Overview of Research on Health Assessment and Fault Prediction of Complex Equipment Driven by Big Data

Can-can Jin¹, Hong Zhang², Jun Liu¹ and Yang Lu¹
¹Nanjing University Of Finance & Economics, Nanjing 210023, P. R. China
²AVIC Beijing Keeven Aviation Instrument Co., Ltd, Beijing 100086, P. R. China
*Corresponding author: Can-can Jin, Tel: +0086 15261883576,892702739@qq. com

Abstract. Health evaluation is an important content of health management for complex equipment, and fault prediction is a key link to realize condition-based maintenance. Based on the literature review, this paper gives a feasible modeling method: The health status of complex equipment is expressed as time series by constructing forward and backward LSTM model, and the health evaluation model is constructed and trained based on the deep learning platform TensorFlow. The improved particle filter constraint algorithm is used to optimize LSTM neural network which is to generate PF-LSTM prediction model. The research on health evaluation and fault prediction of complex equipment based on machine learning discussed in this paper puts forward new requirements and development direction for machine learning of complex equipment, and is of great significance for practical implementation of lean intelligent manufacturing, improvement of stability and economic benefits of complex equipment, and reduction of operation and maintenance costs.

1. Introduction
Complex equipment is a giant system, including complex electromechanical system, control equipment and aero-engine, etc., with characteristics of high cost, large scale, high technology, complex structure principle, complex manufacturing process, complex test maintenance and complex application environment [1-2]. Maintenance and Operation problems of complex equipment include the concept of MRO2 (Maintenance, Repair, Overhaul & Operation) which is planning the Maintenance, Repair, Overhaul and Operation of complex equipment as a whole, so as to optimize maintenance and operating costs for complex equipment over the full life cycle [3].

《The 13th five-year Plan for Technological Innovation in the Field of Advanced Manufacturing Technology》clearly states that the key technology of measurement and control product safety can be used as the key of developmental research, such as Prognostics and Health Management (PHM) and so on. In the recent decades, equipment Diagnostics, Prognostics and Health Management (DPHM) aims to establish predictive maintenance strategy, improve system safety/readiness integrity assessment and fail safe regulation [4]. It has a positive effect on increasing the safety benefits and economic benefits during the normal use period and daily support period of complex equipment, and provides an important guarantee for the reliable safety of major equipment in service. The predictive analysis of DPHM is the core of system performance degradation and residual life prediction.

Along with the modern industrial process, the production manufacture system, the aerospace system and so on modernized equipment informationalized level unceasing enhancement, the complexity and high degree of uncertainty of the system increases accordingly, and monitoring the
process of complex equipment group size amount is larger, observation point is more, the frequency is too high and observation period is long, so when the vast amounts of data to monitoring system, equipment health monitoring in the field of "big data" era is coming\cite{5}. Complex industrial processes are characterized by multivariable, strong coupling, strong nonlinearity, large time delay and frequent change of production boundary conditions. There is a complex nonlinear correlation between subsystems in complex equipment, and the influence of different working conditions on the performance decline of complex equipment is also nonlinear. Only a model with good nonlinear mapping ability can be used to describe the complex nonlinear mapping relationship between complex equipment and its state parameters.

2. New progress in machine learning research and application
With the development of complex equipment automation, flexibility, intelligence and greening, more and more attention has been paid to the study of health evaluation and failure prediction based on machine learning. Machine learning is usually divided into supervised learning and unsupervised learning. The knowledge network diagram of machine learning algorithm is shown in figure 1. Unsupervised learning usually includes clustering, association analysis, recommendation algorithm and density estimation.

Based on the above problems, the article adopts the method of machine learning for complex equipment health status evaluation and fault prediction of combing and analysis of the traditional single, on the basis of condition monitoring and prediction, with the help of advanced condition monitoring method and reliable data processing model and method, from the implied rich system health information mining state of huge amounts of data, the maximum fusion of multi-source information in a more accurate evaluation of complex equipment health status, and thus build a prediction model of the global model selection and parameter optimization function. It is great engineering significance and economic value to promote the transfer of large-scale equipment manufacturing industry in China from the low-end of the value chain to the high-end, to guarantee the safe and reliable operation of major equipment, and to reduce the operation cost, and to improve the core competitiveness of major industries in China.
3. Research status analysis at home and abroad

3.1 Connotation and research status of complex equipment health status evaluation

The health status evaluation of complex equipment is based on the fusion of sensor measurement data, manual measurement data, historical data and so on. Various evaluation algorithms are used to make a correct evaluation of the health status of equipment, providing a basis for the maintenance decision of equipment. The research methods for health status evaluation generally start from two perspectives: model-driven method and data-driven method. With the analysis of the management process of the equipment health management system based on the PHM architecture, the development of health management technology has experienced different developmental processes in the research of system level and component level.

The combined weighting method and similarity evaluation method are based on the common analytical angle of data driven. Among them, the combined weighting method takes the expert scoring method as the core, which has strong subjectivity. The similarity evaluation method takes the similarity correlation degree of test samples and existing samples as the core and normalizes the data to form the health index. The main methods of calculating similarity include Mahalanobis distance, cosine similarity, correlation and Minkowski distance. OMIDI Ali etc. analyzed and studied the state maintenance (CBM) and PHM, and proposed the disadvantages of traditional CBM and wavelet neural network multi-mode filter methods [6].
3. 2 Research connotation and research status of complex equipment health failure prediction

Failure prediction is the prediction of the residual life of the system or the risk of completing a certain function, which includes the tracking and trend analysis of the health state of the system. Along with the modernization level enhances unceasingly, the complexity and uncertainty of the high-end equipment also increase, from the perspective of method system to the existing complex equipment health status change trend prediction method is divided into model driven and data driven: model driven method based on the working condition of engine working principles, material properties are established by factors as cost modeling is generally high. The latter can be directly based on the health data of the engine for modeling, and modeling cost is relatively small.

3. 2. 1 Model-driven approach

This method generally requires the establishment of a mathematical model based on the working principle of equipment, and the parameter estimation technology is used to estimate the model parameters and then predict the trend. The most commonly used parameter estimation method is Kalman filter and its derivative methods. However, the linear model is difficult to describe the characteristics of complex nonlinear, and the solution of nonlinear model is too complex. Secondly, the model is difficult to describe the differences among different individuals in practice. In addition, the number of known parameters obtained by state monitoring may be far less than the number of unknown parameters, which increase the difficulty of parameter estimation of the model.

3. 2. 2 A method based on statistics and time series analysis

In order to solve the problems of large amount of data, high dimension, fast update and noise in time series, Yongwei Yu et al. constructed the time series input vector of the model using the principle of phase space reconstruction, and trained the model with BPTT method to improve the motion accuracy. Yuan Xu et al. introduced multivariate time-delay analysis into high sensitivity local kernel principal component analysis. The correlation and time-delay of process variables were obtained by mutual information estimation and BIC. Time series prediction was carried out by the BP network.

3. 2. 3 A method based on intelligent learning model

Cloud computing fault detection methods based on supervised learning neglect noise data processing, updating training samples and identifying unknown types of faults, which affect the accuracy of cloud computing fault detection. Liu Chengcheng et al. improved the fuzzy kNN according to the fault feature weight and hierarchical detection, and determined the fault detection results based on the maximum membership degree self-learning [7]. Jiansheng Cao et al. according to the data and the improved particle swarm optimization algorithm and BP neural network, the prediction model was established to predict [8].

3. 2. 4 Deep learning method based on big data background

With the development of artificial neural networks, the concept of in-depth learning has emerged. Science published an extremely deep neural network model, to achieve the scientific dimensionality reduction of data. Compared with the traditional principal component analysis method, the model is more effective. The relationship between generalization ability and capacity of model is the main problem of machine learning theory, judging whether a model has generalization ability mainly comes from the data set size of model training. Deep structure model is composed of multi-level non-linear transformation, its capacity is larger than that of shallow model. Artificial intelligence is the key to machine learning. Deep models show strong expressive power in highly nonlinear relationships and complex functions to acquire deep knowledge in large data environments, and gradually show a strong ability in identifying, judging and predicting intelligent behaviour [9-10].

The deep neural network, represented by the real-time recursive learning algorithm neural network has been widely used in the analysis of time-series large data [11-12]. At present, the deep neural network used in large data analysis is mainly feed-forward neural network, which is good at extracting
the correlation of static data and is suitable for data application scenarios based on classification. However, it has limited ability to extract the temporal characteristics of data. Deep neural network is a kind of recurrent neural network with feedback connection. The evolution of the network state with time is the essential attribute of the network. It couples the "time parameters" and is more suitable for extracting the temporal characteristics of data to predict large data.

As a kind of deep learning algorithm, long-term and short-term memory network (LSTM) has increasingly become an important means of time series prediction. Jianpeng Zhao et al. based on the long-term and short-term memory network to predict the state of rotating machinery, and compared with the prediction results of support vector regression machine model, it is proved that the long-term and short-term memory network can achieve better than support vector regression in the state prediction of rotating machinery [13]. Li Ning combines regression algorithm and neural network algorithm in the field of machine learning to study IGBT fault prediction technology, and uses TensorFlow to build a model to train and predict the degradation data of turn-off peak voltage [14].

When there are frames with long time intervals in the sequence, it is difficult to find the relationship between them by traditional cyclic neural network. With the passage of time, the subsequent hidden layer nodes and output layer nodes are less and less affected by the input of each frame. LSTM network can add state to the hidden layer, so that the network has the ability to remember the input of a long time ago in the sequence. TensorFlow platform has high programming speed and data analysis efficiency in large data environment. Therefore, the LSTM model based on TensorFlow platform is analyzed and explored in the following state evaluation and prediction model [15].

4. Research on health status evaluation and fault prediction of complex equipment based on machine learning and large data

4.1 LSTM model

The LSTM model adopted in this paper consists of two input nodes, three hidden nodes and one output node. The input sequence is \(a^1, a^2, \ldots, a^t, \ldots, a^T\), the hidden layer sequence is \(h^1, h^2, \ldots, h^t, \ldots, h^T\), and the output sequence is \(c^1, c^2, \ldots, c^t, \ldots, c^T\). The calculation methods are shown in formulas 3.1 and 3.2.

\[
b^t = Q(R_h \times [a^t, h^{t-1}] + b_h) \quad (3.1)
\]

\[
c^t = P(R_{ho} \times h^t + b_o) \quad (3.2)
\]

\(Q\) and \(P\) represent the activation function of the hidden layer and the output layer, \(R_h\) represents the weight matrix from the input layer and the last hidden layer to the current hidden layer, \(R_{ho}\) represents the weight matrix from the hidden layer to the output layer, \(b_h\) and \(b_o\) represent the skewness vector of the hidden layer and the output layer, respectively. \(t\) denotes the number of frames in the sequence, the current input of the network is \(a^t\), the updated state output is \(J^t\), the forgetting gate is \(f^t\), the input gate is \(i^t\) and the output gate is \(o^t\). The expressions of sigmoid activation function, three new gates and hidden layer output \(h^t\) and status update \(J^t\) are shown in formula 3.3-3.7.

\[
f^t = \mu(R_f \times [a^t, h^{t-1}] + b_f) \quad (3.3)
\]

\[
i^t = \mu(R_i \times [a^t, h^{t-1}] + b_i) \quad (3.4)
\]

\[
o^t = \mu(R_o \times [a^t, h^{t-1}] + b_o) \quad (3.5)
\]
4. Health classification model for complex equipment

Multivariate time series is used as input and health status category as output. The initial output layer of LSTM network is removed, the output of hidden layer \( h^t \) is introduced into the average pool layer, and the vector \( h \) without time information is obtained, that is \( h = \frac{1}{T} \sum_{t=1}^{T} h^t \). In order to remove the output layer and add the average pool layer network, the Bi-mLSTM model is introduced in this paper. Its input sequence is input in reverse order of time. \( h^b \) represents the output of Bi-mLSTM and constructs a logical regression model. The logical regression layer takes the new vector of the output \( h \) and \( h^b \) of mLSTM and Bi-mLSTM as input, and the classification results are obtained through the soft Max layer.

4. 3 Implementation on TensorFlow Platform

InputData is an example of health status of complex equipment. Data-in InputData by dropout operation for prevent model over-fitting. Data information enters two MultiRNNeCells, namely the hidden layer, which is implemented by LSTMCell. Sequence output \( h \) and \( h^b \) by ReduceMean operation. \( h \) and \( h^b \) output the errors of real calss lable by cross entropy error function calculation model and the logical regression of manual construction and activation function softmax.

4. 4 The LSTM training model base on partical filtering algorithm for complex equipment failure prediction

The process of optimizing the function of particle filter algorithm is actually a time-varying dynamic system. The iterations are the discrete time and the state of the system is the optimal solution for each iteration. LSTM model is designed to make the error between prediction and actual value infinitely close. That is to say, the objective function is to minimize \( \sum_{t=1}^{T} (c_t^f - c_t^c)^2 \). Among it, the actual value is \( c_t^c \) whereas the output value of the system is \( c_t^f \). So the measurement equation of the system can be written as:

\[
\text{fitness} = \sum_{t=1}^{T} (c_t^f - c_t^c)^2
\]  

The basic steps of LSTM parameter optimization based on particle filter algorithm are as follows:

Firstly, the definition domain of the initial state is determined by the error range, and the population initialization is generated randomly in the definition domain. The total number of particles is \( N \) and the total number of iterations is \( d \). \( N \) particles are randomly selected from the initial particle set. According to the measurement equation, the coincidence of all particle \( c_t^f \) is evaluated and give the corresponding optimal value \( c_t^c \). If the particle is in conformity to the constraint, the particle weight is 0; if the constraint condition is satisfied, the weight is calculated according to the observed value \( c_t^c \), and the particle weight is normalized. The formula for calculating the effective number of particles is
that is the weight of particles. If \( \hat{N}_j < N_j \), where \( N_j \) is the set threshold, the resampling process is performed according to distance comparison and optimal combination strategy, and obtain a new set of supporting particles \( \hat{a}_k \), and estimated the state and variance of the system. According to the number of iteration steps \( k \) to determine whether the termination condition of the algorithm is satisfied?If the termination condition is not satisfied, the algorithm is re-iterated. If the termination condition is satisfied, return to the output value \( c \) of the optimal particle and the system.

4. 5 Building the LSTM prediction model based on particle filtering

A LSTM cyclic neural network model is built using TensorFlow deep learning framework in this paper. The specific prediction steps are as follows:

Sample data processing with range scaling:

\[
A_s = \frac{A - A_{\min}}{A_{\max} - A_{\min}}
\]

Among them, \( A_{\max} \) and \( A_{\min} \) represent the maximum and minimum of the sample number respectively. Set the network input as \( m \) variables per batch as input \( A = \{a_i | i = 1, 2, \cdots, m \} \), \( n \) variables as output \( C = \{c_i | i = 1, 2, \cdots, n \} \), and the number of hidden layer units is determined by experiment. Parameters are determined by the improved particle filtering algorithm, which makes the model converge to the global optimum and predicts the network flow. Two error analysis methods are used to verify the prediction accuracy, namely root-mean-square error \( X \) and average fractional error \( Y \) to judge the error of prediction model. As shown in formulas 3. 11 and 3. 12:

\[
X(c_i, \hat{c}_i) = \sqrt{\frac{1}{N} \sum_{j=1}^{N} (c_j - \hat{c}_j)^2}
\]

\[
Y(c_i, \hat{c}_i) = \frac{1}{N} \sum_{j=1}^{N} \left| \frac{\hat{c}_j - c_i}{c_j} \right| \times 100\%
\]

Among it, \( c_i \) is the actual value of the sequence sample, \( \hat{c}_i \) is the predicted value, and \( N \) is the total number of samples.

5. Review of research status and development trend

From the above analysis, it can be concluded that the health status evaluation and prediction of complex equipment has been very rich at home and abroad. With the development of MRO2 technology systems, the health status evaluation and prediction technology of complex equipment have been studied and applied. The problems and development trend of health status assessment and fault prediction for complex equipment are summarized as follows.

(1) The changes of health status of complex equipment are often contained in various structured, unstructured and semi-structured data resources, such as BIT, failure cases, maintenance plans and technical data. Therefore, the data of state parameters of complex equipment are complex. The model is not clear and changeable, and the description of the non-linear relationship between complex equipment and state parameters is inaccurate. To solve these problems, the establishment of models and algorithms for health status assessment has become the most important task for system health status classification and estimation.

(2) The health status data of complex equipment is time-series. But the standard cyclic neural network model can not solve the problem of long-term dependence. As the number of layers deepens,
the gradient tends to disappear. Whereas the change of parameters in complex equipment will lead to the change of health status in a long period of time, so it needs more time ahead to predict and analyze the fault in order to deal with the change of health status in this period.

(3) At present, most of the research are aimed at the level of "equipment" or "components". With the enhancement of the breadth and depth of equipment-level condition monitoring of complex equipment, the amount of data collected in the process of condition monitoring increases exponentially. When the traditional shallow model is not good at monitoring and diagnosing massive data and generalization, the ability to extract temporal features of data is limited. The typical feature of health condition is "inadequate ability".

(4) In the large data environment, the large increase of data size makes data and algorithm structure complex, which brings resistance to the establishment of models and data analysis for health status evaluation and fault prediction of complex equipment. Therefore, the efficiency of learning framework construction and algorithm integration needs to be improved urgently. Simplified programming and intelligent computing put forward new requirements for machine learning process of complex equipment.

6. Summary and prospect
This paper gives the classification results of these methods by exploring the modeling idea, model style, reasoning methods and their advantages and disadvantages of machine learning. Then it summarizes the current modeling research status of health status assessment and fault prediction for complex equipment at home and abroad, and summarizes the application objects and research results of these methods. A health status classification and fault prediction model of complex equipment based on LSTM neural network is adopted. A classification model is built on TensorFlow platform by expressing the health status of complex equipment as time series and the gradient descent method is used to train the model automatically to evaluate the health status of complex equipment. The prediction model of LSTM neural network (PF-LSTM) is optimized by improved particle filter constraint algorithm (PF), which accurately grasps the changing trend of complex equipment health status. Finally, this paper put forward the development trend of health status evaluation and fault prediction of complex equipment by summarizing the problems existing in health status evaluation and fault prediction of complex equipment, and considering the characteristics and the requirements of application environment.

References
[1] BO Wenhua, ZOU Hongfu, LV Defeng, YU kai. Maintenance Optimization Method of Civil Aircraft Based on Delay Time Mode [J]. Journal of Nanjing University of Aeronautics & Astronautics, 2015,47(06): 890-896(in Chinese).
[2] MA Xiaojun, REN Shuhong, ZOU Hongfu, WEN Zhenghua. Prediction method of aero-engine life on wing based on LS-SVM algorithm and performance reliability[J]. Journal of Traffic and Transportation Engineering,2015,15(03):92-100(in Chinese).
[3] MO Xinnong. MRO2 drives the shift from manufacturing to productive services [J]. Manufacture Information Engineering of China, 2012, (18): 18-20(in Chinese).
[4] HU Niao-qing, HU Lei, CHEN Ling, GAO Ming. The Status quo, Future, and Challenge of Equipment Health Management[J]. National Defense Science & Technology,2015,36(1):10--15(in Chinese).
[5] YU Chengyuan. Research and Application of Equipment Health and Failure Analysis Based on Industrial Big Data[D]. University of Chinese Academy of Sciences (Shenyang Institute of Computing Technology,Chinese Academy of Sciences), 2017(in Chinese).
[6] OMIDI Ali,LIU Shujie. Overview of the Importance of Intelligent Approaches on Machinery Faults Diagnosis and Prediction Based on Prognostic and Health Management / Condition-Based Maintenance[J]. Journal of Donghua University (Eng. Ed),2019,35(3):270-273
[7] YU Yongwei, DU Liuqing, YI Xiaobo, Chen Gang. Prediction of CNC Machine Tools’ Motion Precision Based on Sequential Deep Learning[J]. Transactions of Chinese Society for Agricultural Machinery, 2018, 12(09): 6-17 (in Chinese).

[8] CAO Jiansheng, LI Kaiqun. Life Prediction of nMOSFET Based on MPSO BP Neural-Network[J]. Chinese Journal of Electron Devices, 2018, 41(5): 1119-1123 (in Chinese).

[9] XU R Z, Geng X F, Zhou F Y. A Short Term Load Forecasting Based on Bagging-ELM Algorithm [C]. Proceedings of the 2012 International Conference on Information Technology and Software Engineering, Beijing, 2013. Springer Berlin Heidelberg: 507-514

[10] GAO F, KOU P, GAO L, et al. Boosting Regression Methods Based on a Geometric Conversion Approach: Using SVMs Base Learners [J]. Neural computing, 2013, 113: 67-87

[11] HUANG Yihua. Research Progress on Big Data Machine Learning System [J]. Big Data Research, 2015, 1(1): 28-47 (in Chinese).

[12] QIAN Yanling, LIU Fujun, YIN Xiaohu. Constructing a “Cloud of Logistics Support” To Realize “Knowledge Centered” Logistics Support [J]. National Defense Science & Technology, 2015, 36(1): 26-30 (in Chinese).

[13] ZHAO Jianpeng, ZHOU Jun. State Prognosis of Rotary Machines Based on Long/Short Term Memory Recurrent Network [J]. Noise and Vibration Control. NOISE AND VIBRATION CONTROL, 2017, 37(4): 155-159 (in Chinese).

[14] LI Ning. Study on Prediction Technology of IGBT Module Based on Machine Learning Algorithm [D]. Beijing: BEIJING JIAOTONG UNIVERSITY, 2017 (in Chinese).

[15] YANG Yu, ZHANG Wei. Human action classification based on LSTM neural network on TensorFlow [J]. Intelligent Computer and Applications, 2017, 7(5): 41-45 (in Chinese).

[16] LI Xiaolin, WU Teng. Efficient network traffic prediction method based on PF-LSTM network [J]. Application Research of Computers, 2019, 36(12) (in Chinese).