An Aeroengine Gas Path Anomaly Detection Method in The Case of Sample Imbalance

Kang Wu
School of Mechanical Engineering
Harbin Institute of Technology (Weihai)
Shandong, China
wukangxxx@126.com

Shisheng Zhong*
School of Mechanical Engineering
Harbin Institute of Technology (Weihai)
Shandong, China
zhongss@hit.edu.cn
*Corresponding author

Xuyun Fu
School of Mechanical Engineering
Harbin Institute of Technology (Weihai)
Shandong, China
fuxuyun@hit.edu.cn

Changtsing Wei
School of Mechanical Engineering
Harbin Institute of Technology (Weihai)
Shandong, China
weichangtsing@163.com

***Abstract*—In the process of aeroengine anomaly detection, there is always an unbalance distribution among the samples of gas path performance parameters, that is, the number of normal samples is much larger than the number of abnormal samples. In addition, this imbalance will worsen with time, which leads to the classifier paying too much attention to normal samples in the process of model training. Thus, the recognition rate of abnormal samples will reduce significantly. To solve the above problems, an adaptive decision threshold support vector machine (ADT-SVM) is proposed and applied to the anomaly detection of aeroengine. Firstly, this paper analyzes the influence of the unbalanced training data on the performance of the traditional classification model. Then the concept of decision threshold is introduced and introduced into support vector machine for anomaly detection. Finally, an adaptive method is proposed to calculate the decision threshold based on the equal expected number of samples, and the performance of the adaptive threshold and the traditional default threshold SVM is compared through experiments, which show that the adaptive threshold is effective in solving the problem of the classification performance degradation of unbalanced gas path performance parameters.**

Keywords-aeroengine; anomaly detection; imbalanced sample; adaptive decision threshold; SVM

I. INTRODUCTION

The working state of aeroengine plays an important role in flight safety [1]. Therefore, it is an effective method to ensure the flight safety of aircrafts to accurately judge the working state of engines and timely detect the abnormal conditions of engines [2]. The working condition of aircraft engines has a very close link to the performance parameters of gas path [3], and through certain methods to rapidly and accurately detect the gas path performance parameters of anomaly samples, which is of great significance to control the work status of engines in real time, prevent and troubleshoot problems, and ensure the flight safety [4]. However, there is always the phenomenon of unbalanced distribution among samples of aeroengine gas path performance parameters [5], that is, the number of normal samples is far greater than the number of fault samples and it deteriorates over time [6]. As a result, the classifier pays too much attention to normal samples in the process of model training, so that the classification performance of abnormal samples declines seriously [7]. For the anomaly detection of aeroengine gas path performance parameters, its essence is a classification problem. And for classification problems, many existing algorithms have been quite mature. But most of the learning algorithm assume or expect data sets with balanced distribution or equal misclassification cost [8]. And the total accuracy of the classifier is set as the evaluation index. Therefore, when dealing with complex unbalanced data sets, traditional algorithms cannot effectively represent the distribution features of the unbalanced data [9]. Because the identification of the minority samples is much more difficult than the majority samples, and it is difficult to determine the true boundary of the minority samples [10]. Therefore, traditional classification algorithms often tend to divide minor class samples into the major class [11]. And the prediction effect of the minor class is not good, which leads to a large number of abnormal samples being mistaken for normal samples in the anomaly detection process [12]. Obviously, it has lost the meaning of anomaly detection.

At present, the research in the field of unbalanced data mining has become more and more important. Because people gradually realize that the distribution of data sets in reality is basically unbalanced, and what's different is just the imbalance ratio of the data [13]. At the same time, this imbalance has seriously affected the performance of classification algorithm, as shown in Fig. 1.

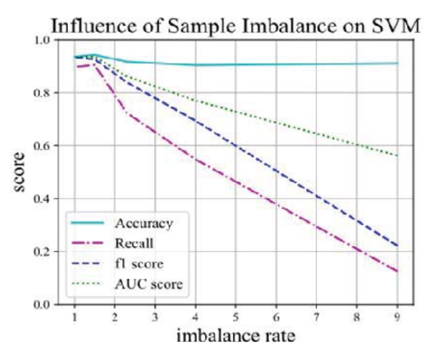

Figure 1. Influence of sample imbalance on svm

In order to improve the recall rate of abnormal samples, an adaptive decision threshold is proposed to replace the traditional default threshold to train SVM for the anomaly detection of aeroengine gas path performance parameters in the case of sample imbalance.

II. Overview of Methods and Theories

A. The Principles of SVM

Support vector machine (SVM) is a common supervised learning method, which is suitable for small samples, nonlinear and high-dimensional data [14]. It is a linear binary classifier that defines the maximum interval in the feature space [15]. And by introducing kernel function, the original sample space is mapped to Hilbert space, so that the problem that is linearly indivisible in the original space is linearly separable in the Hilbert space [16]. In the case of linear separable training data set, SVM solves corresponding convex quadratic programming problems through interval maximization or equivalence to obtain the separation hyperplane [17], as shown in (1).

$$w^{\mathrm{T}}x + b = 0 \tag{1}$$

In (1), x is the sample data, w is the normal vector, which determines the direction of the hyperplane; b is the intercept, which determines the distance between the hyperplane and the origin. The corresponding classification decision function is

$$y = \operatorname{sgn}(w^{\mathrm{T}}x + b) \tag{2}$$

In (2), y is the category label of samples and sgn () is the signum function, whose function image is shown in Fig. 2.

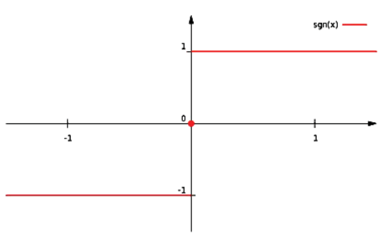

Figure 2. Signum function

There is

$$\begin{cases} w^{\mathrm{T}}x_i + b \geq +1, & y_i = +1; \\ w^{\mathrm{T}}x_i + b \leq +1, & y_i = -1. \end{cases} \tag{3}$$

For the case of linear indivisibility, the main idea of SVM is to map the input vector to a high-dimensional eigenvector space and construct the optimal classification plane in the eigenspace [18].

To take the transformation $\varPhi$ of x from the input space R^n to the feature space H, and get

$$x \to \varPhi(x) = (\varPhi_1(x), \varPhi_2(x), ..., \varPhi_n(x))^{\mathrm{T}} \tag{4}$$

And replace the input vector x with the feature vector $\varPhi(x)$, then the optimal classification function can be obtained as follows

$$y = \operatorname{sgn}(w^{\mathrm{T}}\varPhi(x) + b) \tag{5}$$

B. Decision Threshold

For the classification problem, this paper only discusses the binary classification problem. The essence of the model is the same as that of the regression model. The difference is that the classification model further discretizes the output of the regression model.

Given data set $D = \{(x_1, y_1), (x_2, y_2), ..., (x_m, y_m)\}$, where $x_i = (x_{i1}, x_{i2}, ..., x_{id})$, x_{id} represents the value on the attribute d of the sample i, $y_i \in \mathbf{R}$, and the linear regression model tries to learn a function through the linear combination of attributes to predict its actual label as accurately as possible, as shown in (6).

$$f(x_i) = w^{\mathrm{T}}x_i + b \tag{6}$$

To make $f(x_i) \simeq y_i$, and the classification model introduces activation functions, such as signum function and sigmoid function, on the basis of this linear regression model, to discretize the output of the regression model, so as to achieve the purpose of outputting category labels.

For SVM, the corresponding convex quadratic planning problem is solved by interval maximization or equivalent [19], and the optimal weight vector w^{T} and the optimal bias vector b can be obtained and output $f(x_i) = w^{\mathrm{T}}x_i + b$, which is The distance from a sample point x_i to the classification plane, and then through the signum function map it to the $\{-1,+1\}$, outputting category label. Positive distance is a class of the samples, negative distance is another class of the samples [20].

At this time, the activation function of SVM is changed from signum function to sigmoid function, so that SVM does not directly output the category label belonging to a certain class, but output a continuous value between [0,1], denoted as y_i', that is $y_i' = sigmoid(w^{\mathrm{T}}x_i + b)$, as shown in Fig. 3.

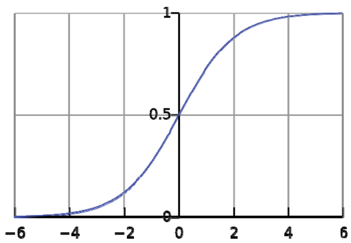

Figure 3. Sigmoid function

Obviously, the closer y_i' is to 1, the higher the probability that it belongs to the positive class; On the contrary, the closer y_i' is to 0, the higher the probability that it belongs to the negative class. In this paper, y_i' is called the class probability, which is the probability of sample x_i belonging to the positive class. The probabilistic boundary value of positive class and negative class is the decision threshold, denoted as t, there t=0.5.

$$\begin{cases} y_i' > t & , y_i = +1; \\ y_i' < t & , y_i = -1. \end{cases} \tag{7}$$

In (7), y_i' is the class probability of positive class of sample x_i, t is the decision threshold, and y_i is the class label of the sample.

Therefore, for the classification problem, we can output the class probability $y_i' = sigmoid(w^{\mathrm{T}} x_i + b)$ of a certain category based on its regression value, such as the above $f(x_i) = w^{\mathrm{T}} x_i + b$, and choose the category with the largest probability as the output category, which is also the practice of most traditional classifiers. The default decision threshold there is 0.5, and the category with probability greater than 0.5 is taken as the output.

However, the premise of this approach is that the sample number of positive class and negative class is equal or almost the same, i.e., their prior probability is equal. However, when these classification models are used to deal with unbalanced samples, e.g., the abnormal detection of gas path performance parameters of aeroengine will result in the aforementioned degradation of classification performance and the reduction of recognition rate of abnormal samples due to the fact that the number of normal samples is much larger than that of abnormal samples.

C. *The Influence of Decision Threshold on Classification Surface*

Traditional classification model with default decision threshold, when dealing with balanced samples, will not be accompanied by performance degradation problems. However, for aeroengine gas path performance parameter data sets with unbalance problems, the default decision threshold makes the surface of the classification model tend to favor the minor class. And more abnormal samples will be mistaken for normal samples, the result is that the recognition rate of the minor samples will be significantly reduced. This kind of phenomenon in the process of aircraft engines anomaly detection is not allowed. Because every leakage alarm could lead to a serious air accident, and the anomaly detection process has lost its meaning.

On the contrary, people can often accept the system to have a few false alarms, i.e., the normal samples are mistakenly divided into abnormal samples. And the result of this is that the higher recognition rate of abnormal samples can be obtained, and more abnormal samples will be detected, which greatly ensures the flight safety of the aircraft.

This paper attempts to correct the classification surface of the traditional classification model by changing the decision threshold. Thus, by allowing a small amount of major class samples are divided into the minor class, the classification surface can return to the neutral position, not because of the imbalance sample distribution to be partial to the minor class. More samples of the minor class can be detected, to improve the recognition rate of abnormal samples while allowing a certain number of false alarms. The specific process is shown in Fig. 4, and the classification surface of SVM with default decision threshold is biased towards to abnormal samples as shown in the dotted line in Fig. 4. Some abnormal samples cannot be detected, but by changing the decision threshold, the classification surface returns to the neutral position or biased towards to the major class. At this time, more abnormal samples will be detected.

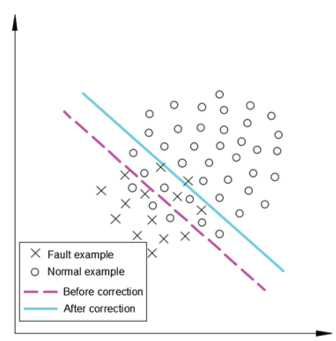

Figure 4. Classification surface correction process

D. *Adaptive Threshold Support Vector Machines*

Most of learning algorithms assume or expect data sets with balanced distribution or equal misclassification cost, and the total accuracy of the classifier is set as the evaluation index. Therefore, when dealing with complex unbalanced data sets, the traditional algorithms cannot effectively represent the distribution features of the unbalanced data. Because the identification of the minority samples is much more difficult than the majority samples, and it is difficult to determine the true boundary of the minority samples. Therefore, the traditional classification algorithms often tend to divide the minor class samples into the major class. And the prediction effect of the minor class is not good, which leads to a large number of abnormal samples being mistaken for the normal sample in the anomaly detection process. Obviously, it is not suitable for unbalanced data classification task. Aiming at the above problems, this paper proposes an adaptive decision threshold SVM (ADT-SVM) method for dealing with unbalanced sample distribution.

For samples with unbalanced distribution, the performance of the classification algorithm will not be affected by the sample imbalance only when the expected sample number of each class is equal, as shown in (8).

$$\begin{aligned} & E(N) = N_n * p_n = N_f * p_f \\ s.t. \quad & p_n + p_f = 1 \end{aligned} \tag{8}$$

In (8), $E(N)$ is the expectation of the samples number in the case of sample balance, N_n is the number of normal samples, N_f is the number of abnormal samples, p_n is the prior probability of normal samples, and p_f is the prior probability of abnormal samples.

By deforming (8), the following equation can be obtained

$$p_n = \frac{p_n}{p_f + p_n} = \frac{N_f}{N_f + N_n} = \frac{1}{1 + IR} \tag{9}$$

In (9), *IR* is the imbalance rate of the data set, i.e., the ratio of the number of normal (majority) samples to the number of abnormal (minority) samples.

When $y_i' < t$, the sample is the normal sample. Obviously, the probability of the normal samples $p_n = t - 0 = t$. Finally, the adaptive decision threshold is set as (10).

$$t = \max(\frac{1}{IR + 1}, t_{\min}) \tag{10}$$

In (10), $t_{\min}$ is the minimum threshold value of the decision threshold. The purpose of setting this value is to prevent decision threshold of extreme unbalanced samples from being too small, resulting in excessive correction of the decision-making process and seriously decrease in *accuracy*.

E. Evaluation Index

1) Confusion Matrix

TABLE I CONFUSION MATRIX OF CLASSIFICATION RESULTS

Fact Situation	Prediction Situation	
	Positive	***Negative***
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

2) Accuracy

$$accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{11}$$

3) Recall and precision

$$recall = \frac{TP}{TP + FN} \tag{12}$$

$$precision = \frac{TP}{TP + FN} \tag{13}$$

In the process of the aeroengine gas path performance parameters anomaly detection, abnormal sample recognition rate may be expressed by *recall*, and the normal sample recognition rate can be expressed by *precision*. The two are a pair of contradictory measurements, generally speaking, *recall* is often low when *precision* is high, while *precision* is often low when *recall* is high.

4) F1 score

$$f1 = \frac{1}{2}(\frac{1}{R} + \frac{1}{P}) = \frac{2 \times R \times P}{R + P} \tag{14}$$

F1 score is a comprehensive index, which is used to measure the performance of classifiers. *f1 score* will increase only when both *recall* and *precision* of the model increase.

III. EXPERIMENTS AND RESULTS

A. Training Data Set Preparation and Preprocessing

The data used in this experiment are all derived from the real engine data. And these data are batch marking generated according to the CNR report. Generally, during the marking process of samples, the first 5 flight cycles before the failure point recorded in the CNR report can be marked as abnormal samples (positive class), and a certain number of samples will be regarded as normal samples (negative class). This experiment selected 70 CNR failure report records, and respectively select 1, 2, 3, 4, and 5 samples before the failure points as abnormal samples and select 9, 8, 7, 6, and 5 samples far before the failure points as normal samples. There are 5 samples between the selected normal and abnormal samples to reduce the systematic error caused by human marking.

TABLE II UNBALANCED SAMPLE SET

SN	Number of Normal Sample	Number of Anomaly Sample	Imbalanced Rate
sample set 1	350	350	1.0
sample set 2	420	280	1.5
sample set 3	490	210	2.3

sample set 4	560	140	4.0
sample set 5	630	70	9.0

70% of fault samples and 70% of normal samples are taken to form the training sample set, and the remaining samples are taken to form the test sample set.

For the training sample set generated above, a cubic polynomial feature is constructed to increase the degree of discrimination between samples. Subsequently, a standardized processing is carried out to eliminate the impact of dimensions between various performance parameters on the performance of the classifier.

B. *Classifier Selection and Overparameter Setting*

In this paper, SVM with gaussian radial basis kernel function is chosen as the classifier. And the super-parameter suitable for this data set was selected through 5-fold cross-validation. And C=100 and gamma=1 were used for the experiment.

C. *The Influence of Threshold on SVM Performance*

The above SVM was used for training on the training sets with different imbalance rates, and then the trained classifier was tested on the test set. The following figures show the performance of SVM with different decision thresholds, as shown in Fig. 5.

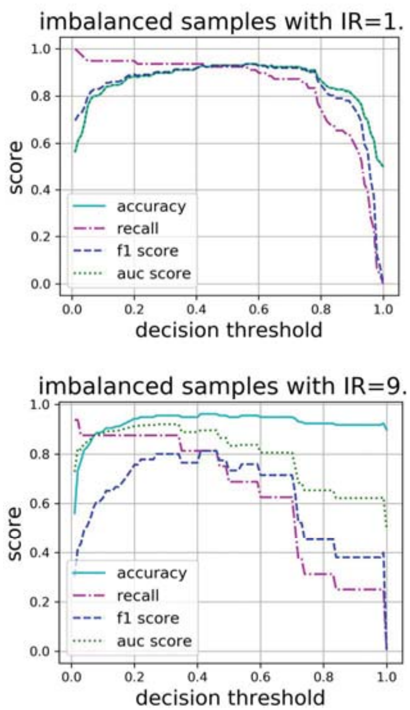

Figure 5. The influence of decision threshold on SVM

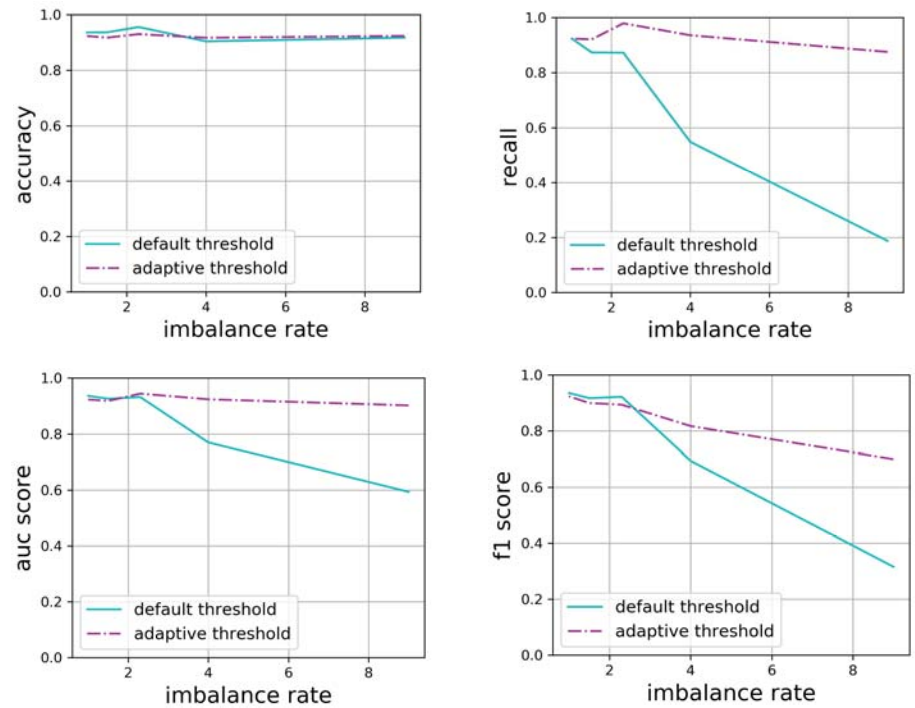

Figure 6. Comparison between ADT-SVM and SVM

It can be found in Fig. 5, for the balance samples (IR=1), when decision threshold is 0.5, the performance is close to the best. But for the imbalanced samples (IR=9), when the decision threshold with the same value, the model is not the best model. For unbalanced samples, reducing the decision threshold, in situations where *accuracy* changes very little, can greatly improve the *recall*, *f1 score* and *AUC score*. On the contrary, when increasing the threshold, them will reduce. It shows that the change of the decision threshold, can impact the performance of the model.

At the same time, it can be seen from Fig. 5 that when the threshold value is lower than a certain value, the *accuracy* of the model will show a cliff-like decline, which is not what we expected. Therefore, when calculating the adaptive threshold proposed above, $t_{\min}$ is introduced to avoid the occurrence of this extreme situation.

D. *The Performance of ADT-SVM*

By using the adaptive threshold proposed in this paper, the ideal decision threshold is generated directly through the imbalance rate (IR) of samples, which is used for the training of SVM. Then, the traditional SVM with default threshold is compared. The experimental results are shown in Fig. 6.

As shown in Fig. 6, with the increase of the imbalance rate, the *accuracy* of the default threshold SVM did not change significantly, but its *recall*, *f1 score* and *AUC score* all decreased significantly, which also verified the assertion that the imbalance of sample has a serious impact on the performance of classifiers.

At the same time, the ADT-SVM proposed in this paper can significantly improve the *recall* of the model under the condition that the *accuracy* remains unchanged. Meanwhile, *f1 score* and *AUC score* are also improved. This shows that the proposed method can be used to solve the problem of low recognition rate of abnormal samples in the aeroengine gas path anomaly detection process in the case of sample imbalance, and the effect is significant.

IV. CONCLUSION

The experiment results in Fig. 5 and Fig. 6 show that, for balanced samples, traditional SVM can show good enough performance. However, with the increase of sample imbalance rate, the *recall*, *f1 score* and *AUC score* are significantly decreased, although its *accuracy* can remain unchanged. This is because when faced with unbalanced training samples, the classifier tends to divide the sample into major class, namely the classification plane tends to favor minor class. Although the *accuracy* is able to remain unchanged, more minority samples are misclassified into the major class, which is the root cause of the significant decline in *recall*, *f1 score* and *AUC score* of abnormal samples. Obviously, such a classifier is not what we expect.

The experiment result in Fig. 5 shows that changing the decision threshold can significantly change the performance of the classifier. At the same time, reducing the decision threshold of minor class can significantly improve the *recall*, *f1 score* and *AUC score* under the condition of maintaining the same *accuracy*. This is crucial for the problem of abnormal detection. On the contrary, when the decision threshold is increased, its corresponding value will decrease.

The experiment result in Fig. 6 shows the superiority of proposed ADT-SVM, compared with traditional SVM with default decision threshold. The method directly generates the optimal decision threshold value by the imbalance rate of training samples, avoiding the one-dimensional search on the threshold space. And it also can achieve satisfactory results. The model can improve the *recall*, *f1 score* and *AUC score* under the condition that *accuracy* does not decrease.

Since the adaptive threshold method proposed in this paper is independent of the model training process, it can also be adapted to other classifiers, such as logistic regression, random forest, etc.

ACKNOWLEDGMENT

The authors would like to thank the editors and the anonymous reviewers for their valuable comments and constructive suggestions. This work is supported by the National Natural Science Foundation of China (No. U15332-02).

REFERENCES

[1] L. A. Urban, “Gas path analysis applied to turbine engine condition monitoring,” J. AIRCRAFT, vol. 10, pp. 400–406, July, 1973.

[2] Y. Hao, “Research on fault diagnosis and life prediction of civil aviation engine based on intelligent technology,” unpublished.

[3] S. M. Zhou, X. F. Zhou, and F. Gao, “An aircraft engine anomaly detection method based on online SVDD,” Metrology & Measurement Technology, vol. 2015, pp. 20–22, 2015.

[4] S. S. Tayarani-Bathaie, Z. N. S. Vanini, and K. Khorasani, “Dynamic neural network-based fault diagnosis of gas turbine engines,” Neurocomputing, vol. 125, pp. 153–165, 2014.

[5] N. V. Chawla, N. Japkowicz, and A. Kotcz, “Special issue on learning from imbalanced data sets,” ACM Sigkdd Explorations Newsletter, vol. 6, pp. 1–6, 2004.

[6] L. Cao, Y. Zhao, and C. Zhang, “Mining impact-targeted activity patterns in imbalanced data,” IEEE Transactions on Knowledge and Data Engineering, vol. 20, pp. 1053–1066, 2008.

[7] M. Bekkar, D. Alitouche, and T. Akrouf, “Imbalanced data learning approaches: a review,” International Journal of Data Mining & Knowledge Management Process, vol. 3, pp. 1–20, 2013.

[8] S. Kotsiantis, D. Kanellopoulos, and P. Pintelas, “Handling imbalanced datasets: a review,” GESTS International Transactions on Computer Science and Engineering, vol. 30, pp. 25–36, 2006.

[9] Y. Sun, A. K. C. Wong, and M. S. Kamel, “Classification of imbalanced data: a review,” International Journal of Pattern Recognition and Artificial Intelligence, vol. 23, pp. 687–719, 2009.

[10] B. C. Wallace, and I. J. Dahabreh, “Class probability estimates are unreliable for imbalanced data (and how to fix them),” Proceedings of the 12th IEEE International Conference on Data Mining, Brusels, Belgium, 2012, pp. 695–704.

[11] P. Cao, “Research on the classification method of unbalanced data,” unpublished.

[12] D. C. Wang, and H. Lin, “A SVM unbalance classification method and its application in fault diagnosis,” Journal of electrical machinery and control, vol. 16, pp. 48–52, 2012.

[13] T. Y. Liu, G. Z. Li, “Research on the problem of unbalanced data in fault diagnosis of rolling bearing,” Computer engineering and science, vol. 32, pp. 34–42, 2010.

[14] Z. F. Wang, J. L. Zarader, and S. Argentieri, “A novel aircraft engine fault diagnostic and prognostic system based on SVM,” IEEE International Conference on Condition Monitoring and Diagnosis, 2012, PP. 723–728.

[15] Q. H. Xu, and J. Shi, “Fault diagnosis of aeroengine based on support vector machine,” Journal of Aerospace Power, vol. 20, pp. 298–302, 2005.

[16] J. W. Lu, R. Wu, and H. S. Chang, “Fault diagnosis of aero-engine gas path improvement neural network,” Automation instrumentation, vol. 36, pp. 5–8, 2015.

[17] X. Wei, C. Lu, and C. Wang, “Applications of support vector machines to aeroengine fault diagnosis,” Journal of Aerospace Power, vol. 19, pp. 844–848, 2004.

[18] P. Rebentrost, M. Mohseni, and S. Lloyd, “Quantum support vector machine for big feature and big data classification,” Physical Review Letters, vol. 113, pp. 1–5, Sept. , 2013.

[19] Q. H. Xu, and J. Shi, “Aero-engine fault diagnosis based on support vector machine,” Journal of Aerospace Power, vol. 20, pp. 298–302, 2005.

[20] C. Y. Sun, L. Liu, and C. W. Liu, “Boosting-SVM algorithm based fault diagnosis in aeroengine,”. Journal of aeronautical dynamics, vol. 25, pp. 2584-2588, 2010.