Fast Fault Diagnosis System Based on Data Mining AR Algorithm

Yahan Yu[1,3], Juan Du[2]

1. Institute of Automation
Chinese Academy of Sciences
Beijing, China
yuyahan2020@ia.ac.cn

Guanghao Ren[1], Yao Tan[2]

2. Chengdu Aircraft Industrial (Group) Co., LTD.
Aviation Industry Corporation of China, LTD.
Chengdu, China

Jian Wang[1], Guigang Zhang[1]

3. School of Artificial Intelligence, University of Chinese Academy of Sciences
Beijing, China

2021 Global Reliability and Prognostics and Health Management (PHM-Nanjing) | 978-1-6654-0131-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/PHM-Nanjing52125.2021.9612883

***Abstract*—Aero-mechanical parts are an important part of the aircraft, and the maintenance of their failures also consumes a lot of manpower and financial resources. Therefore, the fault diagnosis research of aero-mechanical parts is of great significance for ensuring the safety of human life and reducing economic losses. With the development of fault diagnosis technology, the monitoring data is becoming more and more abundant and complex. The traditional methods of processing and analyzing the monitoring data have become more difficult, and it is difficult to establish accurate mathematical models. Therefore, the rapid diagnosis method of aviation machinery parts Become the research focus of fault diagnosis.**

This paper constructs a rapid fault diagnosis system for the construction of aviation machinery parts. Based on the input of past cases, new cases, literature cases, and book knowledge, the case library is refined and the graph library and rule term library are added. AR algorithm is used to mine and obtain Useful association rules between the decision attributes (failure mode, failure mechanism, failure reason, etc.) of the failure information in the database and the basic attributes (basic information other than the decision attributes), to achieve the purpose of assisting failure analysts in rapid fault diagnosis.

Keywords- fault diagnosis system; PHM; AR algorithm; rapid fault diagnosis

I. INTRODUCTION

During the take-off, flight and landing of an aircraft, various faults are prone to occur. Quick diagnosis of these faults and maintenance suggestions are of great significance for improving safety and reducing operating costs. Traditional engine fault diagnosis is mainly regular maintenance, lack of pertinence for troubleshooting, and has high requirements for the professionalism of maintenance personnel, which consumes a lot of manpower and material resources. To solve the above-mentioned problems, this paper has researched fault diagnosis of aero engine. With the development of modern science and technology, the requirements for safety and reliability of aviation equipment are increasing, and the structure, function, load-bearing and service environment of mechanical components are becoming more and more complex, and the subsequent failure of mechanical parts is inevitable. Because the failure of mechanical components is often the result of the coupling of multiple factors, the corresponding analysis work is becoming more and more complex. To quickly and accurately perform fault diagnosis and analysis, relying solely on knowledgeable failure analysis experts is far from meeting engineering needs and fails. Analysts often rely on personal knowledge and experience to analyze the causes of failures. Therefore, the application of advanced, in-depth and systematic rapid failure analysis and diagnosis technology has become inevitable. As a type of high-value and complex equipment, the operational status of aviation machinery parts is closely related to the safety of aircraft flights. More importantly, due to long-term operation in harsh environments such as high temperature and high pressure, aviation components are prone to various failures. Therefore, it is necessary to carry out effective fault diagnosis on aviation machinery parts to ensure that they can operate safely and reliably. As a key technology for the health management of aviation machinery parts, fault diagnosis uses the monitored component parameters to accurately locate and diagnose the fault, which not only helps the staff to grasp the working status of the aircraft in a timely and rapid manner, but also provides the timing for maintenance. Provide strong support for forecasting, maintenance plan formulation, and maintenance cost estimation.

Today's computer technology, database technology and various data mining algorithms are developing rapidly, providing effective methods and means for obtaining and developing efficient fault diagnosis auxiliary tools. Solidify the data (including basic information, attributes, pictures and other data information), knowledge and experience of mechanical component failure analysis into the computer, and automatically mine to obtain potentially useful information from a large amount of incomplete, fuzzy, and random data And knowledge, not only can assist failure analysts in rapid fault diagnosis and analysis, distinguish and identify failure modes and failure

978-1-6654-0131-9/21/$31.00 ©2021 IEEE

causes, make up for the lack of personal knowledge and experience, broaden analysis and research capabilities and eliminate human factor interference, but also can deal with a large number of Invalidation knowledge is inherited and standardized management and utilization. In the field of aviation manufacturing, ensuring flight safety and reducing maintenance costs have always been the goals pursued by many manufacturers. For a long time, the aviation industry has continuously explored various safety and maintenance technologies. These technologies have formed Prognosis and Health Management (PHM) technology in the process of long-term development, evolution and integration. PHM technology was proposed by NASA and the US Department of Defense [1]. Its idea is to use various sensing devices throughout the aircraft to collect real-time status information of aircraft systems, diagnose faults and damages that have occurred, and predict remaining systems. Life, according to fault diagnosis and prediction results, take appropriate mitigation measures and maintenance strategies to avoid accidents and reduce maintenance costs [2]; PHM technology has experienced a rapid development with the support of many scientific research projects and funds and is widely used On multiple types of civil and military aircraft and spacecraft, their role in improving safety and reducing operating costs has also been recognized by the aviation industry [3], but these applications only achieve the integration or integration of part of the health management of the monitoring system. Integration of some health management functions [4]. Domestic research on health management technology is at the technical tracking level. Literature [5] have reviewed PHM technology, and literature [6] have made preliminary discussions on the engineering application of this technology. Up to now, there are no successful commercial cases.

Due to the limitations of design, environment, and operating characteristics, the fault diagnosis data of different systems are aggregated, and cross-system data analysis will cause serious uncertainty. Therefore, data sharing between different enterprises is poor. Data mining requires a huge sample size. The larger the number of samples and the more random the sampling, the stronger the generalization ability of the knowledge base obtained by data mining. Association rules are one of the main algorithms in current data mining research. It focuses on determining the association rules between different attributes in the data, accurately finding the source of the fault, and has a high accuracy rate. Because many faults have parallel, causal and other relationships when the machine When a component fails during operation, the association rules can also be used to provide early warning of possible failures to prevent failures from deteriorating, which is predictive. To sum up, the construction of a rapid diagnostic system for aviation mechanical component failures mainly includes three parts: (1) complete database establishment; (2) data mining and knowledge base formation; (3) rapid fault diagnosis logical relationship establishment [7].

Association rules are a rule-based machine learning method used to find hidden relationships between items from a data set. It can be used to find the connection between shopping basket data to facilitate cross-selling; it can be used for text mining; it can also be used in other fields such as bioinformatics, medical diagnosis, earth science, etc., to find some interesting connections. Association rules are used to describe the relationship between two or more things. It uses one or more things to predict other things and can obtain the connection between valuable data from a large amount of data [8]. Association rule mining is one of the most active research methods in data mining. The typical association rule discovery problem is to analyze the market basket data (Market Basket) in the supermarket. Analyze the customer's buying habits by discovering the relationship between the different products that customers put in the shopping basket.

To solve the above problems, this paper develops a rapid fault diagnosis system based on the data mining AR algorithm. This system is an intelligent program system with a large number of expert knowledge in related fields. It uses artificial intelligence (AI) technology to reason based on domain knowledge, simulates the process of human decision making to solve complex problems, and provide an automatic diagnosis and efficient means of processing knowledge data. By storing failure analysis knowledge and experience knowledge, it is of great significance to the structured management of the failure analysis process in the future. In addition, this paper develops software modules to automatically obtain strong association rules and knowledge bases between fault attributes, which can quickly diagnose and analyze faults, and have a predictive function to a certain extent.

II. Construction of Failure Database of Aviation Mechanical Components

Firstly, a complete aeromechanical component failure database is established. A complete database with sufficient samples is the basis and necessary condition for data mining. The database provides the required knowledge for the inference engine in the inference process, and at the same time provides a storage place for the intermediate results and conclusion information of the inference. It is mainly used to store the three types of information of original information, intermediate information and conclusion information [9].

(1) Original information. It mainly includes the user's prior knowledge in the field, such as failure modes, the normal range of important parameters, the composition of each system, the data files used for diagnosis, and the user management information used to manage the expert system.

(2) Intermediate information. It mainly includes intermediate information generated based on the original information processing design for the convenience of the reasoning and interpretation functions of the inference engine, such as file information tables, parameter data tables, etc.

(3) Conclusion information. Mainly the expert system completes fault diagnosis conclusion information, such as fault reasoning process description, historical fault database, etc., which is easy for users to understand, and can also be used for later report analysis.

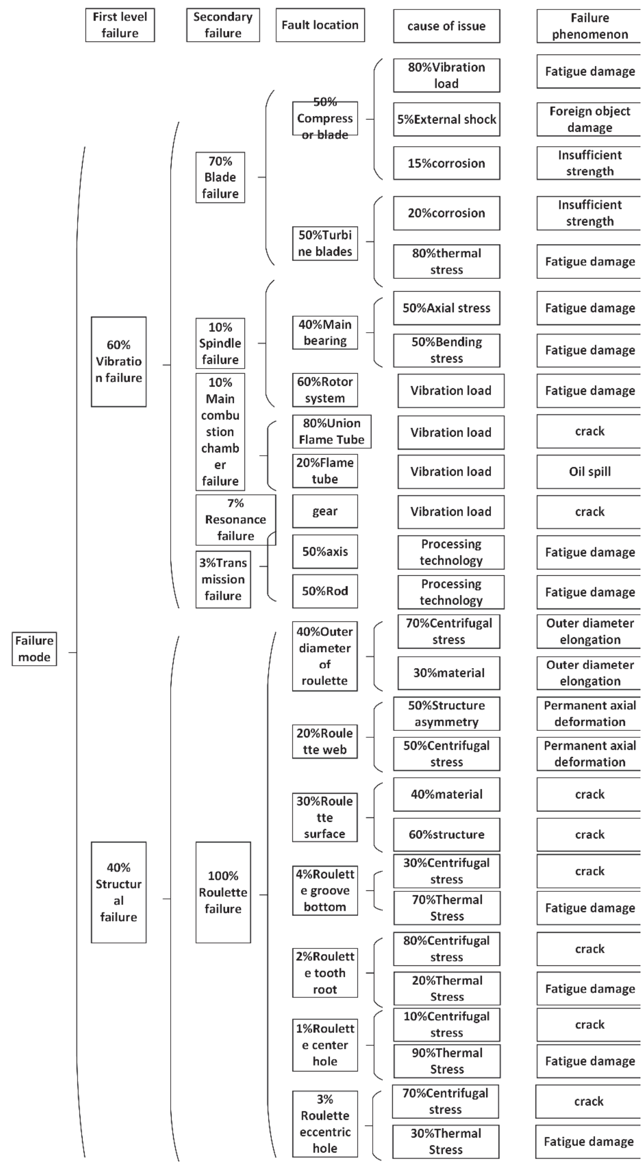
Based on the original failure analysis case library, this paper determines the content of sub-modules at all levels. The definition of fault attributes includes the decision-making attributes of the fault including failure mode, failure reason, and failure mechanism. The basic attributes of the fault include model, field/in-plant, material, service time, service environment, part drawing number, whether repetitive failure, and cause impact Wait. Moreover, the rule termbase means that the standardization of terminology in the failure analysis report is the key to increasing the accuracy of subsequent data mining. To make the attribute description in the failure analysis case accurate and standardized, and prevent ambiguity and inconsistent words Establish a unified rule terminology database to unify terminology for part name, load type, load size, material type, heat treatment process, moulding process, environment, appearance characteristics, etc.

Among all mechanical component failures, the vibration failures accounted for 60%, and the structural failures accounted for 40% [10]. And because rotating machinery is mainly composed of a rotating shaft and various disc-shaped parts (such as impellers, gears, couplings, etc.) installed and fixed on the rotating shaft, its common faults mainly include the following: bending, unbalance, misalignment, Lateral cracks on the shaft, loose connections and rubbing [11]. Gear is the main transmission component and one of the most prone to failure. The requirements for gear quality and transmission ratio are very high, and the factors that cause gear failures are also increasing. Gears can be damaged due to manufacturing errors, improper assembly, or use under inappropriate conditions. Common faults include gear wear, gear misalignment, gear broken teeth, etc. Rolling bearings are important supporting parts and are vulnerable parts. During operation, they may be damaged due to reasons such as assembly problems, lubrication problems, moisture and foreign matter intrusion, corrosion and overload problems. The construction idea of the fault database is shown in figure 1.

III. PRINCIPLE OF AR ALGORITHM

The Apriori algorithm is a commonly used algorithm for mining data association rules. It is used to find data sets that frequently appear in data values. Finding out the patterns of these sets helps us make some decisions. For example, in the common supermarket shopping data set, or the online shopping data set of the e-commerce, if we find a data set that appears frequently, then for the supermarket, we can optimize the placement of the product, for the e-commerce, we can optimize the location of the product The location of the warehouse achieves the purpose of saving costs and increasing economic benefits [12]. For the Apriori algorithm, we use the support degree as our criterion for judging frequent itemsets. The goal of the Apriori algorithm is to find the largest frequent set of K items. There are two meanings here. First, we need to find frequent sets that meet the support standard. But there may be many such frequent sets [13]. The second level means that we need to find the largest number of frequent sets.

There are three commonly used evaluation criteria for frequent itemsets: support, confidence and promotion. Support is the proportion of the number of times that several related data appear in the data set to the total data set. In other words, the probability of occurrence of several data associations.



If we have two data X and Y that we want to analyze the relevance, the corresponding support is:

$$Support(X,Y) = P(XY) = \frac{number(XY)}{num(AllSamples)} \quad (1)$$

Figure 1. The structure of the fault database

Generally speaking, data with high support does not necessarily constitute frequent itemsets, but data with too low support does not necessarily constitute frequent itemsets. Confidence degree reflects the probability of one data appearing, another data appearing [14], or the conditional probability of the data. If we have two data X and Y that we want to analyze the correlation, the confidence level of X to Y is:

$$Confidence(X \Leftarrow Y) = P(XY) = P(XY) / P(Y) \quad (2)$$

The degree of lift represents the ratio of the probability of including Y and the probability of X at the same time[15], and the probability of the overall occurrence of X, namely:

$$\begin{aligned} Lift(X \Leftarrow Y) &= P(X \mid Y) / P(X) \\ &= Confidence(X \Leftarrow Y) / P(X) \end{aligned} \tag{3}$$

The promotion body firstly considers the association relationship between X and Y. If the promotion is greater than 1, $X \Leftarrow Y$ is a valid strong association rule, and if the promotion is less than or equal to 1, $X \Leftarrow Y$ is an invalid strong association rule. In a special case, if X and Y are independent, there is $Lift(X \Leftarrow Y) = 1$, because at this time $P(X \mid Y) = P(X)$.

Generally speaking, to select a frequent data set in a data set, you need to customize the evaluation criteria. The most commonly used evaluation criterion is to use custom support, or a combination of custom support and confidence.

The Apriori algorithm adopts an iterative method, first search for candidate 1 item set and corresponding support, pruning to remove 1 item set lower than support, and get frequent 1 item set. Then connect the remaining frequent 1-item sets to obtain candidate frequent 2-item sets, filter out candidate frequent 2-items sets that are lower than support, and get the true frequent binomial sets, and so on, iterate until it fails Until the frequent k+1 item set is found, the corresponding frequent k item set is the output result of the algorithm. It can be seen that this algorithm is still very concise. The i-th iterative process includes scanning to calculate the support of candidate frequent i item sets, pruning to obtain the true frequent i item sets and connecting to generate candidate frequent i+1 itemsets in three steps.

The Apriori algorithm scans the data set in each iteration, so when the data set is large and there are many types of data, the efficiency of the algorithm is very low. The Apriori algorithm is a very classic algorithm for mining frequent itemsets. Many algorithms are based on the Apriori algorithm, including FP-Tree, GSP, CBA, etc. These algorithms use the idea of the April algorithm, but the algorithm has been improved, and the data mining efficiency is better. Therefore, it is generally seldom to use the April algorithm to mine data directly, but understanding the April algorithm is a prerequisite for understanding other April algorithms. At the same time, the algorithm itself is not complicated, so it is worth studying.

IV. Implementation of Fault Diagnosis Based on AR Algorithm

After establishing a complete database, it is necessary to perform support calculations on different candidate sets by retrieving the type and attributes of each item in the database, and at the same time establish a candidate set that meets the minimum support. Then the frequent itemsets are calculated, and the frequent itemsets are mostly obtained by the Apriori algorithm, and the frequent itemsets obtained by the AR algorithm are used to generate the association rules between the decision attributes and the basic attributes. Figure 2 shows the construction framework of the rapid fault diagnosis system.

It is mainly divided into the following steps:

Data item determination: Select the second, third, and fourth level decision attributes shown in Figure 2 as the candidate data items, and select the main failure modes of the company's parts and failure mechanisms as the main case training sample objects according to database statistics;

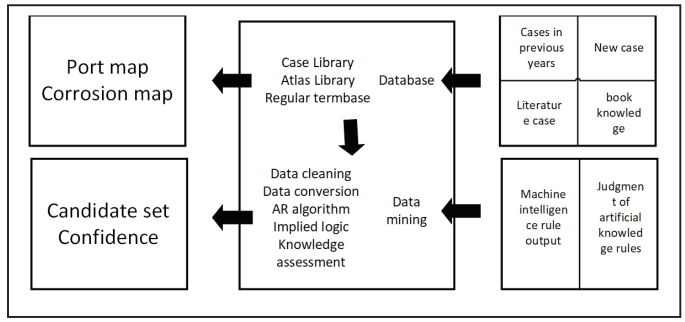


Figure 2. Construction framework of rapid fault diagnosis system

Case training: For all cases in the database, use the AR algorithm to train to obtain strong association rules. The association rule is an implication of the form A=>B. A is called the antecedent data item of the association rule, and B is called the association rule's The subsequent data item, there is a dependency relationship between A and B.

Suppose a set of data items is $I = \{I_1, I_2, I_3, ..., I_k\}$, such as $\left|I\right| = n$, there will be 2 candidate sets. The efficient calculation of the AR algorithm needs to generate a small candidate item set. To improve the efficiency of the AR algorithm, it is necessary to set constraints based on the failure analysis knowledge and experience, and do not generate or calculate those candidate item sets that are unlikely to become frequent itemsets.

The formula for calculating the item set support is $D_{supp}(A) = \|A\| / |I|$, and $\|A\|$ represents the number of A data items included in the item set I. The calculation formula for the support of association rules is $D_{supp}(A => B) = D_{supp}(AUB) = \|AB\| / |I|$.

Frequent set $L = \{A \mid D_{supp}(A) \geq min\ D_{supp}(A)\}$, and then calculate the confidence level of frequent set L $D_{conf}(A => B) = \|AB\| / \|A\|$. Automatically filter and obtain strong association rules through minimum support and minimum confidence.

TABLE I. Association Rule Result (Frequent 1-Itemsets)

ASSOCIATION RULE RESULT	
frequent 1-itemsets	***support***
Structural failure	0.4

ASSOCIATION RULE RESULT	
frequent 1-itemsets	***support***
Vibration failure	0.6
Vibration load	0.306
Roulette failure	0.4
Fatigue damage	0.434
Blade failure	0.42

TABLE II. ASSOCIATION RULE RESULT (FREQUENT 2-ITEMSETS)

ASSOCIATION RULE RESULT	
frequent 2-itemsets	***support***
Vibration load, Vibration failure	0.306
Blade failure, Vibration failure	0.42
Fatigue damage, Vibration failure	0.414
Roulette failure, Vibration failure	0.4
Fatigue damage, Blade failure	0.336

TABLE III. ASSOCIATION RULE RESULT (FREQUENT 3-ITEMSETS)

ASSOCIATION RULE RESULT	
frequent 3-itemsets	***support***
Fatigue damage, Blade failure, Vibration failure	0.336

TABLE IV. ASSOCIATION RULE RESULT

ASSOCIATION RULE RESULT	
rules	***conf***
[' Vibration failure ']=>[' Vibration load ']	0.51
[' Vibration load ']=>[' Vibration failure ']	1
[' Vibration failure ']=>[' Blade failure ']	0.7
[' Blade failure ']=>[' Vibration failure ']	1
[' Vibration failure ']=>[' Fatigue damage ']	0.69
[' Fatigue damage ']=>[' Vibration failure ']	0.95
[' Structural failure ']=>[' Roulette failure ']	1
[' Roulette failure ']=>[' Structural failure ']	1
[' Fatigue damage ']=>[' Blade failure ']	0.77
[' Blade failure ']=>[' Fatigue damage ']	0.8
[' Vibration failure ']=>[' Fatigue damage ', ' Blade failure ']	0.56
[' Fatigue damage ']=>[' Blade failure ', ' Vibration failure ']	0.77
[' Blade failure ']=>[' Fatigue damage ', ' Vibration failure ']	0.8
[' Blade failure ', ' Vibration failure ']=>[' Fatigue damage ']	0.8
[' Fatigue damage ', ' Vibration failure ']=>[' Blade failure ']	0.81
[' Fatigue damage ', ' Blade failure ']=>[' Vibration failure ']	1

V. SYSTEM DEVELOPMENT

A. Import self-built failure mode database

Select the fault.sql file in the AR_PHM folder, which is a temporarily generated fault mode database. First, turn on the MySQL service of the system, then enter the MySQL password in the MySQL command-line client, and then import the MySQL database fault, enter: source <path to the sql file>. After the import is successful, "fault" should appear in the database.

B. Operating system

Run the main.py file in the AR_PHM folder, and the system interface appears. Click "Generate" to generate frequent sets and association rules of failure modes from the database or the local database test.sql. When entering a certain fault feature, all possible diagnosis results and corresponding confidence levels will appear. The result is shown in figure3.

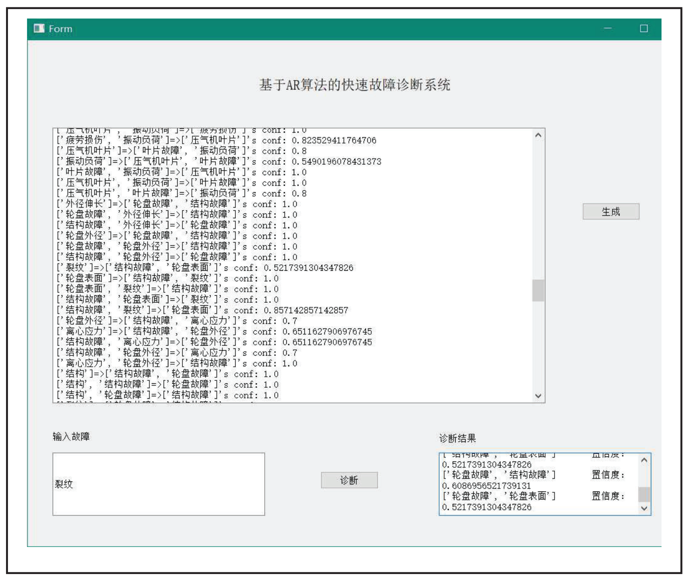


Figure 3. System operation interface

VI. CONCLUSION

This paper constructs a rapid fault diagnosis system for the construction of aviation machinery parts. According to the input of case knowledge, the case library is refined and the graph library and the rule term library are added. The AR algorithm is used to mine the useful association rules between the decision attributes and the basic attributes of the failure information in the database, so as to achieve the purpose of assisting failure analysts in rapid fault diagnosis.

REFERENCES

[1] Zuniga F A, Maclise D C, Romano D J, Jize N N, Wysocki P F, Lawrence D P. Integrated Systems Health Management for Exploration Systems[C]// Proceedings of 1st Space Exploration Conference: Continuing the Voyage of Discovery. Orlando, Florida, USA: AIAA, 2005: 1- 16.

[2] Paris D E, Trevino L C, Watson M D. IVHM Framework for Intelligent Integration for Vehicle Health Management[R]. NASA, Washington D C: NASA, 2005.

[3] Reveley M S, Briggs J L, Thomas M A, Evans J K, Jones S M. An Examination of Commercial Aviation Accidents and Incidents Related to Integrated Vehicle Health Management[C]// Proceedings of 9th AIAA Aviation Technology, Integration, and Operations Conference (ATIO). Hilton Head, South Carolina, USA: AIAA, 2009: 1-14

[4] Gorinevsky D, Smotrich A, Mah R, Srivastava A, Keller K, Felke T. Open architecture for integrated vehicle health management[C]// Proceedings of AIAA Infotech @ Aerospace Conference. Atlanta, GA, USA: AIAA, 2010: 1-8.

[5] Wheeler K R, Kurtoglu T, Poll S D. A survey of health management user objectives related to diagnostic and prognostic metrics[C]// Proceedings of the ASME 2009 International Design Engineering Technical Conference & Computers and Information in Engineering Conference. San Diego, California, USA: 2009: 1-11.

[6] Ofsthun S. Integrated vehicle health management for aerospace platform[J]. IEEE Instruments & Measurement Magazine, 2002, 5 (3): 21-24.

[7] Goldcare: Simplifying the business of airplane ownership[EB/OL].[2006-8-10].

[8] AIRMAN-real-time aircraft health monitoring[EB/OL].[2006-12-13].

[9] Jennions L K. Integrated Vehicle Health Management - Perspective on an Emgerging[M]. Warrendale, PA USA: SAE International, 2011: 126-128.

[10] Larder B D. Helicopter HUM/FDR: Benefit and Developments. Stewart Hughes Limited White Paper[R]. 1-8.

[11] HUMS: Health And Usage Monitoring Systems[EB/OL].[2006-2-1].

[12] Hess A, Calvello G, Dabney T. PHM a key enabler for the JSF autonomic logistics support concept[C]// Proceeding of IEEE Aerospace Conference, 2004, 6: 3543-3550.

[13] Hess A. Prognostics, from the need to reality – from the fleet users and PHM system designer/developers perspectives[C]// Proceedings of IEEE Aerospace Conference, 2002, 6: 2791-2907.

[14] Hess A. The joint strike fighter (JSF) PHM concept: potential impact on aging aircraft problems[C]// Proceedings of IEEE Aerospace Conference, 2002, 6: 3021-3026.

[15] Mangasarian O L, Musicant D R. Lagrangian support vector machines[J]. Journal of Machine Learning Research.2001.1:161-177.