Survey on Advanced Equipment Fault Diagnosis and Warning Based on Big Data Technique

Miao Wang\textsuperscript{a}, Zhenming Zhang\textsuperscript{b}, Kai Li\textsuperscript{*}, Can Si\textsuperscript{c} and Long Li\textsuperscript{d}

School of Mechanical Engineering, Northwestern Polytechnical University, Xi’an 710072, China

*Corresponding author e-mail: taglikai@163.com, \textsuperscript{a}wmiao@mail.nwpu.edu.cn, \textsuperscript{b}zhangzm@nwpu.edu.cn, \textsuperscript{c}m17809164727@163.com, \textsuperscript{d}jaylee@mail.nwpu.edu.cn

Abstract. Faults are main problem of Mechanical equipment. In general, in order to acquire the massive fault information of equipment, a large number of monitoring points need to be established on the equipment, and the number of sensors for the equipment is large, since equipment fault diagnosis enter the era of big data. The successful application of big data technology in track circuit and power equipment system indicates that the big data of equipment contains important information that reveals the evolution and nature of faults. This paper briefly introduces the development of fault diagnosis technology from three aspects: data acquisition, intelligent fault diagnosis approach and remote fault diagnosis. In addition, This paper expounds the development history of fault diagnosis technology, analyzes the challenges of intelligent fault diagnosis in the era of big data, and discusses the development trend of equipment fault diagnosis according to the existing foundation and challenges. Finally, the development trend of equipment fault diagnosis and early warning are pointed based on existing research.

1. Introduction

Advanced equipment manufacturing industry is one of the strategic emerging industries, high and new technology as the leading, including aviation equipment, rail transportation equipment and intelligent manufacturing equipment and so on. The industry is the core of the industrial chain, promoting industrial transformation and upgrading, and is also an important symbol of measuring international competitiveness. The next few years will be a vital period for high-end equipment manufacturing industry. Facing the huge market demand space, it is imperative to accelerate the development of high-end equipment manufacturing \cite{1,2}. High-end manufacturing equipment has the characteristics of complex structure, long manufacturing cycle, huge scale, large volume and complex operating environment, and its engineering design, quality control in manufacturing process have higher requirements, especially fault diagnosis of equipment.

In the previous research, a large number of sensors and data transmission lines were deployed in the rail circuit transportation system, and the electrical service department could regularly obtain more detailed daily operation data of rail circuit, forming the big data containing various operation conditions \cite{3}. In power system, researchers developed a safety assessment and fault diagnosis system based on data mining platform PDMiner that based on cloud computing, to improve the data
processing efficiency and accurately extract the information of fault types and risk region, having much higher value in power grid renovation and planning [4]. In the mechanical equipment industry, large equipment group size, multiple measuring points for a single equipment, high data sampling frequency and long monitoring duration enable massive data to be acquired, thus pushing the field of high-end equipment fault diagnosis into the era of big data [5]. Big data of high-end equipment is characterized by large amount of data and low value density, but it contains various information of equipment fault evolution mechanism and essence, which promote the equipment fault diagnosis technique on the existing basis.

2. Research on Fault Diagnosis Technology

In order to adapt to the rapid development of industry, fault diagnosis and early warning technology of mechanical equipment need to be continuously reformed. Equipment fault diagnosis and early warning mainly through various types of sensors by virtue of intelligent diagnosis method to obtain fault information and make timely fault warning. Different types of equipment and different components within the same equipment show different fault characteristics which often presents a complex nonlinear mapping relationship. A fault feature may be caused by multiple reasons, which brings great challenges to the field of equipment fault diagnosis, as shown in Fig.1 Feature-Cause map.

![Feature-Cause map](image)

Fig 1. Feature-Cause map

In the 1960s, fault diagnosis technology started its worldwide implementation in the United States. Over the past few decades, workers judge whether the equipment fails or not, mainly relying on the accumulated experience of long time and the observation of the equipment [6]. With the development of fault diagnosis technology for a long time, researchers engaged in fault diagnosis of mechanical equipment have conducted in-depth research on signal acquisition, fault characteristics, fault treatment technology and other aspects, and fruitful research results have been achieved.

In recent decades, due to the increasing complexity of equipment structure, fault diagnosis methods have changed from the signal processing and analytical model to knowledge-based diagnosis, and the research has also become integrated, precise and intelligent. Intelligent fault diagnosis technology mainly includes three aspects, the flow path is presented in Fig.2.

Fault signal acquisition: the multi-source signals from equipment are collected to detect the running state of the equipment.

Fault feature acquisition: based on the collected signals, fault features are extracted to identify fault types.

Fault identification and early warning: the model established by intelligent method is used for fault identification and early warning.
2.1. Information Collection

It is the premise of mechanical fault diagnosis to use advanced sensing technology to acquire response signal to represent the running state of mechanical equipment. Yang et al. [7] established a multi-source information acquisition system for diesel engine, which collects vibration and noise signals. Based on multi-sensor information acquisition, the multi-source information of different time and space is fully utilized, thus improving the reliability of decision-making. In [8], in order to realize the mechanical equipment condition monitoring based on wireless sensor network, problems such as high-speed synchronous acquisition, real-time reliable transmission and energy supply must be solved, and low cost, low power consumption, networking and multi-function will be the future direction. Zhang et al. [9] proposed an online fault diagnosis and early warning model for wind turbines combining Storm real-time streaming data processing and Spark batch processing technology.

Furthermore, it is proposed to monitor every equipment by adding data acquisition module or data capture card, but the method has some requirements for equipment interface[10]. In [11], a hybrid system of radio frequency identification and wireless sensor network (HRW) was proposed, which combined the traditional radio frequency identification system with the wireless sensor network system to achieve efficient collection of equipment data. H. Shen et al.[12] for lab large wireless sensor networks, a reliable and efficient data acquisition combined clustering routing design method is proposed. A large number of heterogeneous data types are generated during the operation of complex equipment. In [13], The mathematical models of measurable variables and unmeasurable variables are established by using soft measurement technology, and finally the unmeasurable state information of coal mill is obtained.

2.2. Intelligent diagnostic method

In [14], taking a certain advanced equipment integrating optical, mechanical and electrical components as the research object, an intelligent diagnosis method based on BP neural network and fuzzy theory was proposed to solve the problems of monitoring equipment aging, lack of independent learning and renewal function. BP network diagnosis has obvious effect on the system fault diagnosis of quantitative data, but it can't deal with qualitative data. MA et al [15]. study intelligent fault diagnosis methods on dealing with qualitative data of BP neural network equipment, switching qualitative data into quantitative data by applying some pretreatment technology like coding of fault diagnosis. Based on the knowledge base, the automatic and intelligent fault diagnosis is realized, which greatly improves the efficiency of equipment fault diagnosis, and expands the application scope of BP neural network diagnosis methods. Generally, the fault diagnosis system adopts the service machine mode, which requires on-site diagnosis. Due to the sharing of technical diagnosis information, this diagnosis support level is limited.

In the face of big data of equipment that large quantity, multi-source, complex and rapid growth, Zhang [9] designed the Topology structure of stream data processing, introduced Spark, and used the elastic distributed dataset (RDD) programming model to realize fault diagnosis and early warning of equipment. A new adaptive multi-core combined RVM prediction method was proposed and applied to the residual life prediction of mechanical equipment after simulation [16]. In [17], the limitations of time-based equipment maintenance methods and the advantages of predictive or online maintenance techniques for identifying the onset of equipment failure were discussed.

Real-time initiative maintenance strategy (MRAM) was proposed in [18], set up the equipment maintenance array. The information of each element is fed back into the array in real time, and the array threshold was set. When the data value in the array exceeds the threshold, the maintenance
equipment will be called for maintenance. If it cannot be solved, the relevant maintenance personnel will be notified for maintenance. When the equipment operation returns to normal, the state value in the array will be set to normal (0). MRAM has the characteristics of active fault reporting, which is superior to the traditional mode that hierarchical fault reporting. In MRAM algorithm, as shown in Fig3 Pseudocode of the data processing, N represent the number of items. SEERA et al. [19] applied a hybrid intelligent model composed of fuzzy minimum-maximum neural network and random forest model in motor fault diagnosis. Zhang Lin et al. [20] studied and analyzed intelligent equipment fault diagnosis method based on rule reasoning to monitor the equipment health status, aiming at solve practical problems such as the complexity of a certain missile equipment system, lack of maintenance experience and difficulty in fault diagnosis. SHEN et al. [21] proposed a general multi-class analysis method based on support vector machine to identify different fault modes of rotating machinery.

```plaintext
While
  for i ← 1 to N
    if (alarm(i)==1||state(i)>=stateThreshold(i))
      alarm(i) ← 1; //alarm or state is greater than threshold
      //call corresponding available maintenance equipment to item
      for j ← 1 to numFac[i] //available maintenance
        if (maintainFac[i,j]==0)
          call maintainFac(i,j);
          maintainFacility(i,j) ← 1;//set to work status
          //wait for maintenance equipment to work
          wait();
        if (state(i)<stateThreshold(i))
          alarm(i) ← 0;
          maintainFacility(i,j) ← 0;
        else //equipment has not completed the task
          for k ← 1 to numStaff[i] //idle personnel
            if (maintainStaff[i,j]==0)
              call maintainStaff(i,j);
              maintainStaff(i,k) ← 1;//set working status
              alarm(i) ← 0;
            maintainStaff(i,k) ← 0;
            Break;
          End for
      End if
  End for
End While
```

Fig 3. Pseudocode of the data processing

It can be seen from the above research that there are endless ways of feature recognition and processing mode in fault diagnosis. With the development of the manufacturing industry, the research in the field of fault diagnosis is deepening. The requirements for real-time and prediction accuracy of fault diagnosis are higher. The fault diagnosis algorithm is more intelligent, and the diagnosis function is more and more perfect. However, the existing researches mostly focus on the spare parts of equipment, such as the fault and life prediction of motor or gear box in a certain equipment, but the overall monitoring of equipment is insufficient.
2.3. Remote Fault Diagnosis
With the development of information technology, especially the emergence of 5G technology, the technical framework of fault diagnosis will change rapidly. Due to the sharing of technical diagnostic information, local diagnostic capability is limited.

According to [22], a kind of cloud message-based intelligent remote diagnosis service architecture was introduced, which can be used for remote diagnosis of multiple remote equipment distributed in different regions, and make diagnostic analysis by experts in different regions or artificial intelligence experts. The architecture is suitable for remote collaborative diagnosis of different types of equipment and has good scalability and adaptability. Some production sites do not have network environment and thus cannot diagnose faults on site. The Remote fault diagnosis network architecture is shown in Fig.4.

Li et al. [23] proposed to use remote data transmission terminal DTU and GPRS to set up wireless communication environment for remote transmission of data, so as to realize remote diagnosis of fast forging press.

With the increasing complexity of equipment structure, many enterprises are not able to monitor the running state of equipment in real time and give early warning of faults. Under the new manufacturing mode, the maintenance cost of equipment increases sharply. More and more researchers want to improve the accuracy of equipment fault diagnosis through new information and communication technology and intelligent diagnosis algorithm.

Fig 4. Remote fault diagnosis network architecture

3. Challenges and prospects of fault diagnosis based on big data

3.1. Industrial big data
With the development of decades, fault diagnosis and early warning technology has been improving its efficiency in data collection and algorithm model establishment. Information physical systems(IOT), cloud computing and industrial Internet have led the development of industry 4.0. The growth rate of industrial manufacturing data is unprecedented, and the academia has a scientific definition of big data [24]: big data refers to the collection of data that cannot be perceived, captured, managed, processed and served in an acceptable time by traditional software and hardware technologies or other tools. Scholars have proposed 3V, 4V and 5V to describe, namely Volume, Velocity, Variety, Value and Veracity. Manufacturing big data can be roughly divided into three types: equipment data, production data and manufacturing command data. Through the collection and in-depth analysis of equipment information, the fault information can be quickly mined, and the rapid diagnosis and prediction during operation can be realized.
3.2. Challenges Faced

1) In order to grasp the status of equipment more accurately and find the potential fault information in time, the range of equipment detected gradually becomes larger, the time is extended, and the sampling frequency of monitoring points also increases, which also enables the monitoring center to process massive data. The collected data contains a variety of useful information from various aspects and different types, but it seems inadequate to rely only on experts and other technical personnel to search for useful information in a large amount of data.

2) Traditional intelligent diagnosis methods usually acquire signal features first and then train the model of diagnosis. In terms of fault information extraction, a large number of signal processing technologies are required, and experts are required to design feature extraction algorithms, which largely rely on the prior knowledge of signal processing and expert diagnosis. The shallow intelligent model is often used to identify mechanical faults, but it does not adapt to the equipment fault characteristics in the era of big data.

3) The existing equipment intelligent diagnosis and life prediction system is mostly oriented to the level of equipment parts. Equipment is composed of multiple functional systems, and parts are coupled to each other to form a whole. The failure of a single module will inevitably affect its coupled parts. Therefore, using big data to realize fault diagnosis and early warning of the whole equipment operation is still a big challenge.

4) Remote fault diagnosis system is based on computing, information communication and other technologies to connect the distributed equipment through industrial LAN/WAN. However, in the case that the internal network system of many manufacturers has not been established or cannot be connected with the external network due to confidentiality, the timeliness of fault diagnosis will be lost if the equipment status information is separately packaged and transmitted. The advantages that instant messaging has brought to remote troubleshooting will no longer exist.

3.3. Future Work

Fault diagnosis based on big data is facing a series of challenges. Researchers have improved equipment fault diagnosis methods, such as combining big data characteristics of mechanical equipment with deep learning theory, training deep learning network with a large amount of real-time test data, constructing in-depth model and extracting hidden features, so as to complete equipment fault diagnosis. As a branch of machine learning, deep learning is characterized by hierarchical structure, consisting of input layer, hidden layer and output layer. The deep learning neural network structure is presented in Fig.5. The most commonly used deep belief network avoids the process of artificial feature extraction and selection, and reflects fault information through network characteristics [25].

![Deep learning neural network structure diagram](image-url)
Multiple data fusion. There are many kinds of equipment in manufacturing industry, and the transmission is complicated, so the fault information produced has the characteristics of complexity, concurrency and non-stationarity. Data fusion based on deep learning is an optimal method to solve the uncertainty of equipment fault diagnosis. The method is used to fuse multi-source data, overcome the characteristic limitation and increase the accuracy of decision diagnosis. Therefore, the introduction of data fusion method in the field of equipment fault diagnosis is an inevitable trend of the fault diagnosis technology. The Fig.6 represents fusion diagnosis strategy, (a) represents the network topology of fault information fusion diagnosis, and (b) represents the fault information fusion diagnosis strategy.

In addition to more intelligent and precise algorithms, cloud message-oriented middleware technology can be used to improve the big data diagnosis service architecture in the communication, storage and analysis of equipped big data. Research ZeroMQ and other fast message communication technologies, as well as use its characteristics of high throughput and low latency can meet the transmission requirements of equipment data in different regions, so as to realize the rapid processing from big data of equipment fault to the generation of cloud diagnosis results. In addition, attention should be paid to the construction of operational database for equipment cluster, so that it has the characteristics of data sharing, low redundancy, high independence and convenient centralized control.

4. Conclusions
Taking the fault data of rail transit and power system in the new era as an example, this paper expounds the opportunity of intelligent fault diagnosis of equipment in the era of big data, and summarizes the research results of researchers on intelligent fault method of data collection and remote fault diagnosis mode.

The paper summarizes the basic types of manufacturing big data and 5V characteristics. Finally, the challenges faced by intelligent fault diagnosis technology and system construction under the background of big data are proposed, and it is believed that in-depth studies can be carried out from the aspects of deep-learning-layer fault identification, multi-data fusion, data transmission and storage, etc. Thus, the equipment fault diagnosis technology driven by big data will be pushed forward.
Acknowledgments
This work was financially supported by the Key Research and Development Program of Shaanxi, China (Grant No. 2019ZDLGY01-01-01) fund. Many thanks are expressed to Qian Feng for her kind help during the preparation of the manuscript.

References
[1] Zhang Tieshan, et al. "Research on the influence of big data on the business model of high-end equipment manufacturing." Mall modernization, 2016(18):2-4.
[2] Yuan wei, From "big province of equipment manufacturing" to "strong province of equipment manufacturing" -- interpretation of zhejiang high-end equipment manufacturing development plan (2014-2020), Zhejiang Today, 2015(3):30-31.
[3] Wang Peijian, D - pro: dynamic data center operations with demand - responsive electricity prices in smart grid. IEEE Transactions on Smart Grid, 2012, 3(4): 1743 - 1754.
[4] Cheng, Chang. "Power System Security Assessment and Fault Diagnosis System Based on PDMiner Large Data Mining Platform." Power & Energy (2016).
[5] Yang Luo, Lei, et al. "Opportunities and Challenges of Machinery Intelligent Fault Diagnosis in Big Data Era." Journal of Mechanical Engineering (2018).
[6] Wang ling, "Research on Intelligent Fault Diagnosis Technique of Complex Equipment." Diss, Shenyang normal university, 2011.
[7] YANG Lijin, Yang Ye, Tian Jingyi, "Intelligent Fault Diagnosis of Large and Complex Equipment." Coal Mine Machinery, 33.1(2012):267-269.
[8] TANG Baoping, HUANG Qingqing, DENG Lei, et al. "Research progress and challenges of wireless sensor networks for machinery equipment condition monitoring". Journal of Vibration, Measurement & Diagnosis, 2014, 34(1): 1-7.
[9] Shaomin, Zhang, M. Dong, and W. Baoyi. "Application of Big Data Processing Technology in Fault Diagnosis and Early Warning of Wind Turbine Gearbox." Automation of Electric Power Systems (2016).
[10] 2016. [Online]. Available: http://www.advantech.com.cn/.
[11] H. Shen, Z. Li, L. Yu, and C. Qiu, "Efficient data collection for large scale mobile monitoring applications," IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 6, pp. 1424–1436, Jun. 2014.
[12] Z.Xu, L.Chen, C.Chen, and X.Guan, "Joint clustering and routing design for reliable and efficient data collection in large-scale wireless sensor networks, "IEEE Internet Things, vol. 3, no. 4, pp. 520–532, Aug. 2016.
[13] Liang Xuifan, "Research Oil coal mill condition monitoring and fault detection based on first principle approach and on—site data" [D]. 2016.
[14] Wang changchang, et al. "On-line monitoring and diagnosis of power equipment.Beijing: tsinhua university press,2008.
[15] MA Changgang, et al. " Research on Intelligent Fault Diagnosis Methods for SomeEquipment Based on Neural Network." Naval Electronics Engineering, 10 (2017):110-113.
[16] Yaguo, Lei, et al. "A Relevance Vector Machine Prediction Method Based on Adaptive Multi-kernel Combination and Its Application to Remaining Useful Life Prediction of Machinery." Journal of Mechanical Engineering 52.1 (2016):87.
[17] H. Hashemian and W.Bean, “State-of-the-art predictive maintenance techniques,” IEEE Trans. Instrum. Meas., vol. 60, no. 10, pp. 3480–3492, Oct. 2011.
[18] Wan, Jiafu, et al. "A Manufacturing Big Data Solution for Active Preventive Maintenance." IEEE Transactions on Industrial Informatics (2017):1-1.
[19] SEERA M, LIMCP, NAHAVANDIS et al. "Condition monitoring of induction motors: A review and an application of an ensemble of hybrid intelligent models". Expert Systems with Applications, 2014,41 (10) :4891-4903.
[20] Zhang, L., et al. "Intelligent fault diagnosis of certain missile equipment." Journal of Central South University (Science and Technology) (2013).
[21] SHEN Changqing, WANG Dong, KONG Fanrang, et al. Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier. Measurement, 2013, 46 (4):1551-1564.

[22] Wang Huatao and Yang Lili, "A Kind of Intelligent Remote Diagnosis Service Architecture Based On Cloud Message", Computer and Information Technology, v.26; No.153.03 (2018):12-17.

[23] Li Chaohui, et al. "High-speed Forging Hydraulic Press Fault Diagnosis System." Industrial Control Computer (2016).

[24] Li Guojie, Cheng Xueqi, "big data research: research status and scientific thinking of big data: a major strategic field for future scientific, technological and economic and social development", Proceedings of The Chinese Academy of Sciences (2012).

[25] GENG Zhi-qiang,ZHANG Yi-kang. An Improved Deep Belief Network Inspired by Glia Chains [J]. Acta Automatica Sinica,2016,42 (5):643-654. [2] B. P. Abbott, R. Abbott, T. D. Abbott et al., “Observation of Gravitational Waves from a Binary Black Hole Merger,” Physical Review Letters, vol. 116, no. 6, Article ID 061102, 2016.