Research on Early Warning of Hoist Failure based on Big Data and Parallel Simulation

Yuyan Zhang\textsuperscript{1,a}, Sihai Zhao\textsuperscript{2,b*}, Dan Li\textsuperscript{3,c}

\textsuperscript{1,2,3} School of Mechatronics and Information Engineering, China University of Mining and Technology (Beijing), Beijing, 100083, China

\textsuperscript{a}email: zhangyuyan527@163.com

\textsuperscript{b*}Corresponding author’s e-mail: zsh@cumtb.edu.cn

Abstract: Aiming at the problems of easy fault diagnosis and difficult early warning of mine hoist, a parallel system architecture of hoist fault early warning based on big data is proposed, the structure of each subsystem of hoist is analyzed, and a parallel simulation system of hoist fault early warning is established; secondly, the Hadoop ecosystem of hoist is established on the virtual machine, and the massive data is mined by using clustering algorithm and association rule algorithm, so as to speed up the calculation speed and improve the reliability of early warning; finally, the safety state evaluation rules of hoist are proposed, and the system decision is made according to the fault early warning results. The experimental results show that it can achieve the purpose of fault prediction.

1. Introduction:
As one of the large-scale production equipment, the operation quality and efficiency of mine hoist not only affect the economic benefit of coal mining enterprises, but also affect the life safety of workers.

At present, most of the research on hoist fault is fault diagnosis or fault early warning of some parts, and there is little research on fault prediction of the whole machine. Moreover, the soft fault of hoist system is difficult to predict by simple analysis, and the fault hazard can be diagnosed only after data mining of a large number of data \cite{1}.

The hoist fault early warning system based on big data technology and parallel simulation established in this paper obtains the real state and a large amount of data presented by the hoist real fault early warning system through the combination of real system and manual system. After the joint analysis of Hadoop platform and mathematical model, it drives the hoist fault early warning system of parallel simulation, then the simulation results are analyzed to assist the decision-making of the real system. Thus, the artificial system and the real system complement each other, so as to achieve the best fault early warning effect of the hoist. It also provides a new research idea for the study of hoist fault.

2. design of hoist fault early warning data analysis platform based on Hadoop
During data mining and analysis of the hoist, because the hoist has a large amount of accumulated historical data, although there are some real-time dynamic data, the hoist failure will not occur in a moment; on the other hand, considering the difficulty of obtaining development tools, this paper selects Hadoop as the network infrastructure of data mining. Using the database of massive data that can be controlled in Hadoop and powerful computing power, combined with machine learning, this
paper mines and analyzes the massive data of hoist.

2.1 hoist data construction design
(1) Data preprocessing and ETL

Take the auxiliary shaft lifting system of hoist in Cheji mine of Yongcheng Electromechanical Group as an example. Through the analysis of the data records of the mine monitoring system, it can be seen that most of the data of the mine hoist are structured data, and the general big data processing database is enough to express the relationship between the data of the hoist clearly. The historical data monitored by each sensor in the past 5 years are classified and sorted to facilitate database reading and writing.

First of all, according to the large amount of raw data collected from the hoist, the collected data may not be complete, the existence of noise, fuzzy signal and other characteristics, a large number of historical data of the hoist is preprocessed to improve the data quality [2]. After data preprocessing, in order to integrate all the monitoring system data into a unified database, we need to carry out data integration. Considering that the data collected by the hoist monitoring system is multi-source and heterogeneous, ETL technology is finally selected for data integration.

The data preprocessing method mainly uses the twice data filtering method, and the data preprocessing flowchart is shown in Figure 1.

![Fig. 1 preprocessing flowchart](image)

(2) Hoist fault information storage based on Hbase and Hive

The Hadoop platform ecosystem consists of HDFS, the file management system for data analysis, MapReduce, the underlying logical architecture for data computing, Hbase, the relational database, and Hive.

Firstly, the data of mine hoist is processed by filtering twice to get the normal data of each monitoring point sorted by time. Then, the pre-processed data is classified, different categories and formats are corresponding to the corresponding data, and the ETL tool is used to extract the data as required, and the data is stored in HDFS. After a large amount of data is cleaned using Hive, the data is stored in Hbase to complete data storage for the hoist.

2.2 Modeling design of hoist fault prediction based on MapReduce
(1) Hoist fault prediction algorithm based on MapReduce

The original data of the hoist is based on time series data. Time series data is characterized by complexity, high dimension, real time, high noise and large scale [3-4]. Considering that the data volume of the hoist is large but the overall data structure is clear and clear, the K value can be preset according to the experience, so the most classical k-means clustering algorithm is selected. In the
further analysis of the fault, Apriori algorithm in association rules is selected to save calculation time. Its core is based on the two-stage top-down recursion method in association rules.

Based on the powerful computing power of MapReduce, the k-means algorithm and association rule algorithm are combined with MapReduce, and k-means is used to mine the data of the hoist. The normal data and the fault data are classified. After the clustering algorithm data mining out the abnormal data of the hoist monitoring institutions, it is necessary to further analyze them. Confirm the location and location of the fault. The fault data can be processed in parallel by association rules, and rules and rule confidence can be generated. By combining the Apriori algorithm with MapReduce and using the parallel computing efficiency of MapReduce, the computational complexity of association rules is reduced, the time of data analysis is saved, and the accuracy of data analysis is improved. To achieve the purpose of analyzing the data to predict the fault.

(2) Fault prediction data model modeling

Based on the clustering algorithm and association rules, the existing historical data and real-time data of the hoist can be mined and analyzed. If there is no obvious fault, the future data will be predicted through the current certain amount of data, and then data mining will be carried out on the forecast data.

For the fault state prediction of the hoist, the structure of the hoist is known, the monitoring system and data of the hoist are known, the common fault classification and fault characteristics of the hoist are known, and the running data of the components in the future state is unknown. At the same time, considering the low accuracy of data prediction using a single model, this paper uses the variance-covariance weight method to combine the grey model GM(1,1) and the differential integrated moving average model ARIMA, that is, the AG combination model is used for modeling, so as to achieve the optimal prediction effect. The modeling is shown in Figure 2.

![Fig. 2 AG model](image)

The predicted value of ARIMA model is $y_1$, the predicted value of GM(1,1) model is $y_2$, and the model weight coefficients are $\omega_1$ and $\omega_2$ respectively. There are:

$$y_d = \omega_1 y_1 + \omega_2 y_2$$  \hspace{1cm} (1)

Then the error test is carried out and the test results are evaluated.
3. Design of parallel system for hoist fault early warning

3.1 Parallel system of hoist

Based on the system framework idea proposed above, a platform with multiple functions is built as shown in figure 3. The platform is mainly built around the idea of ACP, including database, tool library, Hadoop platform, condition evaluation, fault early warning, maintenance, visualization and other functional modules [5-6]. The database contains the historical data, real-time data, environmental data and detailed equipment data of the hoist; the tool library contains a big data processing platform, which can perform parallel operations on various data in the database and improve the ability of analysis and processing [7-8]. The clustering algorithm can be used to calibrate and extract the feature data. The data is introduced into the AG mathematical model for machine learning and further evaluation. Based on the comprehensive evaluation results, the actual system is maintained and optimized. The visual function module is used to display the processing process interactively.

3.2 Design of hoist production process safety state evaluation system based on parallel control

Considering the underground safety, system reliability, lifting equipment reliability and economy, the state safety evaluation set of hoist is obtained = {motor operation stability, spindle work quality and efficiency, mechanical equipment safety, and temperature of important parts}.

For the data that have obtained the analysis results after data mining, the analytic hierarchy process evaluation method is selected to obtain the safety state evaluation results of the hoist.

Through the corresponding weight calculation of the selected evaluation indexes, the description of the safety state level of the hoist can be obtained. In this paper, the safety status of the hoist is divided into four levels: Health (Ⅰ), good (Ⅱ), deterioration (Ⅲ) and failure (Ⅳ). That is, the safety state set of the hoist H = {Ⅰ Ⅱ Ⅲ Ⅳ}.

Fig. 3 Parallel system framework for early warning of hoist failure
3.2.1 Establishment of safety state evaluation model of hoist

The establishment process of hoist safety state evaluation model is shown in Figure 4:

According to the established state set and evaluation set, the hoist safety evaluation matrix is as follows:

\[
B = \begin{bmatrix}
B_1 \\
B_2 \\
\cdots \\
B_n
\end{bmatrix}
\]

The safety evaluation vector of the i-th component is:

\[
B_i = W_i \cdot R_i = [b_{i\,\text{I}} \, b_{i\,\text{II}} \, b_{i\,\text{III}} \, b_{i\,\text{IV}}]
\]

The safety state evaluation vector of the hoist is:

\[
E = W \cdot R = [b_{\text{I}} \, b_{\text{II}} \, b_{\text{III}} \, b_{\text{IV}}]
\]

3.2.2 early warning decision based on hoist safety evaluation system

After data mining, the current state of each component is introduced into the safety state evaluation model in the hoist manual system to obtain the calculation results, which guide the decision-making in the actual system of the hoist. If it is healthy and good, continue to repeat the evaluation of the current state. If it is degraded, it indicates the degraded part. If it is a fault, it will give an early warning and shut down for maintenance according to the actual situation.

The complete workflow of the manual system of the hoist includes test design, calculation, fault prediction and state evaluation. As shown in Figure 5.
4. experimental verification and analysis
In order to verify the effect of Hadoop platform combined with clustering algorithm and association rules on massive data mining of hoist, and whether the established hoist safety state evaluation rules are reasonable for the decision-making of real system, the data processing platform and data model of hoist fault early warning parallel system are verified and analyzed here.

4.1 data mining model validation

4.1.1 Clustering model verification
After the Hadoop cluster is set up successfully, eclipse can communicate with the cluster, program the clustering algorithm through Eclipse, and calculate the results through MapReduce.

After program debugging, the data of each parameter of the hoist can be classified as normal data and fault data (lower or higher than normal data).

Partial clustering results of hoist data are shown in Figure 6 and 7:
As shown in the figure, the red part of clustering 1 for speed and voltage is the data within the normal range; the green part of clustering 2 is the data higher than the normal range; and the blue part of clustering 3 is the data lower than the normal range.

The normal data and the fault data are classified by clustering, and then the fault data is imported into the association rules for further confirmation.

### 4.1.2 Association Rule Verification

#### Table 1 Failure comparison table

| Serial number | fault            | Fault performance                                      | FM expression |
|---------------|------------------|--------------------------------------------------------|---------------|
| 1             | Motor fault      | Speed overrun                                          | F1            |
| 2             | Abnormal temperature | The temperature exceeds the normal range               | F2            |
| 3             | Brake failure    | The speed exceeds the normal range                     | F3            |
| 4             | Oil pressure fault | The pressure exceeds the normal range                  | F4            |
| 5             | Wire rope fault  | The current exceeds the normal range                   | F5            |

After the above clustering analysis, the data beyond the normal value range are divided into different clusters. At this time, the abnormal data output will no longer use the original data, but mark and classify according to the fault comparison table in Table 1, generate new key value pairs, input them into the association rules, and enter a new round of MapReduce. The treatment results are shown in Table 2:

#### Table 2 Apriori results

| condition | conclusion | Support | Confidence |
|-----------|------------|---------|------------|
| F4        | F3         | 0.29    | 0.67       |
| 5         | F1         | 0.28    | 0.5        |
| F3        | F4         | 0.40    | 0.67       |
| 5         | F5         | 0.29    | 0.67       |
| F4        | 5          | 0.27    | 0.5        |
It can be seen from the data in the first row that when the oil pressure fails, the probability of abnormality of the brake is 0.29 and the accuracy is 0.67; It can be seen from the third line that when the brake fails, there is a probability of 0.4 that the brake failure will occur, and the support degree is 0.67. High support indicates high probability of occurrence, high importance, and high confidence indicates high accuracy.

4.2 validation of data prediction model

Taking the elevator motor speed data of Cheji mine of Yongcheng Electromechanical group as an example, it is imported into the prediction model of elevator fault early warning manual system for experiment. Some data are combined with Ag prediction model to make the prediction model achieve the target effect.

![Fig. 8 GM(1,1) model prediction results](image1)

![Fig. 9 ARIMA model prediction results](image2)

![Fig. 10 AG model prediction results](image3)

It can be seen from the prediction figure 8 ~ 10, the error of the single model forecast data and the actual data is larger than that of the AG combination forecast data. It is proved that the AG combination model can improve the accuracy of future data prediction.

4.3 verification of combination of virtual reality and early warning of parallel system

Through the above calculation results in the hoist manual system, the hoist safety evaluation system is applied. The results are as follows:

![Fig. 11 Diagram of early warning results of hoist failure](image4)
As shown in Figure 11, after the calculation results processed by the association rules are evaluated through the evaluation model, the result diagram can confirm that the final fault location is brake fault, and it is also related to the faults of oil pressure, voltage, and steel wire rope. Through this result, the brake maintenance decision is made in the actual system of the hoist, and the steel wire rope and hydraulic system are selectively maintained.

5 Conclusion

In view of the traditional methods to solve the problems of low utilization rate, low accuracy, easy diagnosis and difficult prediction of hoist fault early warning data, a hoist fault early warning parallel system is proposed. Based on big data technology, the parallel system theory is applied to the hoist system. Using the combination of virtual and real, combined with the advantages of Hadoop platform in data processing, the operation data of the hoist is analyzed and processed to achieve the purpose of fault early warning.

The combination of clustering algorithm and association rules is used to mine the established hoist database, classify the fault data, and improve the confidence of the predicted fault through association rules. The combination of ARIMA and GM (1,1) is used to predict the data of the hoist in the future.

The problem of combining hoist fault prediction with big data is preliminarily solved. Through the analysis and evaluation of parallel execution, it is verified that the parallel system theory can be used in mining machinery and equipment, which provides a certain reference significance for subsequent research.

References

[1] Zhou, C.B. (2009) Research on Mine Hoist Fault Diagnosis Based on RBF Neural Network. Harbin Engineering University.
[2] Vinutha, D.C., Raju, G.T. (2021) Budget Constraint Scheduler for Big Data Using Hadoop MapReduce. SN Computer Science, 2(4).
[3] Yin L.F., Li S.Y., Liu H. (2020) Lazy reinforcement learning for real-time generation control of parallel cyber-physical-social energy systems. Engineering Applications of Artificial Intelligence, 88.
[4] Alex, K., Dillon, K., Pinle, Q., Hamid, Z. (2019) A Hadoop/MapReduce Based Platform for Supporting Health Big Data Analytics. Studies in health technology and informatics, 257.
[5] Zaheer, S., Malik, A.W., Rahman, A.U., Khan, S.A. (2019) Locality-aware process placement for parallel and distributed simulation in cloud data centers. The Journal of Supercomputing, 75(11).
[6] Adeeb, A. (2021) Utilizing Parallel Computer Systems to Examine Seismic Reliability of Constructions. Journal of Steel Structures & Construction, 7(1).
[7] Mohammad, R.S., Mohsen, K., Samad, P. (2021) Big data fuzzy C-means algorithm based on bee colony optimization using an Apache Hbase. Journal of Big Data, 8(1).
[8] Chen, Y.C., Wang, P.P., Xu, H. (2018) Analysis and establishment of equipment fault warning model based on big data. Power Station Auxiliary Equipment, 39(04):4-7.