

Research on Application of Computer Big Data Technology in Real-time Monitoring System of Bridge Structure Safety

Lianfa Wang^{1,2,3,4, *}, Yufeng Zhang^{1,2,3,4, a}, Yu Cheng^{1,2,3,4, b}, Yichao Xu^{1,2,3, 4, c},
Jiayi Peng^{1,2,3,4, d}

¹ State key Laboratory of Safety and Health for In-service Long Span Bridge, Nanjing, 210012, China;

² Key Laboratory of Large Span Bridge Health Inspection & Diagnosis Technology Ministry of Communications, Nanjing, 210012, China;

³ Center of Data Measured by Structural Health Monitoring System of Long-span Bridge in Province of Jiangsu, Nanjing, 210012, China;

⁴ JSTI Group, Nanjing 211000, China.

*Corresponding author e-mail: wlf746@jsti.com, ^aYufengzhang2022@163.com, ^byucheng2022@163.com, ^cyichaoxu2022@163.com, ^djiayipeng2022@163.com

Abstract. The bridge structure safety real-time monitoring system based on big data evaluates the bridge safety status and real-time warning under the comprehensive consideration of various factors that affect the service performance of the bridge. The system platform uses a distributed file system with high fault tolerance and a parallel data processing engine with high computational efficiency. It has high reliability, availability and storage efficiency, and is easy to expand. Furthermore, the multi-factor analysis method is used to fully explore the implicit correlation between the real-time data of the various sensors of the bridge, and the bridge service performance evaluation model is accurately established by analysing the data correlation, and real-time warning of the bridge safety status is given. In addition, the system platform uses a variety of model verification methods to evaluate the effectiveness of the model to ensure the reliability of the big data analysis method.

Key words: Big data technology, bridge structure, real-time monitoring.

According to statistics, the actual life of a large number of bridges is shorter than the design life. Under this background, in order to more effectively monitor and evaluate the carrying capacity, operating status and durability of the bridge during operation, the large-scale bridge structure health monitoring system has attracted academic circles at home and abroad. This area of research has attracted a lot of attention in the engineering community. In order to improve the accuracy of monitoring and evaluation, the current large-scale bridge structure health monitoring system integrates modern sensing technology, network communication technology, signal analysis and processing technology, data management methods, knowledge mining, prediction technology and structural analysis theory, etc. Big data is the key technology of bridge management digitization and is highly valued in today's society. Many countries and international organizations recognize the importance of big data and regard big data as the commanding heights of a new round of high-paying technology competition. This paper presents an important perspective on how to effectively deploy and implement a big data strategy.

1. The application status of big data technology

Domestic research on bridge monitoring is later than that in other countries, but bridge health monitoring researchers have successively and quickly carried out scientific research on health monitoring systems, and have successively conducted practical research on major domestic bridge structures. And under the current wave of big data and artificial intelligence, research on bridge health monitoring combined with big data platforms has gradually emerged. Someone built an accurate finite element model of the bridge structure based on the bridge monitoring data, and proposed a method to

evaluate the service state of the bridge structure using a five-layer deep learning network [1]. There are also researchers who use dynamic networks and genetic algorithms to analyze the pattern classification technology using neural networks, conduct actual measurement analysis of its effect in identifying bridge structure damage, and prove the feasibility of the pattern classification technology [2].

2. Big data technology research

2.1 Big data storage technology

Distributed database relational database theory has brought together the wisdom of many computer scientists for decades. Nowadays, relational database systems such as Oracle, MySQL, and Microsoft SQL Server are widely used in all walks of life. For the bridge health monitoring system based on big data, a distributed relational database cluster can be built through open source middleware. The specific implementation steps are: first, install MySQL or Oracle database on each node; second, use Mycat middleware on the proxy server to connect all data nodes to manage distributed relational database clusters. Among them, Mycat is an open source distributed data storage middleware designed to solve the shortage of big data storage in traditional relational databases, and is compatible with current mainstream relational databases. The principle of Mycat lies in "interception". It can intercept SQL statements and perform specific analysis on them, such as fragmentation analysis, routing analysis, read-write separation analysis, cache analysis, etc. Then send this SQL to the back-end real database, and do appropriate processing on the returned result, and finally return it to the user.

2.2 Distributed File System

Common distributed file systems include GFS, HDFS, Ceph, FastDFS, etc. Among them, HDFS is an open source implementation of GFS, and it is also the system that this section focuses on. HDFS is the core sub-project of the Hadoop ecosystem. It is the foundation of data storage management in distributed computing [3]. It is developed based on the needs of streaming data mode to access and process large files. Its high fault tolerance, high reliability, high scalability, high availability, high throughput and other characteristics provide fault-tolerant storage for massive data and bring many conveniences to the application and processing of very large data sets. HDFS uses a master-slave (Master/Slave) architecture for data storage [4].

3. Data mode stored in the bridge health monitoring system

The real-time bridge monitoring system data volume often reaches the level, and the frequency of sensor data collection is relatively high. The daily data volume of each attribute (such as acceleration, dynamic strain, etc.) is large, and there is a strong correlation between bridge health monitoring data. It can make full use of the big data platform for data mining. The big data analysis method is mainly to establish multi-parameter, multi-source data consistent acquisition, fusion and storage technology at each stage of the bridge structure's life, forming a big data evaluation model and method of bridge service performance [5].

3.1 Data input of big data platform

Data input mainly includes numerical data, text data, drawing data, image data, etc., researching suitable algorithms for different types of data, improving the efficiency of data extraction, cleaning, fusion and loading at various stages, the goals of the big data management platform include three aspects, Including the realization of the storage function of bridge monitoring data, the realization of the query function of bridge monitoring data, the realization of bridge service performance evaluation and the real-time warning function. These functions ensure that the bridge big data analysis method

can read and write the near data volume within a second response time is relatively high. The system structure diagram is shown in Figure 1.

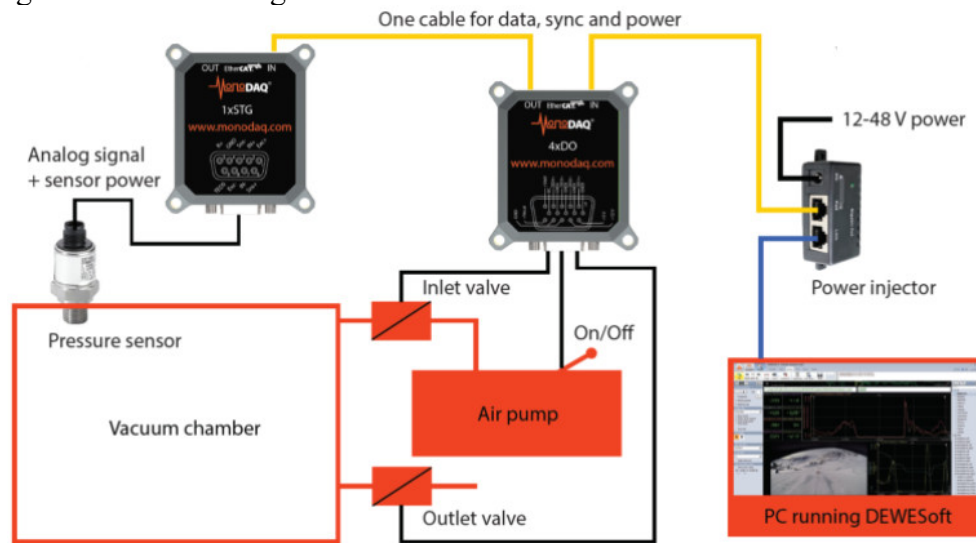


Figure 1. Bridge structure safety inspection system based on computer big data

3.2 Data standardization

Environmental and operating loads will cause autocorrelation, non-stationarity and nonlinear effects in real-time monitoring of bridge structure time series data, thereby reducing the effect of machine learning. One type of method that can be used to deal with this problem is to separate the load components from the data characteristics. This process is the main content of the structural state assessment and is also defined as the standardization of monitoring data. Due to the limited number of real-time monitoring points of the structure in the real bridge, and the randomness of the temporal and spatial distribution of environmental and operational loads, it is not easy to accurately separate various load components. Among the three load factors of temperature, wind, and vehicles that act on the structure, the temperature effect presents a significant annual and day-to-day cyclical change trend. Therefore, the data standardization of real-time monitoring of bridge structures is mostly aimed at the temperature effect.

3.3 Platform model evaluation

A series of advanced technical methods used to accomplish the above functional objectives are evaluated. First, through the dynamic capture and extraction of multi-source time-varying data such as bridge environmental load data, disaster data, design, construction, monitoring, transportation and nutrition, etc., the filtering, compression, fusion, and storage methods of bridge heterogeneous data are proposed; second, according to different data The application characteristics of the big data platform under the type, and the building of the module cluster, the component bridge big data system platform in the software environment; finally, the correlation model under the multi-factor analysis of the bridge monitoring data is proposed to form the bridge service under the big data analysis technology Performance evaluation model and early warning method.

4. Data correlation analysis and early warning model establishment

Perform multi-factor analysis on many attribute data of the bridge, dig out the implicit correlation between each attribute, and then use the regression model to fit and calculate the fitting formula. After the data relevance analysis is completed, the characteristic index is selected to construct the bridge service performance evaluation model, and the reasonable numerical interval of the bridge attribute is set, and the real-time warning is given when there is an abnormal value in the bridge monitoring data. It is divided into the following 4 parts.

4.1 Determine the correlation between bridge attributes

This part first conducts a preliminary analysis of the data, analyzes some descriptive parameters such as the mean and variance of the numerical data, and then displays it graphically to get a preliminary understanding of the laws of the data. If you find that the data has many dimensions, use principal component analysis to perform data dimensionality reduction operations. Then analyze the degree of influence between each attribute of the bridge. On the basis of combining the explicit correlation of the professional knowledge of the bridge structure, the algorithm of data analysis is used to deeply dig the invisible correlation. There are three standards for calculating the correlation degree: Spearman rank correlation coefficient, Pearson Correlation coefficient and Kendall correlation coefficient.

4.2 Establish a regression model of related attributes

This part mainly uses regression methods to fit the data and compares the fit of various methods. When performing data regression analysis, it is necessary to consider methods such as parameter adjustment to improve the fit, and at the same time conduct in-depth analysis of the data. Find out a method suitable for regression fitting to the data. The second is the evaluation of regression model, including the calculation of regression error, such as mean square error. There are two main methods for model evaluation. One is to use cross-validation; the other is to fit all the data and calculate the fitting error of all the data. The idea of cross-validation is to group the original data, select part of the data as the training set, and the other part of the data as the validation set. The first step is to use the training set data to learn the regression model, and the second step uses the validation set data to test the training model obtained. Cross-validation can be used for model evaluation methods of classification models and regression models. There are three methods for directly evaluating the regression model, the mean square error and the mean absolute value error. The mean square error is the expected value of the square of the difference between the true value of the parameter and the estimated value. It is a more convenient way to measure the average error. If the value is smaller, the prediction model is more accurate; R^2 represents the degree of change in the dependent variable explained by the regression equation. The value of R^2 is the degree of fit of the equation to the observed value; the average absolute error refers to the average degree of error between the true value and the predicted value. The schematic diagram is shown in Figure 2.



Figure 2. Schematic diagram of bridge structure safety inspection system under computer big data

4.3 Early warning design

Multi-factor analysis is performed on many attribute data of the bridge, and the implicit correlation between each attribute is explored, and then the regression model is used to fit and calculate the fitting formula. After the data relevance analysis is completed, the characteristic indicators are selected to construct a bridge service performance evaluation model, and a reasonable numerical interval of the bridge attributes is set to provide real-time warning when there is an abnormal value in the bridge

monitoring data. It is divided into four parts. First, determine the correlation between bridge attributes. This part conducts a preliminary analysis of the data, analyzes some descriptive parameters such as the mean and variance of the numerical data, and then displays it graphically to initially understand the law of the data. If you find that the data has many dimensions, use principal component analysis to perform data dimensionality reduction operations. Then analyze the degree of influence between each attribute of the bridge. Based on the explicit correlation of professional knowledge of the bridge structure, use the algorithm of data analysis to deeply explore the invisible correlation.

4.4 Establish service performance evaluation model and early warning model

The first is the evaluation of service performance. In the data analysis, the use time of the bridge needs to be considered. The load capacity of the bridge to various temperatures and pressures is the strongest in the early stage of construction, but after a long-term use of the bridge, damage will occur. The damage of the bridge needs to be considered. The initial stage of the bridge can be assumed. The damage is zero. Analyze the relationship between the bridge deflection and other properties every year. Because the bridge may be damaged every year, the relationship between the deflection and other properties will be different each year. For example, the bridge is used in the first year, a certain temperature and load. In this case, the deflection is expressed as x , and under the same temperature and load in the third year of use, the deflection is expressed as y . It is necessary to define the degree of bridge damage according to the corresponding relationship, and remind relevant units to make corresponding bridge inspections and maintenance when the bridge is seriously damaged. At the same time, it can also be combined with the deflection regression model of multi-factor analysis to verify whether the deflection data conforms to the regression model. If there is a large deviation, it will prompt the bridge abnormality. Secondly, the establishment of the early warning model needs to consider two factors: the degree of relevance of the influencing factors; the current damage situation of the bridge structure. The purpose of studying the degree of relevance of influencing factors is to obtain effective early warning indicators; and the current situation of bridge structural damage can be reflected in the residuals after the sensor data excludes the data changes caused by factors such as the environment and vehicles, and the residuals fall within a small range of changes. , If the residual error changes greatly, it means that the mechanical properties of the bridge structure are degraded or the bridge components are damaged. The early warning threshold needs to be graded, and the extreme value or root mean square value of the probability of exceeding is the normal value and the limit value. If the sensor data value does not exceed the warning limit, the bridge structure is in a safe state; if the sensor data value exceeds the limit, the bridge structure is in a dangerous state, and an alarm message needs to be sent out immediately.

5. Application of big data technology in real-time monitoring of bridge structure safety

The big data research in the field of real-time monitoring of bridge structure safety is still in its initial stage, and the content and direction of the research are still being explored. Existing research focuses more on the application of big data processing technologies such as cloud computing, and there are not many researches focusing on big data analysis processes.

5.1 The application of big data processing technology is to avoid the limitation of computing power

Traditional bridge structure safety data analysis generally uses only a certain period of time or subsampled monitoring data for analysis. This will bring about the following problems: the time resolution of the data after downsampling is reduced, and it cannot reflect the changing law of the data on a smaller time scale; only using the data of a certain time period will introduce some contingency factors, making the analysis results not universal. The data itself has a low signal-to-noise ratio, and a reduction in the amount of data will result in a corresponding reduction in the amount of

information in the real signal, thereby increasing the analysis error. Considering the above shortcomings, efficient analysis of large-volume monitoring data has always been a requirement in the field of real-time monitoring of bridge structure safety. Big data processing technology that can greatly expand data analysis capabilities through distributed storage computing has therefore attracted attention and applications. Among them, Hadoop is a software framework that supports distributed applications to build a distributed platform and perform specific functions such as storage and calculation through other program groups. mapreduce is a computing model running on Hadoop, which can split large data sets and perform calculations on the hard disk of a computer cluster to increase data processing speed; spark is another distributed computing framework that processes data in the memory of a computing cluster, Comparable to mapreduce to obtain dozens of times the computational efficiency improvement. Cloud computing is a virtualized computing method based on the Internet that is implemented through the Hadoop ecosystem, providing scalable and cheap distributed computing and storage capabilities. It is a general term for distributed, non-relational databases. Because of its flexible horizontal scalability, it can support the storage of massive multi-source heterogeneous data.

5.2 Aiming at the problem of insufficient computing power or low computing efficiency in real-time monitoring data analysis.

Some researchers have shown in the damage identification of a continuous rigid frame bridge laboratory model that the clustering algorithm through mapreduce parallelization has higher operating efficiency than the single-machine serialization; mapreduce is used to process the 2-layer frame numerical model and the 3-layer frame. The data of the test model can be used for damage identification, which can save calculation cost and improve calculation efficiency; some people use mapreduce to cluster the strain monitoring data of a continuous beam bridge to identify vehicle crossing conditions, including more than 200 million points of data. The calculation process takes less than 14 hours, and the calculation speed can be further improved by adding computing nodes; a data-driven online evaluation framework is established through cloud computing, which can execute machine learning algorithms in near real-time, and its effectiveness lies in a high-level framework. The damage identification and location of the model has been verified.

6. Research on data missing repair method of bridge monitoring big data

The phenomenon of missing data is also a common type of abnormal data in sensors. According to the missing period, it can be divided into short-term data missing and long-term data missing. The reasons for these two types of data missing are different. Short-term data loss, also known as data omission, is manifested in that the device does not collect signals for some reasons in a short period of time. For example, the sampling frequency of the temperature sensor of the Dashengguan Bridge in Nanjing is 1 Hz, so the data volume per day is more than 80,000 points. In the actual collection process, a small number of points are missing or even exceed the theoretical data volume. The main reason for this phenomenon is the instability of the input and output signals of the front-end equipment of the bridge detection system or the equipment, which causes the server's clock to be inaccurate. Short-term data loss may be the cause of the system startup. Long-term data loss means that the device has not collected signals for some reasons for a long time, and some data is not saved for one or more days. The main reason for this phenomenon is equipment failure, acquisition card failure, server crash and power failure, etc.

7. Conclusion

The bridge structure safety real-time health monitoring system platform framework based on big data is used to monitor and evaluate the service performance of the bridge in real time. Compared with the traditional closed bridge database and the processing method for bridge health monitoring

data, the proposed bridge health monitoring big data analysis platform has the following advantages: the use of distributed storage technology can effectively meet the needs of large-scale storage of health monitoring data, and is based on the big data health monitoring database, which store monitoring data across platforms and companies. Big data distributed systems have high fault tolerance and fast parallel processing and calculation speed. Therefore, the computing efficiency of bridge health monitoring data mining technology is greatly improved, and multiple factors of monitoring data can be fully considered. Therefore, the bridge service performance evaluation model can be established more accurately, and the real-time warning of the safety status of the bridge structure can be carried out.

Acknowledgements

This research was supported by the Project of Industry Foresight and Key Core Technologies (Grant No. BE2021021), for which the authors are grateful.

References

- [1] Lei Yaguo, Jia Feng, Kong Detong, et al. Opportunities and challenges of mechanical intelligent fault diagnosis under big data. *Chinese Journal of Mechanical Engineering*.12(5) (2021) 121-124.
- [2] Min Zhihua, Sun Limin, Dan Danhui. Analysis of Environmental Factors Affecting Modal Parameters of Cable-Stayed Bridges. *Vibration and Shock*.12(4) (2020) 136-138.
- [3] Jiao Meiju, Sun Limin, Li Qingfu. Reliability assessment of bridge structure based on monitoring data. *Journal of Tongji University: Natural Science Edition*.10(4) (2020) 35-36.
- [4] Li Lei, Application of Computer Big Data Technology in Bridge Inspection, *Bridge Technology*.09(05) (2018) 210-215.
- [5] Wang Liang. Research on Bridge Safety Checking and Testing Technology, *Quality and Commodity*.08(6) (2019) 52-55.