A Fault Diagnosis Method of Engine Rotor Based on Random Forests

Qi Yao,Jian Wang*, Lu Yang, Haixia Su,Guigang Zhang
High-tech Innovation Center
Institute of Automation Chinese Academy of Sciences
Beijing China
{yaoqi2013,jian.wang,guigang.zhang}@ia.ac.cn

Abstract—**Rotor is the main part of the engine, the vibration fault is very common in the process of running, it must be monitored, checked, excluded in a timely manner for improving the reliability of engine and aircraft safety. This paper mainly studies four kinds of rotor fault, including unbalance, misalignment, surge, bearing failure. The frequency spectrum of the vibration signal of a rotor system is an important basis for rotor fault diagnosis, using the spectrum of rotor to build decision tree analysis is an important method for rotor fault detection. As the single decision tree's anti-interference ability is very poor, this paper presents an engine rotor fault diagnosis method based on Random Forests. Experimental results show that the accuracy of this diagnosis method is high, the failures can be diagnosed timely and effectively to keep the engine in normal operation. To evaluate the validity of Random Forests, a SVM classifier is trained for comparison. Compare with SVM, we obtain better classification in Random Forests algorithm. This result demonstrates that Random Forests algorithm is a valid method for engine rotor.**

Keywords-fault diagnosis; engine rotor; Random Forests; SVM

I. INTRODUCTION

As a core component of the aircraft, the aviation engine has a complicated structure and failures happen easily with different kinds of causes, which have a direct impact on flight safety. To make correct diagnosis and prediction, taking maintenance strategy in timely are great significant for flight safety and decreasing maintenance cost. There came up with a large number of engine monitoring diagnostic systems since 1970s. In recent years, the Prognostics and Health Management (PHM) based the concept----Condition Based Maintenance (CBM)was proposed, which focus on predicting the failure in order to making more proactive troubleshooting. The United States has put some advanced maintenance techniques, methods and procedures introduced into the practice, at the same time they have proposed the "Enhanced Condition Based Maintenance (CBM +)" based on CBM.

In general, aviation engine fault diagnosis methods can be divided into three categories: model-based method, knowledge-based method and data-based method. The model-based method can reconstruct the target Mathematics Model by forming the principle analysis and taking the fault diagnose through residues cooperation. This method is sufficiently required for high precision, which was mainly used in the relatively simple system, including parameter estimation, state estimation and the parity space method. The main feature of the Knowledge-based method is presenting the system by summarizing and transforming the experts' prior knowledge, and then monitoring the model fault in detail based on the "knowledge". Fault Tree Analysis and Expert System belongs to this type of method. Data-based method, which doesn't need to establish precise mathematical model, is the major method that has been promoted and applied in many field. Its main feature is that collecting a lot of useful information, taking the pre-processing data, then implementing further data analysis and mining, and diagnosing the system faults at final, it mainly includes statistical analysis and machine learning.

The common faults of aviation engine including vibration fault, gas path fault, wear failure, and so on. According to statistics, 70% of engine faults come from vibration faults, 80% of which come from the rotor vibration faults. The main rotor faults including unbalance, misalignment, surge, bearing failure and so on. The feature of the rotor system fault is that multiple faults have same signs, which caused by varied reasons. These features increase the difficulty of the rotor system status monitoring and fault diagnosis. How to timely detect the fault information early or even before the rotor vibration faults occur and reduce the severity of the problem, and it has a special significance for the entire engine health management. Essentially, fault diagnosis technology is a pattern classification and recognition problem, namely that the operation of the system state is divided into two types of normal and abnormal, and that abnormal sample belongs to which fault exactly, is also a classification problem. The traditional classification algorithms include decision tree, logistic regression, Bayes, neural networks and so on. The algorithm of this paper is Random Forests (RF), which is a combination of classification and has been widely used in many areas. In this paper, RF is applied to the fault diagnosis of engine rotor and compared with SVM. The algorithm has superior performance of classification. The experimental results are very good.

II. RELATED WORKS

Engine is complicated in structure. Unbalance, misalignment, surging, rolling bearing faults are common vibration faults in rotor system. and there is a strong correlation between faults, so it's difficult to build the simulation model. In

978-1-5090-0382-2/16/$31.00 ©2016 IEEE

addition, noise and the measured signal are similar and there is bias in the measured signal, so it is difficult to distinguish them. Thus, there are many problems in the rotor vibration fault diagnosis. At present, the research on the fault diagnosis of rotor system is mainly focused on the following three aspects:

A. Research on model of rotor system

Research on dynamic model of rotor system is mainly a physical or mathematical model from the angle of system mechanism, failure cause and effect of the state of the rotor system. The reference [1] established early crack rotor vibration analysis model, put forward to early crack spectrum analysis of the vibration signals by using $1\frac{1}{2}$ dimension spectrum based on the propagation mechanism of rotor crack. The reference [2] built coupling dynamic model with rotor containing rubbing faults, rolling bearings, supporting and casing to do the experiment of rubbing fault. The reference [3] and the reference [4] studied on the relevant fault based on the simulation platform of ZT-3.

B. Fault feature extraction

Fault feature extraction is composed of fault feature selection and structure. Feature selection mainly includes the original fault feature and the second feature. The original fault feature mainly includes the system temperature, pressure, vibration signal and other parameters signal information, and the second feature is to get the new features with transformation and combination of the original parameters. When the characteristic dimension and amount of data are large, the principal component analysis is used for dimensionality reduction. The reference [5] studied the feature extraction and quantitative evaluation method of vibration signal by using information entropy analysis method. A method of fault diagnosis of turbine rotor based on isometric feature mapping (ISOMAP) and support vector machine (SVM) is proposed in [6]. The vectors form low dimension data space are extracted, which produced by ISOMAP from high dimension space of vibration signal, and consider them as feature vectors.

C. Research on fault diagnosis method

The traditional fault diagnosis is usually completed by the field of experienced personnel, in recent years, artificial intelligence technology has been rapid development, and it has been widely used in the field of fault diagnosis of rotor. The expert system is widely used, and the expert knowledge were pooled and transformed into the information which computer can identify to support dynamic expansion, simulate experts in the field to fault reasoning, complete intelligent fault diagnosis. It can greatly reduce manpower and material inputs. The reference [7] designed the fault diagnosis expert system based on fault tree and rule. The reference [8] used fuzzy neural network to diagnose faults of aero engine. The reference [9] researched on how to design and implementation aero-engine fault diagnosis expert system knowledge base based on fuzzy inference and then design the eliminating strategies rules competition based on the fuzzy inference rule, achieve real-time update and online fault diagnosis. The method based on artificial neural network and D-S evidence theory is applied to aero-engine fault diagnosis in [10]. The [11] used Kalman filter and genetic algorithm to diagnose faults.

III. RANDOM RORESTS

Decision tree is a classical classification algorithm, which has been widely used for decades in many applications. The main variants of decision tree include ID3, C4.5 and Classification And Regression Tree (CART). The difference between these variants lies in the way they choose feature to split at each node. ID3 chooses the feature which has the most entropy gain in the remaining features at each node. Applying this strategy, the tree tends to choose the feature which has the most attributes first. This may be not wise strategy in many situations. To overcome this, the C4.5 algorithm use entropy gain ratio to choose feature at each node. CART adopts a different strategy to select features, which select feature which has the minimum Gini index or the least square error. Decision tree is proven to have infinite VC dimension, which means it can fit any data. While in practical applications, decision tree tend to over fit due to the noise in the data, i.e. the tree can have very small train error while have large test error, which cannot be used as a qualified model to predict in real applications. So pruning is necessary to lower the complexity of the tree. Pruning is the process to delete the node which has low entropy gain. To some extent, pruning can lower the rate of over-fitting, but it still not applicable in big data sets. In recent years, Random Forests as a variant of decision tree has been proven to be powerful in big data classification and regression tasks. It is less likely to be over-fitting and easy to conduct. The Random Forests is also based on decision tree, which has multiple trees to make a forest. For a single tree in the Random Forests, it uses the CART strategy to build a tree. The flowchart of Random Forests is shown in Fig 1. Given the date set S and the feature dimension M, the concrete steps to build a Random Forests is given below:

A. Training

- Generate C sub training data sets $S_1, S_2,\ldots S_C$ from the original data set S based on the re-sampling technique Bootstrap. The subset is required to have the same size.
- For each subset $S_k(k=1,\ldots,K)$, Randomly select m features from the M features to form the data set to train the decision tree T_k.

B. Testing

- For a test sample X, let it go through all the decision trees in the Random Forests and get the corresponding result
- Adopt the majority voting strategy to get the final result of the test sample X.

IV. ROTOR FAULT DIAGNOSIS

This paper mainly studies four kinds of rotor fault, including unbalance, misalignment, surge, bearing failure.

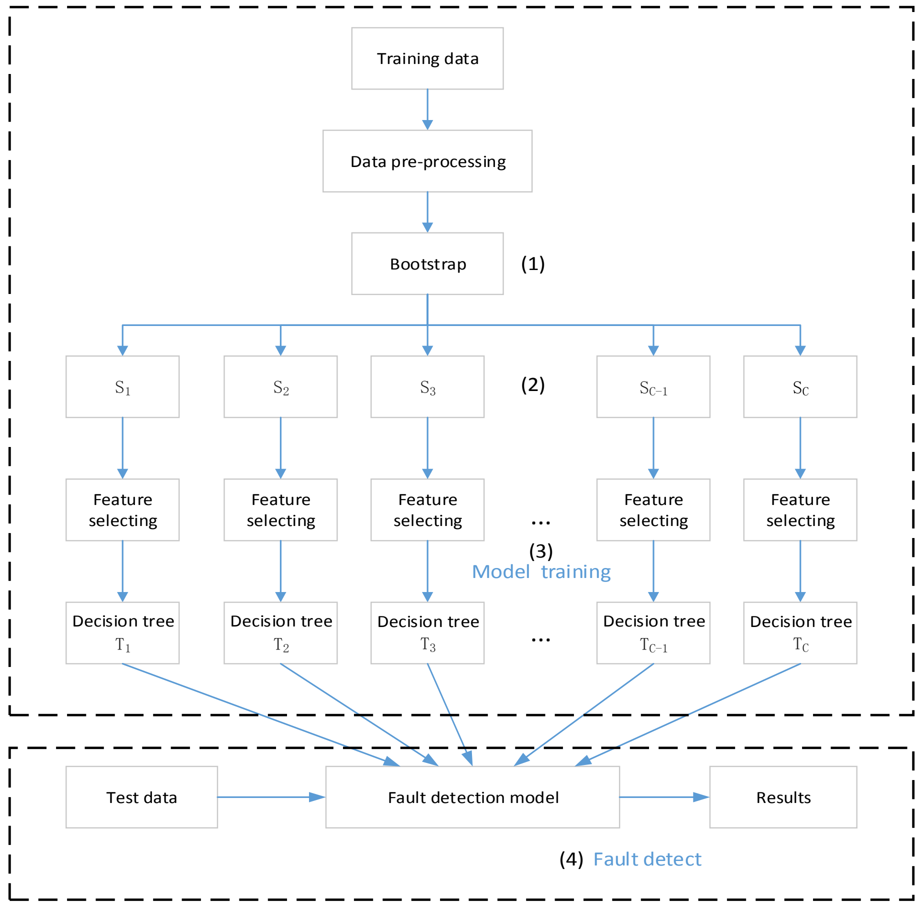

Figure 1. flowchart of Random Forests

A. feature selection

In the fault diagnosis of engine rotor, the spectrum diagnosis signal is used as one of the important basis, this paper chooses $0.2f_0$, $0.25f_0$, $0.43f_0$, $0.5f_0$, $0.75f_0$, f_0, $2f_0$, $3f_0$, $4f_0$, $5f_0$ as the features (f0 is the work frequency of the rotor). These features proved to have a strong correlation with four kinds of faults above.

B. Data preprocessing

There exists 300 training samples and 40 test samples, each sample contains 11 data, the first column to the 10th column corresponds to the value of ten features mentioned above, the 11st column is corresponding label of the sample, 1 means unbalance, 2 means misalignment, 3 means surge, 4means bearing failure, 5 means normal.

C. Simulation test

The Open Source Toolkit--Random Forests-matlab, developed by Abhishek Jaiantilal, is used to construct the model in this paper. Two key parameters can be adjusted in the training function of this model, one is the number of decision trees, the other is number of features per tree (the default value equals the square root of M, M is the total number of features). Due to features of sample used are few, we uses the default value when training this algorithm, this paper focus on exploring performance of the algorithm when number of the decision trees grow. The number of decision trees range between 5 and 100, with an increase of 5 each step, then use the training set for training and use the test set for verification.

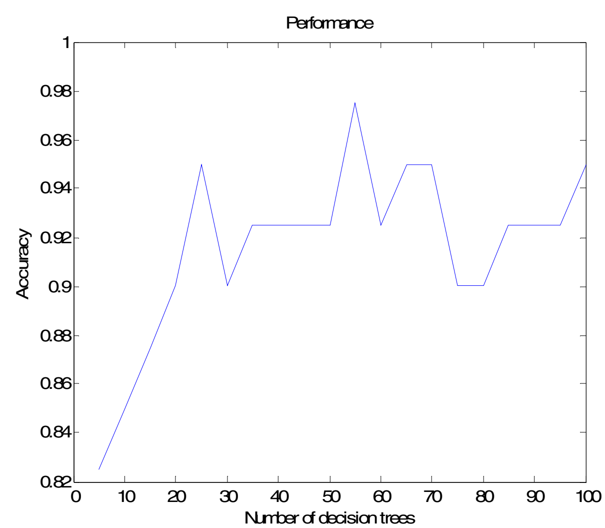

Figure 2. Performance of Random Forests

The performance of “Accuracy--Number of decision trees” diagram is shown in Figure 2, it can be seen from the figure that the model has the best performance when the number of decision trees is 55, it also shows that the accuracy of the model is roughly positively related to the number of decision trees. Due to the randomness of the algorithm, the performance of the algorithm and decision tree is not a linear relationship, in other words, not simply more trees, higher accuracy.

To evaluate the validity of Random Forests, a SVM classifier is trained for comparison. SVM is widely used for its advantages in solving small sample, nonlinear and high dimensional pattern recognition problems. The LIBSVM package is chosen for classification in this paper. As the dimension of feature is limited, linear kernel function may fail to distinguish different class of samples, we use SVM model with radial basis function as kernel to predict. In this case, parameters *g* and *c* have an effect on the final classification results. Cross-validation is used to seek out the best *g* and *c*. From Figure.3, we obtain the best accuracy when $c = 0.25$, $g = 0.75786$ and the accuracy for train set is 91.25%.Then predict the test set with the trained model and the accuracy for test set is 85% (34/40). The accuracy in test set is lower than in train set, which means that over-fitting may occur in SVM to some extent. Compared with SVM, we obtain better classification performance in Random Forests algorithm. This result demonstrates that Random Forests algorithm is a valid diagnosis method for four kinds of rotor fault mentioned above.

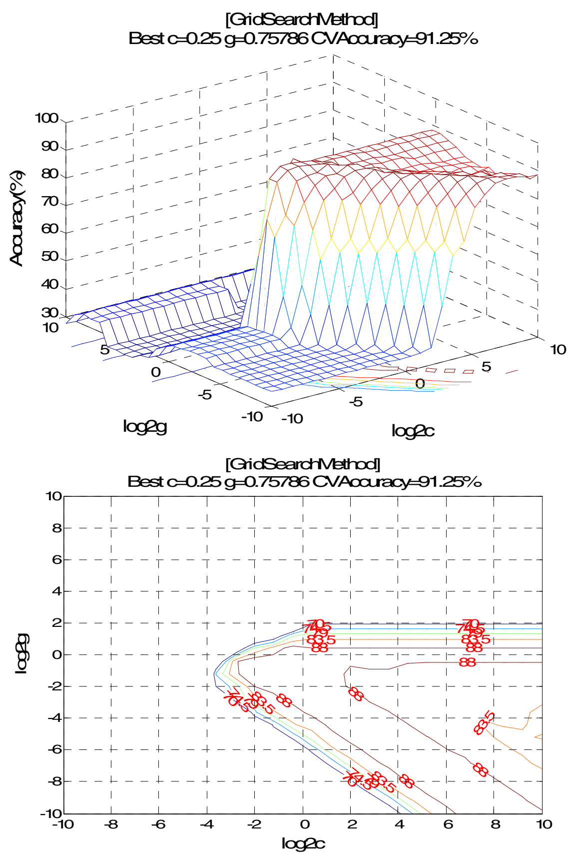

Figure 3. Parameter selecting

V. CONCLUSION

Firstly, a brief introduction of fault diagnosis for engine rotor is given in this paper, secondly, the related research of engine rotor fault diagnosis is reviewed, and then the Random Forests algorithm is described in detail, finally, a comparative study using SVM and Random Forests algorithm is demonstrated. From the results, we can get the following conclusion: the Random Forests is effective for fault diagnosis of engine rotor. In the future, it will be focused on how to improve the shortcomings and generalization ability of the method.

ACKNOWLEDGEMENT

This research was supported by: Special Project for Civil-Aircraft, MIIT.

REFERENCES

[1] Wang Yanfeng, Zhu Jing, Teng Guangrong. 1½-dimension spectrum analysis on early cracked fault characters on aero engine rotors[J]. Journalof vibration and shock , 2015, 34:88-93.

[2] CHEN Guo, LI Chenggang, WANG Deyou. Nonlinear dynamic analysis and experiment verification of rubbing faults of rotor-ball bearing-support-stator Coupling system for aero-engine[J]. Journal of vibration and shock , 2015, 34: 88-93.

[3] Zhou Yang. Rotor Faults Diagnosis Based on Vibration Signal Analysis. [D]., Nanjing University of Aeronautics and Astronautics , 2014.

[4] Wang Dongyun. STUDY ON FAULTDIAGNOSIS METHOD FOR ROTOR-BEARING. [D], Yanshan University, 2012

[5] YANG Juan. Extraction Method Research of Quantification Feature For Rotor Fault Signal. [D].Lanzhou University of Technology, 2010。

[6] SUN Bin, XUE Guangxin. Method of Rotor Fault Diagnosis Based on Isometric Feature Mapping and Support Vector Machine[J]. Journal of Mechanical Engineering, 2012, 48(9):1-10

[7] Hou An-hua, Qin Hong-lei. Fault Diagnosis Expert System Structure based on Fault Tree and Rule Based Resoning[J]. Micro computer information, 2008, 24(19):191-193.

[8] Cai Kai-long, Yan Bin-zheng, Xie Shou-sheng. A Method for Fault Diagnoses of Aero-engines Based on Fuzzy Neural Network[J]. MECHANICAL SCIENCE AND TECHNOLOGY, 2004, 23(1), 96-99.

[9] Yan Yang.Aero-engine knowledge base construction for the fault diagnosis expert system. [D].Shengyang Aerospace University.2013

[10] Chen Tian, Sun Jian-guo, Hao ying. Neural Network and Dempster-Shafter Theory Based Fault Diagnosis for Aero-engine Gas Path[J]. ACTA AERONAUTICA ET ASTRONAUTICA SINICA, 2006, 27(6):1104-1108.

[11] Yuan Chunfei, Yao Hua. Development of Kamlan filter and genetic algorithm for aero-engine performance diagnostics[J]. JOURNAL OF PROPULS ION TECHNOLOGY, 2007, 28(1):9-13.