Big Data Technology Application in Mechanical Intelligent Fault Diagnosis

Huiteng Cao

Shandong Vocational College of Light Industry, Zibo, China

*Corresponding author e-mail: caohuiteng@sdlive.com

Abstract. With the rapid implementation of made in China 2025 plan and the rapid development and application of information technology such as artificial intelligence, big data technology, industrial Internet of things and 5G, information technology has been integrated into every link of the whole life management cycle of mechanical products, such as tool condition detection and mechanical fault diagnosis in machining process. Based on this, the purpose of this study is to study the application of big data technology in mechanical intelligent fault diagnosis. In the process of this study, the decision number algorithm and data mining algorithm are used to study the experiment, and some mechanical faults in the past are analyzed and studied. Summary of the experimental results show that the use of decision number algorithm and data mining algorithm in the experiment has achieved good results, through these methods and big data technology, we can quickly diagnose the fault of mechanical equipment, accurately locate the fault location of mechanical equipment. Mechanical intelligent fault diagnosis based on big data technology can improve the efficiency of fault diagnosis, reduce enterprise costs and improve economic performance.

Keywords: Big Data, Fault Diagnosis, Bayesian Classification Algorithm, Mechanical Intelligence

1. Introduction

Nowadays, big data technology and industry IOT have been widely used in the mechanical industry. This paper studies the application of big data technology in mechanical fault diagnosis.

The machinery operators see the mechanical fault, record it and send it to the technical specialist, and then the specialist finds out the mechanical fault and repairs it through the fault symptoms provided by the operators before the big data technology is used. However, the operators fail to record the fault symptoms immediately when the failure happens, and they forgets some information usually, which makes that the specialist can’t identify the failure quickly. Meanwhile, the technical specialist solves the mechanical fault passively

The traditional mechanical fault diagnosis technology has a lot of disadvantages, which is that the technical experts need to diagnose the possible places of the machinery after the faults are exposed, and because the faults are diagnosed through the oral description of the operators, it is difficult to predict the faults in the machinery. Moreover, the staff's description of the mechanical failure also has
certain subjectivity. They may describe the failure through experience, so there will inevitably be errors. There is a lot of the subjective and obvious fault description provided by the operator, and the subtle and inner failure which can’t be seen from its appearance can’t be recorded, so it will make that the specialist make the errors of fault diagnosis and the fault prediction. Most failures are led by the effectiveness of small part, and the specialist and the operator can’t find the small part effective lost.

Big data technology can record the running data of machine in real time, which indicates the health of the machine, so the fault diagnosis can be finished through the detailed data. The failure regularity of machine may be found out through the real-time data of device running state, the failure model may be set up by the data then. The real-time data will be analyzed and sent to the operator and specialist, and they can maintain the machine in advance of machine fault on the one hand, on the other hand the obvious fault can be diagnosed and repaired in time by the hand of running data. The production can be ensured by big data technology, and the unnecessary loss can be minimized.

With the rapid development of national information technology, big data technology will be widely used in various fields in the future. In the field of mechanical fault diagnosis, experts and scholars have begun to focus on it. In the research process, Yan proposed to use big data analysis methods such as time series/unsupervised learning to monitor mechanical equipment. During this period, big data mining algorithm was used to monitor and calculate the anomalies of mechanical equipment, and then verified with the existing faults in the database. In case of possible fault diagnosis, a fault warning was given and the staff could repair it in time. This method can quickly detect the possible or impending faults in machinery, realize the timely diagnosis of faults, and avoid unnecessary losses [1]. In the research process, Qu proposed to use MapReduce parallel framework to improve the original Bayesian classification algorithm, which improves the efficiency of Bayesian classification, and can effectively improve the efficiency of mechanical fault diagnosis, and make Bayesian classification algorithm more suitable for the big data environment, so as to obtain more effective diagnosis rules and make fault diagnosis more accurate [2].

With the increasing integration of mechanical manufacturing, the condition and maintenance of mechanical equipment is particularly important. Because big data technology has been applied in the field of mechanical intelligent fault diagnosis very well, big data technology has become more and more important in the field of mechanical intelligent fault diagnosis.

2. Experimental Tool for Mechanical Intelligent Fault Diagnosis

2.1. Big Data
Big data refers to the diversified, high growth rate and a large amount of information data collected by software tools from many sources. These data come from many places, such as the Internet, social tools, phone records, etc. Moreover, big data does not use random analysis, which is from the sampling survey method to collect and analyze data. Big data has five characteristics: massive data, high processing speed, diversification, low value density and authenticity. Big data technology can mine and analyze massive data. For people, big data will mine and analyze people's information, and then provide reliable information data automatically in the next service. Today's society is a highly information-based world. Science and technology are developed, information flow is very fast, and information exchange between people is becoming closer. All these are closely related to big data [3-4].

2.2. Data Mining Algorithm
"Data mining algorithms are a set of heuristics and calculations that create data mining models based on data" [5]. In this paper, due to the large amount of data generated during on-line monitoring of capacitive equipment, I used the data mining method of big data to conduct statistical analysis on these data, and find out the beneficial and useful information hidden therein, so that the monitoring can more effectively find the problems existing in the equipment and reduce the unnecessary loss of enterprises or power plants [6].
2.3. Fault Diagnosis
Fault diagnosis refers to the process of using various inspection and testing methods to find equipment or system faults. Fault detection refers to the process of checking equipment or system to find out whether there is a fault, and then further to determine the approximate failure site; fault isolation refers to the process of isolating the fault location after the fault is located and repairing until the fault is completely repaired. Nowadays, the fault diagnosis is a kind of technology to understand and master the movement state of the machine, determine whether the machine is abnormal, and also can detect whether the machine has fault in advance and give a quick warning [7].

2.4. Bayesian Classification Algorithm
Bayesian classification algorithm is a statistical method, which uses probability statistics and knowledge for classification calculation. It is a common classification method in processing classification tasks. Bayesian classification algorithm needs to build a Bayesian classifier, first of all, it should assume that one of the specified categories, the values of its attributes are independent of each other. This assumption is called class condition independence, which can ignore the parameter of each attribute in the calculation, thus effectively reducing the amount of computation in Constructing Bayesian classifier. For Bayesian probability theory, let $X$ be a data sample that has not been classified, and $H$ is a hypothesis. If the data sample $X$ has a specific classification attribute, then the classification problem can be understood as solving $P(h|x)$, that is, for a specific data sample $x$, the probability that h hypothesis holds.

$P(h|x)$ denotes the post event probability, that is, the H probability based on condition $x$ [8-9].

Since the probabilities of $P(x)$, $P(H)$ and $P(XH)$ can be calculated from the data information provided in the data set, Bayesian theorem is used to describe how to calculate $P(HX)$ according to $P(x)$, $P(H)$ and $P(x \mid h)$. The specific formula is defined as follows:

$$p(H \mid X) = \frac{p(H)p(X \mid H)}{p(X)}$$ (1)

It can be understood that the task of the basic Bayesian classifier is to assign the sample $X$ of the unknown class to the class $C_i$, which is valid if and only if $P(c \mid x) > P(c \mid x)$, where $1 \leq j \leq m, j \neq i$, that is to make $p(c \mid x)$ value maximum. The attribute category $C_i$ is called the predicted value of maximum post event probability. It can be concluded that:

$$p(C_i \mid X) = \frac{p(X \mid C_i)p(C_i)}{p(X)}$$ (2)

3. Experiments

3.1. Subjects
The experimental object mainly carries on the mechanical fault diagnosis research to the wind turbine mechanical group. In this experiment, the research group participated in the data collection of failure times of mechanical equipment in wind power enterprises in the past years. This experiment is split into three groups.

The experimental group is the researcher who will analyze the mechanical fault data collected in wind power enterprises. In this process, big data technology is used to settle and separate the fault diagnosis data of mechanical equipment in order to obtain the fault cause data, and set up the failure model then.

The comparison group is to obtain the different data in the aspect of fault diagnosis and fault prediction between the traditional fault diagnosis method and the mechanical intelligent fault diagnosis based on big data technology, then find out the advantage of the mechanical intelligent fault diagnosis based on big data.
The analysis group is to analyze the data of the whole experimental process and to draw a conclusion.

3.2. Experiment Design
In the process of experiment research, the experiment scheme of mechanical intelligent fault diagnosis based on big data is designed after lots of the paper result and relevant medical research are considered. First, the four experiment schemes are designed, which are on the basis of the consideration of the implementation objet. The machine health condition can be monitored in real-time, and the prediction about the machine failure can be drawn after the wind power machine running-data is analyzed and simulated. Second, the running data is divided into different categories in order to obtain the experiment data. Third, the Bayesian classification method and decision tree algorithm are used to calculate the useful data separated. Finally, all the useful data are analyzed as a whole, and the conclusion can be drawn about the deficiencies in the research process of mechanical intelligent fault diagnosis. These deficiencies is discussed in the paper and different ways are used to eliminate them.

3.3. Statistical Methods
The SPSS software was used to analyze the collected data, LSD method was used to calculated the variance, and the test level was $\alpha = 0.05$.

4. Discussion

4.1. The Mechanical Fault Diagnosis Experiment of Wind Turbine
The object of this experiment is the wind turbine of wind power enterprise. In this experiment, the research group participated in the work well. As the experimental object of this experiment is wind turbine, so we do a statistical experiment on the wind turbine of national wind power enterprise. According to the statistics of public data collected on the Internet, by the end of 2018, China's cumulative wind power grid connected capacity was about 170 million kilowatts. Figure 1 shows the statistics of China's new wind power grid connected installed capacity from 2014 to the first quarter of 2018. At the same time, with the improvement of the operation life of wind turbines, most of the existing wind turbines have exceeded the warranty period, so the proportion of operation and maintenance costs in the total revenue is increasing. How to improve the machine quality and efficiency of repairing and maintaining the power machine has become an urgent problem in the field of wind power, and the healthy operation of wind turbines is particularly important [10].

![Figure 1. 2014-2018 Wind power grid capacity statistics](image-url)
4.2. Wind Turbine Gear Fault Test
The gear of wind turbine is connected with the main shaft and the wind turbine, which is an important part of the unit. Moreover, the internal structure and stress condition of the gear are very complex, and the operating environment is bad. Because the wind turbine must work for a long time, when the gear is running, it is often under overload, and it is easy to fault [11]. In this regard, we will gear common fault types of statistics, the statistical results are shown in Table 1.

Table 1. The fault type and fault frequency

| Fault location     | Number of failures | Average downtime/h | Number of wear failures |
|--------------------|--------------------|--------------------|------------------------|
| Bearing            | 45                 | 560                | 35                     |
| Gear               | 5                  | 270                | 3                      |
| Seal up            | 10                 | 50                 | 5                      |
| Lubrication System | 15                 | 25                 | 5                      |

4.3. The Comparative Experiment of Wind Turbine Gear State
This experiment is to compare the state of gear failure with that of normal condition, so that it is easier to distinguish whether the gear is abnormal and which kind of fault occurs in the future work. In this experiment, the motor speed is set at 1000r/min, and the sampling frequency is 10.25khz, and the data of gears in different states are collected [12]. As shown in Figure 2 below is the effect picture of sample state recognition. The results of 100 repeated experiments show that the average accuracy of the experiment is 99.5%.

![Figure 2. Classification results of four states](image)

As shown in the figure 2 above, the initial signal value under normal state is 1, and the signal value of the other three fault states is 0. In addition, we find that the time when the signal value of the other three fault states is 1 is different. From this, we can also diagnose what kind of gear fault occurs when the gear fails.

5. Conclusions
This paper mainly focuses on the application in the big data in intelligent fault diagnosis field of mechanical equipment. In recent years, big data technology has also been applied in various industries, and mechanical intelligent fault diagnosis technology has also joined the big data technology, which provides the possibility and foundation for the realization of mechanical intelligent fault diagnosis. Nowadays, it has been possible to comprehensively grasp the running state of the whole machine or
system, comprehensively monitor the deterioration law of mechanical equipment, grasp the health status of mechanical equipment, diagnose possible mechanical faults, and then repair the faults timely and fast manner to prevent the impending failure, which can reduce the unnecessary losses of the factory. Therefore, in the research process, combined with the distinguishing feature of the mechanical equipment intelligent diagnosis based on big data technology, the application data and the technical route of big data in mechanical intelligent fault diagnosis are established, and the problems obtained in the process of mechanical equipment intelligent fault diagnosis are studied, so as to improve the application of the big data technology.

References
[1] Yaguo L, Feng J, Detong K, et al. Opportunities and Challenges of Machinery Intelligent Fault Diagnosis in Big Data Era. Journal of Mechanical Engineering, 2018, 20(1):1-8. (in Chinese).
[2] Liu X. Research on Automatic Fault Diagnosis Technology of IT Equipment Based on Big Data International Conference on Human-Computer Interaction. Springer, Cham, 2019,7(1).00-38.
[3] Cao H. Big Data Attribute Selection Method in Distributed Network Fault Diagnosis Database. Journal of Intelligent and Fuzzy Systems, 2020,11(2):1-12.
[4] Qu J T, Liu F, Meng H. A Method for CIR Fault Diagnosis Based on Improved Tri-Training in Big Data Environment. 2018,11(1):213-218.
[5] Chen, Ganlang. Research on Big Data Attribute Selection Method in Submarine Optical Fiber Network Fault Diagnosis Database. Polish Maritime Research, 2017, 24(3):74-84.
[6] Kai Y, Zhenyu Y, Jiahao G. Research on Operation Condition Assessment and Fault Diagnosis of Power Metering Equipment Based on Big Data. Chinese Journal of Electron Devices, 2019, 37(11):2441-2452. (in Chinese).
[7] Xu Q, Zhang P, Liu W, et al. A Platform for Fault Diagnosis of High-Speed Train based on Big Data. 2018, 51(18):309-314.
[8] Onel M, Kieslich C A, Guzman Y A, et al. Big Data Approach to Batch Process Monitoring: Simultaneous Fault Detection and Diagnosis Using Nonlinear Support Vector Machine-based Feature Selection. Computers & Chemical Engineering, 2018, 115(4):503-520.
[9] Jiang B, Guo Z, Zhu Q, et al. Dynamic Minimax Probability Machine-Based Approach for Fault Diagnosis Using Pairwise Discriminate Analysis. IEEE Transactions on Control Systems Technology, 2019, 27(2):806-813.
[10] Zhang S, Bi K, Qiu T. Bidirectional Recurrent Neural Network-Based Chemical Process Fault Diagnosis. Industrial & Engineering Chemistry Research, 2019, 61(6):719-723.
[11] Wang Y, Yan J, Yang Z, et al. Partial Discharge Pattern Recognition of Gas-Insulated Switchgear via a Light-Scale Convolutional Neural Network. Energies, 2019, 12(1):147-149.
[12] Wang L, Shang L, Ma M, et al. Fault Diagnosis and Trace Method of Power System Based on Big Data Platform. IOP Conference Series Materials ence and Engineering, 2018, 394(4):042-116.