Intelligent Mining and Parallel Processing Technology of Tunnel Construction Data under Complex Geological Conditions

Si Wei2*, Men Yanqing1, Ge Zhenzhen1

1Jinan Rail Transit Group Co., Ltd., Jinan 250101, Shandong, China
2Shandong Jianzhu University, Jinan 250101, Shandong, China
*Corresponding author’s e-mail: 876027479@qq.com

Abstract. In order to achieve the safe and rapid construction of tunnels under complex geological conditions, it is necessary to fully and timely grasp the key information to carry out effective safety control. Massive information and multi-source heterogeneous data have become the bottleneck for timely and accurate risk assessment. Thus, the utilization of advanced information technologies such as the Internet of Things and data fusion is one of the important ways to improve the safety risk control level of subway tunnel construction. Building risks have a prominent role. This paper builds a tunnel risk information database and uses data mining technology to realize the structured expression of tunnel multi-source structural safety information data and apply it to engineering practice with good results.

1. Introduction
The tunnel construction in karst area of spring field is a large-scale underground engineering with complex technology. Once the accident happens in the construction process, the loss is difficult to recover, and such accidents exist in large numbers all over the world. There are a lot of key information in the process of tunnel construction, and it is the first step for in a timely manner to master these effective information. Therefore, the use of advanced information technologies such as data mining and parallel computing can effectively improve the level of safety risk control in tunnel and subway construction under complex geological conditions, and significantly improve the ability of collecting, summarizing and mining key data information in subway construction, so as to ensure the safety of tunnel and subway and to effectively reduce the construction safety risk[1-4]. At present, the main risk factors, such as mechanical indicators of tunnel surrounding rock, equipment and facilities, environment, construction personnel, can be directly captured and perceived by multi-sensor. The mutual deduction pattern and complex association between massive multi-source heterogeneous data involved in tunnel security management force the level of tunnel data mining to be further improved in the future. The massive information collected in the tunnel can be divided into numerical structured data which is easy to extract and multi-source heterogeneous data which is rich in information. Due to the lack of complete technology for deep mining of heterogeneous structure data, it is difficult to transform rich information into knowledge and rules. Therefore, this paper constructs a tunnel risk information database, uses data mining technology to comb out the tunnel multi-source structure of information data, with the help of parallel processing mode to quickly convert it into a structured and visual display mode, and applies it to engineering practice, and achieves good results.
2. Contents and methods of data mining

2.1 Data Mining Content
Data mining aims to extract key information and its inherent laws from massive data based on optimization algorithm[5]. In practical application, data mining is to analyze, sort out and convert metadata into data sets suitable for mining, and then through training and learning the specific data sets of mining model, according to different characteristics of the data set to form a "mining model" to predict its inherent laws, provide decision-making reference.

Data mining data processing has the basic idea of ordinary data processing: collection, collation, analysis, induction. At the same time, it also has its own characteristics and requirements.

(1) Massive data screening algorithm: under complex environment, the data introduced are not all accurate, and even inevitably have a variety of unexpected interference, and there are a large number of similar data, which need screening algorithm to achieve the purpose of redundancy and precision. Whether the algorithm is suitable or not directly affects the volume of data to be processed.

(2) Data processing hardware requirements: the massive data in the shield tunnel construction is that is updated, once there is a problem need to make a quick decision, which has a high demand for data processing hardware. Parallel algorithms working with clustered hardware is an important factor to determine the efficiency.

(3) Visualization requirement: Visualization presentation is the bridge between computer and manager, and also provides a solid foundation and platform for managers to make decisions quickly.

2.2 Data Mining Methods
The data mining method is the tool which manifests the data set the intrinsic rule and the logic. Usually, data mining is realized by two ways: descriptive mode and predictive mode [6]. Among them, description mode is to find rules in data information set according to category, concept and association degree, and to describe the characteristics of data set, focusing on reflecting the internal logic and relationship of data set elements. The prediction model refers to the reference of historical data, according to a certain data statistical algorithm, to find rules in historical data and predict the value of future data items. Among them, data mining methods mainly include correlation model, historical backtracking prediction model and classification and clustering combination model.

3. Intelligent data mining of tunnel construction

3.1 Tunnel construction risk information
There are many kinds of information in tunnel construction, such as displacement meter, acceleration sensor, strain gauge and so on. These together constitute a reserve database for tunnel construction. In the process of shield tunneling under complex geological conditions in spring karst area, there are many factors and data that affect the changes of stratum, such as compression of stratum, disturbance of stratum stress, lining convergence, soil consolidation, soil creep and so on. Under special circumstances, geological mutation, inadequate support at excavation, leakage of shield tail grouting, crossing adjacent bridges and other factors are likely to cause engineering accidents such as excessive soil deformation, surface subsidence exceeding the standard, underground pipeline dislocation, obvious uneven settlement of buildings and so on. These information volumes are huge and most of them are non-structural data, among which text materials are the majority.

3.2 Data Mining of Tunnel Construction Information
With the development of tunnel construction informationization, the main storage mode of safety monitoring data has realized the construction of database based on continuous and structured data model, and the agglomeration analysis and prediction of collected data with clear rules by means of mature data mining software[7]. However, in the complex geological conditions of tunnel construction, for example, in shield tunneling, the data that can really reflect the degree of risk and affect the
analysis and decision-making are often unstructured data such as text, pictures, videos and so on, in which text data is particularly important. In the previous construction, technicians face massive non-structural data, it is very heavy work to extract effective information only by manual, and can’t cope with the current complex construction environment and almost harsh time requirements. Therefore, with the help of the ever-changing cutting-edge information technology, referring to the latest bio-mathematical algorithms, relying on mature parallel hardware facilities, different types of data structures have been refined and optimized to eliminate the false and retain the true. It condenses the core data and rules, presents them in front of managers in the way of multimedia visualization, and hopefully realizes human-computer communication in the way of dialogue. The stability and safety of tunnel construction can reach a new height and perspective.

Conventional risk assessment has obvious deficiencies in timeliness and accuracy, the reason behind it is that the application of new technologies and new algorithms is lagging behind. Accordingly, based on conventional risk assessment, data mining can optimize it in the following aspects:

(1) Real-time dynamic data information is used to sense real-time safety status. The time sequence prediction algorithm in data mining can be used to predict the next data evolution trend based on a large number of historical data learning samples. Therefore, the effective data mining model of tunnel construction time sequence can deal with the complex nonlinear system that can’t be solved by conventional methods, and improve the accuracy of real-time dynamic data prediction.

(2) Reasonable analysis of multi-source structural data and comprehensive grasp of tunnel safety information. Text type of unstructured data in multi-source structured data contains more complex and hidden information, which can be studied more. For text-based data, we can mine keywords with greater weight and professional words with higher risk level, and find the internal rules and links between them, so as to achieve functional conversion. In addition, for the shield machine tunneling pictures, construction site pictures and other data, mining from the image data can reflect the law of the relevant text data, and thus realize the image target recognition, and combined with other knowledge using association rules.

(3) The collocation of parallel algorithm and hardware. Parallel computing technology adopts distributed computing mode, which can efficiently and accurately deal with large data environment, and the ability of saving time is far beyond the general single calculator. From a long-term point of view, the savings of engineering resources far exceed their input costs. It can be said that it is the best solution to save time and cost. The operation of large data can’t be separated from the software and hardware optimization, relying on the powerful hardware platform, using massive data to train optimization algorithm in advance, improve its adaptability, and make it more suitable for efficient processing of tunnel construction environment under complex geological conditions.

(4) Visual data mining process can better display the risk of tunnel construction, and present it in a simple and intuitive way, which improves the intuitive understanding of risk and the grasp of risk management and control.

3.3 Text Mining
Tunnel project in different stages of management will produce a large number of unstructured text data, such as engineering construction progress, survey report, evaluation scheme, shield tunneling machine feedback, etc. Although these texts contain a large number of effective real information, but can’t be described and displayed with structured data. At present, the research on risk management of tunnel construction based on text mining technology is nearly blank. In this paper, the text mining technology optimization algorithm and the corresponding parallel technology are applied to the actual risk management of tunnel construction.

Text mining[8-10] means to retrieve a large number of unstructured text information, and to obtain the effective information and inherent rules hidden in it, and to use the rules to guide future action. Text mining technology involves many knowledge categories, including: multi-dimensional informatics, intelligent digital technology, database retrieval technology, statistical analysis, text
recognition, data mining fusion and many other fields. Text mining process can be divided into five steps, followed by: text data collection-> preliminary processing-> reduction dimension processing-> data analysis mining algorithms-> visualization of the results and other components.

Text mining technology has a clear application-oriented. Because the general lexicon is not able to analyze and understand the text data in the professional field comprehensively and accurately, it is necessary to choose the appropriate algorithm to construct the proprietary lexicon for this field. In order to realize the expectation of text mining efficiently and accurately, this paper constructs an optimized text mining segmentation thesaurus which is very suitable for tunnel construction risk information data under complex geological conditions, as shown in Figure 1.

![Fig.1 Structure of unstructured text data mining system](image)

4. Engineering Application
From investigation and design to tunnel construction preparation, a large number of text information will be generated, among which the report data that can fully reflect the tunnel hydrogeological risk status and other relevant information include: "Tunnel Static Risk Assessment Report", "Periodic Inspection Report", "Expert Demonstration Conference Summary Report" and "Tunnel Site Hydrogeological Survey Report". Base the above material on a mature professional word segmentation package (such as wordCloud2 for R) convert to plain text files. In actual processing, the frequency of words that can reflect the overall situation of the text content will be relatively high, extract the top high-frequency words according to experience, and display the results according to the word cloud package.

Need to pay attention to is: the results of the analysis if only for a single risk factor, which is inconsistent with the actual situation. In fact, the risk of accidents often leads to a number of risk factors, but the weight is different. Based on the actual situation, further optimization can be made: the introduction of co-occurrence analysis method in the text mining analysis method to deal with the co-occurrence of risk sources. The so-called co-occurrence analysis refers to the use of a reasonable algorithm to count the frequency of two or more risk source keywords in the same file set, and give them appropriate weights, so as to reveal the coupling relationship with the occurrence of dangerous situations.

5. Epilogue
In this paper, the advanced intelligent information algorithm and technology such as data mining are used to analyze and process the multi-source non-structured data rich in tunnel construction information, and the structured application is realized quickly with the parallel algorithm.
Visualization of tunnel safety information is realized by using visualization information processing algorithm, and applied to engineering practice, which can effectively improve the ability of real-time analysis and mining of safety information in the process of tunnel construction, and effectively reduce the risk of construction safety. It is embodied as follows:

1. Active perception of real-time security status using dynamic and real-time data information. According to the effective tunnel construction time sequence data mining model to deal with conventional methods can’t solve the complex nonlinear system, and the introduction of parallel algorithms and hardware with the operation, from both hardware and software point of view to improve the accuracy and effectiveness of real-time dynamic data prediction.

2. Based on multi-source structure data, fully grasp the key information of the tunnel. Reasonable combination of text mining technology and tunnel construction safety risk management and control, mining the keywords with greater weight and professional words with higher risk level, find the inherent law and relationship between each other, and quickly put forward early warning strategies.

3. Visualization of data mining process. Tunnel construction risk information visualization processing, can really affect the tunnel construction management index of complex data in the form of visual graphics surface with the management and technical personnel and can effectively reduce the difficulty of tunnel construction dynamic risk management.

References

[1] Wang Mengshu. Tunnel and Underground Engineering Technology and Its Development [M]. Beijing Jiaotong University Press, 2004.

[2] Li Shucai, Zhou Zongqing, Li Liping, et al. Risk assessment of water inrush in karst tunnels based on attribute synthetic evaluation system[J]. Tunnelling and Underground Space Technology, 2013, 38: 50-58.

[3] LI Shucai, XUE Yiguo, ZHANG Qingsong, et al.Study on key technology of comprehensive prediction and early warning of tunnel construction geological hazards in high risk karst area [J].Chinese Journal of Rock Mechanics and Engineering, 2008,27 (7): 1297-1307.

[4] Liu Shang. Dynamic Safety Risk Assessment of Tunnel Construction Process Based on Data Mining and Its Engineering Application [D]. Shandong University, 2018.

[5] Wang G H, Jiang P.A survey of data mining [J] .Journal of Tongji University (Natural Science Edition), 2004,32 (2): 246-252.

[6] Liu Hongyan, Chen Jian, Chen Guoqing. A survey of data classification algorithms in data mining [J]. Journal of Tsinghua University (Natural Science Edition), 2002,42 (6): 727-730.

[7] Xu Feng, Wang Yang, Du Juan, et Al. Study on landslide displacement prediction model based on time sequence analysis [J]. Chinese Journal of Rock Mechanics and Engineering, 2011, 30 (4): 746-751

[8] Song Y, He J, Shu HP, et al.Research on disease symptom weight mining based on word frequency and inverse document frequency method of text mining [J] .Journal of Chengdu Institute of Information Engineering, 2014,29 (1): 52-58.

[9] ZHANG Rong-yu. Research on Space Engineering Technology Risk Classification Based on Text Mining [D]. Harbin Institute of Technology, 2016.

[10] ZHANG Wen, TANG Xijin, YOSHIDA Taketoshi, et al. An approach to Web information processing based on Web text miningAIS, text mining-based enhanced Web Information Processing Technology [J]. systems engineering theory and practice, 2010,30 (1): 96-104.