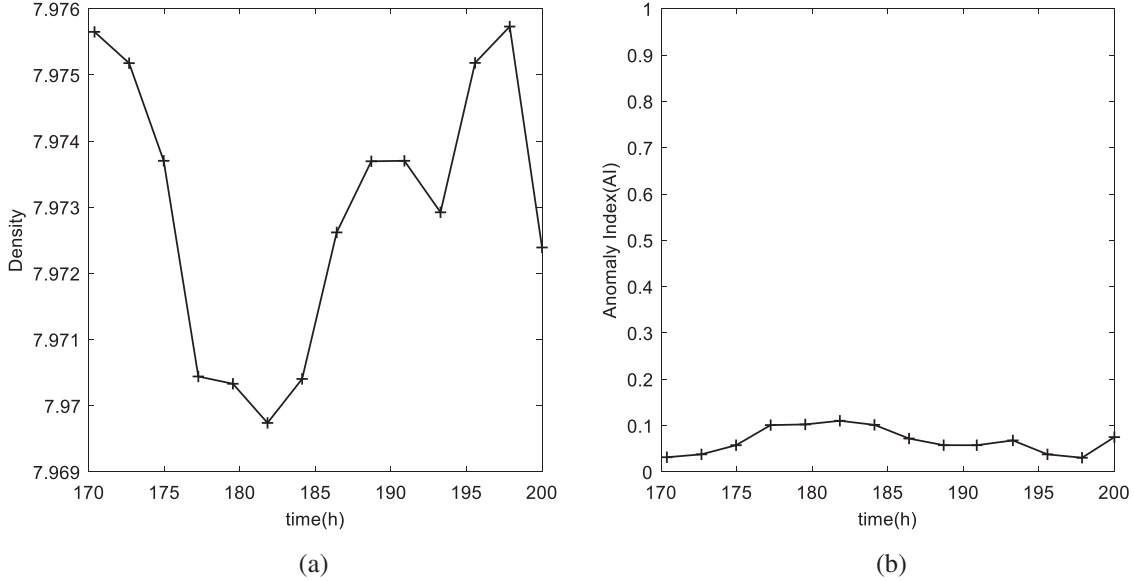


Figure 4: Result of anomaly indexes. (a) Density estimation (b) Calculation of AIs

5.3 Experiments for a Precision Inertia System Platform

The inertia system includes three gyroscopes, and we can get 8 drift ampere values of the precision platform by testers, which are the I_{Y-up} , I_{Y-down} , I_{X-up} , I_Z^{X-up} , I_{X-down} , I_Z^{X-down} , I_{Z-up} , I_{Z-down} . Based on the 8 values, 6 drift coefficients of the three gyroscopes can be deduced, the methods are: $K_{OY} = \frac{1}{2}(I_{Y-up} + I_{Y-down}) \cdot K_c^Y$, $K_{OX} = \frac{1}{2}(I_{X-up} + I_{X-down}) \cdot K_c^X$, $K_{OZ} = \frac{1}{2}(I_{Z-up} + I_{Z-down}) \cdot K_c^Z$, $K_{SY} = -\frac{1}{2}(|I_{Y-up}| + |I_{Y-down}|) \cdot K_c^Y + \Omega \sin \varphi$, $K_{SX} = -\frac{1}{2}(|I_{X-up}| + |I_{X-down}|) \cdot K_c^X + \Omega \sin \varphi$, $K_{IZ} = \frac{1}{2}(I_Z^{X-up} + I_Z^{X-down}) \cdot K_c^Z$, where φ is the geographical latitude of testing point, K_c^X , K_c^Y , K_c^Z , Ω are constants if the testing target is known.

The sample form of the precision platform is $X = \{K_{OY}, K_{OX}, K_{OZ}, K_{SY}, K_{SX}, K_{IZ}\}$. There are 130 groups of data with the 6-dimensional drift coefficient gathering from 2001 to 2007, as shown in Fig. 5. The values in this figure are all practical values of the precision platform's drift coefficients. The horizontal ordinates are the sample series numbers, and the vertical ordinate are the practically measured data.

The goal of the experiment is to analysis the platform's working state in 2007 based on the data from 2001 to 2006. There are 74 samples matching the system performance requirement in

the six years, and these data are considered as the training samples to build the supersphere model with OCSVM. Fig. 6a is the support vector pre-selection result. In this subfigure, the horizontal ordinates are the sample series numbers of the drift time series, and the vertical ordinates are the values of Lagrange multipliers, the selected support vectors are used to build the supersphere model. The variable values representing the relative location of the samples and the supersphere are then calculated, as shown in Fig. 6b. In this subfigure, the horizontal ordinates are the whole samples including the training and the testing samples, and the vertical ordinates are the relative distances between the samples and the supersphere. If the relative distance is negative, it means the sample is inside the supersphere; if the distance is 0, it means the sample is on the supersphere surface; and if the distance is larger than 0, it means the sample is outside the supersphere.

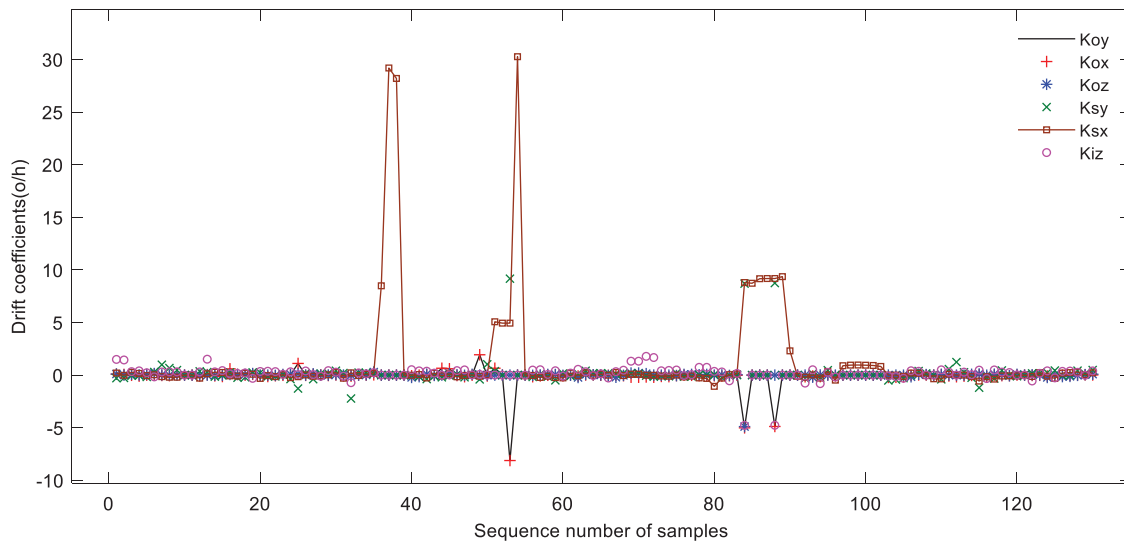


Figure 5: Practical values of the precision platform’s drift coefficients

Utilizing the modeling result, we can calculate the sample’s anomaly indexes (AIs) by probability density estimation method, as shown in Fig. 7. It can be found that the AIs of the anomaly samples are all below 0.5 among the 118 samples from 2001 to 2006, which means the AIs can distinguish the normal samples and the anomaly samples. Considering the 4th sample in the 12 samples from 2007, namely the 122th sample in Fig. 7, the AI value is 0.5268, the AIs of the remaining 8 samples are all below 0.5. By analyzing the drift data of the platform’s normal state from 2001 to 2006, we can obtain the practical drift range is $[-0.5423, 0.5508]$. It means that although the 122th sample’s AI value is bigger than 0.5, namely the platform is beyond the normal working range in the past, but it still didn’t exceed the drift range of the engineering permission. So, we can conclude that the precision platform had no obvious drift fault trend and had a stable running state in 2007.

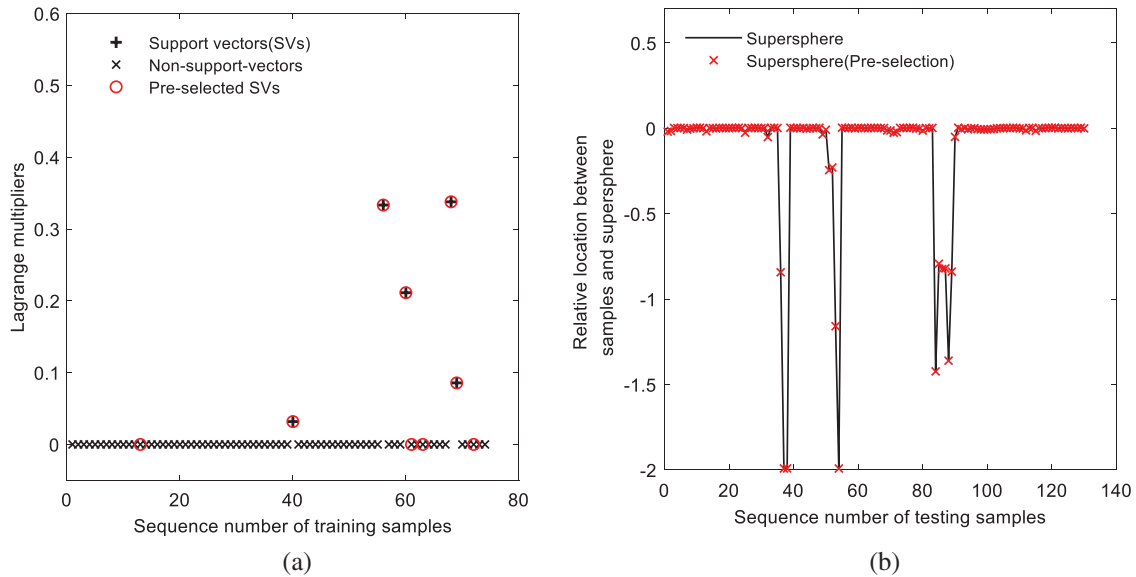


Figure 6: Modelling result of supersphere. (a) Support vector pre-selection (b) Location relationship between the samples and the supersphere

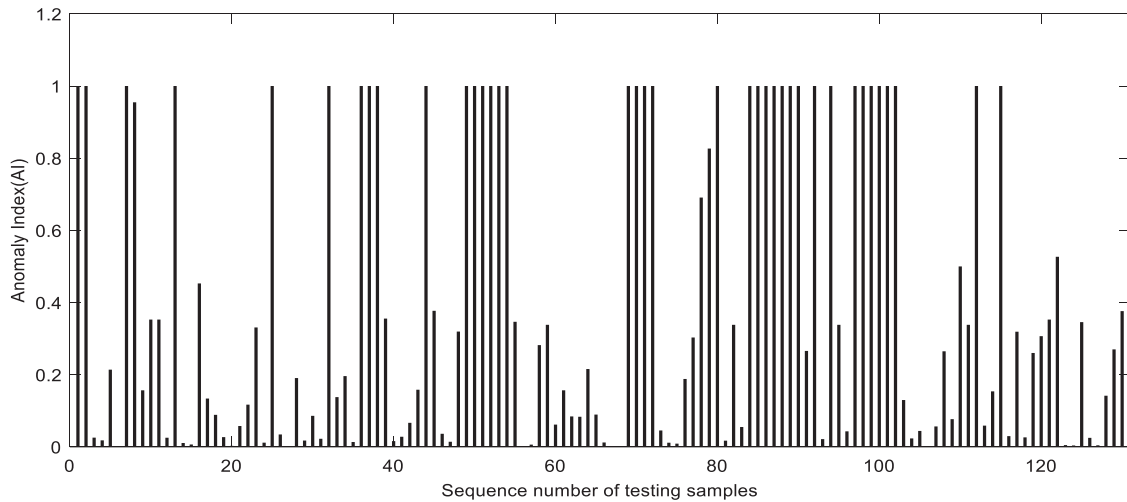


Figure 7: Result of anomaly indexes (AI)

6 Conclusions

In this paper, a fast small-sample supersphere one-class SVM modeling method using SVs pre-selection is systematically studied, aiming at the data-driven fault state prediction and quantitative degree measurements. (1) The essence of the method lies in two aspects. firstly, construct the supersphere space model based on the samples from the system's normal state; secondly, compare the relative location relationship between the new sample and the supersphere to judge whether the sample deviates from the system's normal state. As a result, the fault prediction goal is reached. (2) The advantages of the method lie in the easily acquisition of data samples, the

fast modeling and the strong sensibility to the system's anomaly. (3) The disadvantage is that it cannot track the fault trend continually, so there is a lack of further understanding for the fault trend. (4) When the system's fault trend occurs, how to build the efficient prediction model is the key problem to be solved. The OCSVM fault prediction model is mainly studied based on SV pre-selection. To improve the modeling efficiency, the SV pre-selection method for the supersphere model is proposed. The methods can extract the outside points and boundary points of the supersphere, and can give good fault prediction capability and effectiveness of quantitative anomaly measurement for precision inertial devices. The experiments on precision inertial systems testify that the proposed method has good fault prediction speed and precision, and can effectively quantitative measure the anomaly level of systems.

Acknowledgement: We thank the editors for their rigorous and efficient work, and we also thank the referees for their helpful comments.

Funding Statement: This work was jointly supported by the National Natural Science Foundation of China (Grant No. 61403397) and the Natural Science Basic Research Plan in Shaanxi Province of China (Grant Nos. 2020JM-358, 2015JM6313).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Zhou, D. H., Xu, G. B. (2009). Fault prediction for dynamic systems with state-dependent faults. *Proceedings of the 4th International Conference on Innovative Computing, Information and Control*, pp. 207–210. Piscataway, NJ, USA.
2. Zhang, X. Y., Li, C. S., Wang, X. B., Wu, H. M. (2021). A novel fault diagnosis procedure based on improved symplectic geometry mode decomposition and optimized SVM. *Measurement*, 173. DOI 10.1016/j.measurement.2020.108644.
3. Kumari, S., Sachin, K., Saket, R. K., Sanjeevikumar, P. (2020). Open-circuit fault diagnosis in multilevel inverters implementing PCA-WE-SVM technique. *IEEE Transactions on Industry Applications*, 1–9 (IEEE Preprint).
4. Li, Z. Z., Wang, L. D., Yang, Y. Y. (2020). Fault diagnosis of the train communication network based on weighted support vector machine. *IEEE Transactions on Electrical and Electronic Engineering*, 15(7), 1077–1088. DOI 10.1002/tee.23153.
5. Yang, K., Kpotufe, S., Feamster, N. (2021). An efficient one-class SVM for anomaly detection in the Internet of Things. arXiv: 2104.11146.
6. Wang, H. Q., Cai, Y. N., Fu, G. Y., Wu, M., Wei, Z. H. (2018). Data-driven fault prediction and anomaly measurement for complex systems using support vector probability density estimation. *Engineering Applications of Artificial Intelligence*, 67, 1–13. DOI 10.1016/j.engappai.2017.09.008.
7. Wang, J. S., Chiang, J. C., Yang, Y. T. (2007). Support vector clustering with outlier detection. *Proceedings of the 3th International Conference on Intelligent Computing*, pp. 423–431. Qingdao, China.
8. Davy, M., Desobry, F., Gretton, A. (2006). An online support vector machine for abnormal events detection. *Signal Processing*, 86, 2009–2025. DOI 10.1016/j.sigpro.2005.09.027.
9. Shi, Q., Zhang, H. (2020). Fault diagnosis of an autonomous vehicle with an improved SVM algorithm subject to unbalanced datasets. *IEEE Transactions on Industrial Electronics*, 68(7), 6248–6256. DOI 10.1109/TIE.2020.2994868.
10. Rasheed, W., Tang, T. B. (2020). Anomaly detection of moderate traumatic brain injury using auto-regularized multi-instance OCSVM. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(1), 83–93. DOI 10.1109/TNSRE.7333.

11. Dai, X. W., Gao, Z. W. (2013). From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis. *IEEE Transactions on Industrial Informatics*, 9(4), 2226–2238. DOI 10.1109/TII.2013.2243743.
12. Fang, Y., Min, H., Wang, W. (2020). A fault detection and diagnosis system for autonomous vehicles based on hybrid approaches. *IEEE Sensors Journal*, 20(16), 9359–9371. DOI 10.1109/JSEN.2020.2987841.
13. Du, X. (2019). Fault detection using bispectral features and one-class classifiers. *Journal of Process Control*, 83, 1–10. DOI 10.1016/j.jprocont.2019.08.007.
14. Gharoun, H., Keramati, A., Nasiri, M. M., Azadeh, A. (2019). An integrated approach for aircraft turbofan engine fault detection based on data mining techniques. *Expert Systems*, 36(2), 1–18. DOI: 36.10.1111/exsy.12370.
15. FernaNdez-Francos, D., MartiNez-Rego, D., Fontenla-Romero, O., Alonso-Betanzos, A. (2013). Automatic bearing fault diagnosis based on one-class v-SVM. *Computers & Industrial*, 64(1), 357–365. DOI 10.1016/j.cie.2012.10.013.
16. Amer, M., Goldstein, M., Abdennadher, S. (2013). Enhancing one-class support vector machines for unsupervised anomaly detection. *Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description*, pp. 8–15. Chicago, IL, USA.
17. Yin, S., Zhu, X. P., Jing, C. (2014). Fault detection based on a robust one class support machine. *Neurocomputing*, 145, 263–268. DOI 10.1016/j.neucom.2014.05.035.
18. Xiao, Y. C., Wang, H. G., Xu, W. L., Zhou, J. W. (2016). Robust OCSVM for fault detection. *Chemometrics and Intelligent Laboratory Systems*, 151, 15–25. DOI 10.1016/j.chemolab.2015.11.010.
19. Yan, K., Ji, Z. W., Shen, W. (2017). Online fault detection methods for chillers combining extended kalman filter and recursive OCSVM. *Neurocomputing*, 228, 205–212. DOI 10.1016/j.neucom.2016.09.076.
20. Huang, N. T., Chen, H. J., Zhang, S. X., Cai, G. W., Li, W. G. et al. (2016). Mechanical fault diagnosis of high voltage circuit breakers based on wavelet time-frequency entropy and one-class support vector machine. *Entropy*, 18(1), 7. DOI 10.3390/e18010007.
21. Perera, P., Oza, P., Patel, V. M. (2021). One-class classification: A survey. arXiv preprint. arXiv:2101.03064v1.
22. Kim, S., Lee, K., Jeong, Y. S. (2021). Norm ball classifier for one-class classification. *Annals of Operations Research*, 303(1), 433–482. DOI 10.1007/s10479-021-03964-x.
23. Kumar, B., Sinha, A., Chakrabarti, S. (2021). A fast learning algorithm for one-class slab support vector machines. *Knowledge-Based Systems*, 228, 107267. DOI 10.1016/j.knosys.2021.107267.
24. Juhamatti, S., Daniel, S., Jan, L., Allan, T. (2019). Detection and identification of windmill bearing faults using a one-class support vector machine (SVM). *Measurement*, 137(4), 287–301. DOI 10.1016/j.measurement.2019.01.020.
25. Yan, X. A., Jia, M. P. (2018). A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing. *Neurocomputing*, 313(11), 47–64. DOI 10.1016/j.neucom.2018.05.002.
26. Yi, Y. G., Shi, Y. J., Wang, W. L., Lei, G., Dai, J. Y. et al. (2021). Combining boundary detector and SND-SVM for fast learning. *International Journal of Machine Learning and Cybernetics*, 12(3), 689–698. DOI 10.1007/s13042-020-01196-2.
27. Miao, X. D., Liu, Y., Zhang, H. Q., Li, C. G. (2018). Distributed online one-class support vector machine for anomaly detection over networks. *IEEE Transactions on Cybernetics*, 49(4), 1475–1488. DOI 10.1109/TCYB.6221036.
28. Bhland, M., Doneit, W., Grll, L. (2019). Automated design process for hybrid regression modeling with a OCSVM. *AT-Automatisierungstechnik*, 67(10), 843–852. DOI 10.1515/auto-2019-0013.
29. Luo, X., Sun, J., Wang, L., Wang, W., Zhao, W. et al. (2018). Short-term wind speed forecasting via stacked extreme learning machine with generalized correntropy. *IEEE Transactions on Industrial Informatics*, 14(11), 4963–4971. DOI 10.1109/TII.2018.2854549.
30. Chalapathy, R., Menon, A. K., Chawla, S. (2020). Anomaly detection using one-class neural networks. *Machine Learning*. arXiv: 1802.06360v2.

31. Ruff, L., Vandermeulen, R., Goernitz, N., Deecke, L., Siddiqui, S. A. et al. (2018). Deep one-class classification. *Proceedings of the 35th International Conference on Machine Learning (ICML)*, pp. 4393–4402. Stockholm, Sweden.
32. Liznerski, P., Ruff, L., Vandermeulen, R. A., Franks, B. J., Kloft, M. et al., (2021). Explainable deep one-class classification. *Proceedings of ICLR'2021*. arXiv:2007.01760v3.
33. Oza, P., Patel, V. M. (2019). One-class convolutional neural network. *IEEE Signal Processing Letters*, 26(2), 277–281. DOI 10.1109/LSP.2018.2889273.
34. Sabokrou, M., Khalooei, M., Fathy, M., Adeli, E. (2018). Adversarially learned one-class classifier for novelty detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3379–3388. Salt Lake City, UT, USA. arXiv:1807.02588v2.
35. Plakias, S., Boutalis, Y. (2019). Exploiting the generative adversarial framework for one-class multi-dimensional fault detection. *Neurocomputing*, 332, 396–405. DOI 10.1016/j.neucom.2018.12.041.