A Hybrid Structural Building Data Mining System Based On the Data Tracking Through Devices

A Sangeerani Devi*, B Ravi Prasad2, Venkateshwaran Loganathan3, Swapnil Saurav4

1Department of Computer Science and Engineering, Sri Sairam Engineering College, Chennai, Tamil Nadu, India
2Department of Computer Science and Engineering, Marri Laxman Reddy Institute of Technology, Dundigal, Telangana, India
3Department of Information Technology, KGiSL Institute of Technology, Coimbatore, Tamil Nadu, India
4Department of Problem Management, Service Now Software India Pvt. Ltd, Hyderabad, Telangana, India
Email: sangeerani.cse@sairam.edu.in1*, rprasad.boddu@gmail.com2, anandvenkat4@gmail.com3, swapnil.saurav@gmail.com4

Abstract. One of several traditional megaprojects is underground construction, given its long building time high building expense and possible risks. In tunnel engineering, trench boring devices are generally used to increase work performance and safety. During the tunnelling process, system has recorded vast volumes of tracking data to ensure building safety. The processing of vast real-time surveillance data also lacks successful techniques, and, in many situations, it must be performed manually that pose possible safety hazards. This paper suggests an approach for hybrid data mining (DM) to automatically process the TBM data for real-time tracking. Three separate DM strategies are merged in order to improve the operation of mining also to help security management. The sequential pattern method is executed to remove connections between TBM parameters in order to give people the expertise needed for an irregular on-site judging. A random forest model is built to identify training data in order to complement knowledge needed for building decision-making system. Finally, neural network models measure the penetration rate (ROP) in order to detect irregular data and to alert early. In the case of a tunnel project in China, the suggested technique was applied, and the findings of the application concluded that the approach offered a reliable and effective way of evaluating TBM protection management data in real time during buildings.

Keywords: Machine Learning, Data Mining, tracking, Association rules, ANN, KNN

1. Introduction

The large-scale high-cost and long project life cycle of tunnel boring devices has been commonly used as an important method for tunnel construction and the entire project is often linked with complicated technological challenges and possible risks [1]. Modern TBMs generally implement a range of methods of sensing and tracking, recording operational parameters [2] such as tension, current, gas pressure, flow, etc. These tracking information display the operating status of the TBM also can be used to minimise building risk, improve building activities and ensure building safety. By developing wearable sensors, cyber-physical systems have been developed for productive data Control [3], namely automated recording, storing and display of the data.
The microscopic examination however is contradictory and raises risks for the creation of a tunnel. There have also been several catastrophic building incidents due to the lack of a timely and precise analyzed result. [4] In a tunnel in China, Guangdong, on February 2018 a massive water inrush crash resulted in over 10 deaths and huge economic damages. The cause of the incident was that the Blade penetrated a leaky layer not found by environmental analysis. This might be stopped if for example real-time tracking data were evaluated if the development irregularities were detected in advance. However, in terms of the following features of TBM data monitoring, real-time automated data collection is also met with many difficulties.

It is hard to evaluate factors pertaining to important parameters. Tunnel design must report data for all its facets and subsystems, in order to minimize risk, as Megaprojects of immense uncertainties. There can be over a hundred monitoring parameters and the interaction between them is complicated to remember and manually model [5]. It is problematic to classify altogether the associated variable quantity from a large range of parameters if important parameters are to be estimated and evaluated. This impacts the precision of the data analysis.

Soil variability does not have reliable results. Geological knowledge plays a critical role in the design of underwater tunnels in relation to projects above ground level. However, geological data are collected by means of borehole sampling and formations are typically measured using a linear fitting between the sampling points. The findings vary often from the real situation and impact the precision of the data analysis. There can also be no sampling and finding of some poor formations which carry potential pitfalls to the tunnel design process.

For data processing and modelling, the dynamic non-linear relation between monitoring variables is difficult. As a TBM functions, the soil, shield in addition internal machinery interplay, complicating and tough to explain the correlation between limitations in a simple formulation [6]. Consequently, the mathematical research techniques widely used do not work well.

This research aims at addressing these issues and at ensuring that TBM tracking data are processed automatically and effectively to ensure construction safety. DM methods are provided in this paper to solve this problem. DM is an empirical data-driven approach that is particularly appropriate for extracting necessary patterns from big, rapidly changing and complicated data [7]. In Engineering Projects DM approaches have been validated to support the importance of data monitoring [8].

However, due to stringent demands on performance and precision, implementation of DM has met challenges in the field of TBMs. This paper suggests a new hybrid DM approach to solve these difficulties. This paper includes the five topics that follow: (1) Association guidelines are drawn up to complement the expertise of individuals and help to judge irregularities. (2) Training data is refined to assist TBM operator’s decision-making by classification review. (3) ANN model measurement of TBM results can be achieved by measuring the assessment value to real value and an unhealthy condition can be observed. (4) The current DM algorithms were modified to increase performance on the basis of TBM characteristics. (5) In addition to solving individual problems, the three DM algorithms are paired with hybrid DM architecture.

Enhance the method of data mining. The suggestion instructions derived are implemented in subsequent algorithms for the collection of criteria, besides the confidential formation is used to maximise data output in RPM forecasts. The performance and precision of the algorithms is tested by a case study and strengthened by the proposed hybrid DM architecture. The mining findings are integrated to promote safety management in tunnel engineers, including construction rules, refining preparation, and updated ROP.

2. Related Work
As far as safety tunnelling is concerned, how effectively to measure the efficiency of TBMs is still one of the most difficult problems for both clinicians and academics. Precise evaluation of results in TBM will help pick the machine and estimate the length of the project, thus reducing the expense of projects.
Generally, the performance estimation of TBMs includes predicting those parameters, these included ROP and advanced (AR), while rock properties and system parameters are used in the input variables [10]. The prediction methods can usually be categorized into three categories: conceptual, analytical and computational. In a recent study, Rostami [10] has established theoretical and methodological approaches.

Models focused on force balance between rock, cutter and external machine structure establish theoretical methods. The Colorado College of Mining (CSM) proposed framework by [11] is one of the most commonly used theoretical models. The model is founded on rock wounding theory through disc cuts besides considers the effect of rock dynamics, disc cutter geometric parameters in addition slashing limitations. The model is based on by adding additional rock mass parameters in [12] updated the CSM model.

The numerical methods are focused on construction know-how involving multiple laboratory experiments, field observations and building records. The Norwegian Science and Technology University (NTNU) model [14] has been widely used as an observational model. The NTNU model has been tested and improved in a variety of respects, with Macias doing the new work [15]. Another common analytical template to predict the success of TBM is the Field Saturation Index in [16] and have recently modified the model.

In recent years, the use of numerical computational methods has attracted ever more exposure to the literature, with the exponential growth of computer theory and equipment. Nonlinear regression analyses are a simple but powerful method of simulation that allows an adaptation formula to be obtained between TBM output and input parameters. [13] Artificial neural networks (ANNs) are a common way of designing models for their strong teaching and nonlinear fitness. In the meantime, fuzzy logic has already become a neuro-fuzzy approach in modelling and was further merged with ANN. In addition, it was attempted to forecast the efficiency of TBM by means of other procedures such as supporting vector machines (SVM), programming of gene expression (GEP) and optimization of particles swarm (PSO). In the TBM efficiency forecast, [5] analysed implementations of Soft Computing technologies and considered soft Computing methods to be effective in handling complex connections between TBM parameters.

The observational data also represent any other workplace practices of TBMs that are worthy of consideration in addition to evaluation of TBM results. The measurement of the soil-to-TBM interaction will forecast system risk and atmosphere risk and lead to optimizing the architecture and the selection of TBMs. The Cutting Forces besides energy cost of disc-cutters Acaroglu established a fluffy logic model. The model was created on practise also linear cutting experiments, with rock characteristics and mechanical measurements included input parameters.

In order to measure the scale of the heavy factor besides its time and space delivery by elimination of the reported data from TBM Festa et al. have developed a provision model. In addition, the Random Forest Algorithm has was cast-off to propose a complex load forecast model. Excavation in soft soil can central to settlements that impact the landscape also buildings around it. It is suggested that the theory of soil motions be related to complex and geometric features of TBM. And the soil redistribution can then be accomplished by means of data records of amplitude and distribution. Geological knowledge and numerous approaches like ANN are a crucial consideration for tunnel safety management.

The prediction of geological formation, based on operating TBM results, was suggested, and the supporting vector classifier (SVC). Shi et al suggested clustering the TBM tracking data by a flouted c-means algorithm.

The data clusters will then extract knowledge about training and organizational behavior. Furthermore, it developed a model decision tree for the creation of a method to identify rocks. In the TBM performance assessment the framework may have a classification criterion for geological results. Ses experiments have been used to show the effectiveness and convenience of TBM data tracking research in shield tunnelling automation. DM is a series of technologies for data mining to retrieve unknown facts from large data sets. DM algorithms can be classified further into different groups.
according to mining functions, such as grouping, regression, clustering besides exploration of relations. DM involves the general method of information collection, statistical analysis, DM algorithms implementation and mining interpretation.

A significant volume of data has long been gathered for the architectural engineering and building industry (AEC). Sensors also Internet of Things (IoT) have expanded the information volume additionally. Sustainability, optimizing the procedure, intelligence acquirement, and risk recognition and costs savings are some significant factors for practitioners to implement DM. At present, DM is implemented during the lifecycle of the house.

During the design process DM strategies were used by Kim et al. to determine the design choices for houses. Multiple building materials, namely roofs, walls, heating and ventilation system (HVAC) and building orientation were tested for an energy effect. The findings were assessed. The main decorations were created to assistance the management team enhance the construction of the house. In addition, it is used DM approaches to forecast structure energy efficiencies. In order to excerpt information from text strategy data a semantic graph has been developed. Applied to the association rule algorithm, the relation among options available and building efficiency was investigated after construction.

During design, DM was used to determine building risks and to classify building faults and to predict capital expenditure. In order to estimate the likelihood of deaths, Cheng et al. used a conclusion tree model. The parameters used in the model are the type of accident, type of project, age, gender, etc. The approaches ANN besides Case-Based Reasoning (CBR) have been used to measure megaproject development danger. The model design has taken into account heterogeneous data and several parameters, and Latent Clustering Analysis (LCCA) was introduced in order to decrease the data size. Finally, the model generates the danger level of such building actions.

In the study of associations between design faults and inspection markers, Lin also Fan used the suggestion rule methodology. The regulations could help inspectors concentrate better on key indicators. DM showed positive findings in the area of energy efficiency measurement in the operation and maintenance (O&M) process. The Decision Tree Model was used by to investigate what climate response approach would make an indoor atmosphere pleasant. The layout of the random forest produced should easily illustrate each factor's effect. Ashouri et al. employed an energy savings advisory hybrid data mining technique. In order to examine the energy consuming reports of occupants and provide advice, Association rules and cluster analyses were used. ANN was eventually interested in estimating the energy saved. DM techniques may be used for maintenance reports in addition to its high measurements. Innovative studies on removing secret trends from keeping archives of large public buildings through hybrid DM techniques has been carried out by Peng et al., in addition Wen et al. The regulations derived should be made available as guidelines to building administrators.

3. Proposed System
For tunnel construction data, a number of models and algorithms have been used for the modelling of the complex non-linear TBM relationship. In certain cases, these data-driven approaches have increased the performance of tunnelling and shortened operating hours. In today's literature, however there are many shortcomings. Many TBM presentation estimate study is not intended for real-time review. The information rummage-sale in these experiments was confined mostly to rock and TBM geometry with little real-time information tracking. The analytical method often requires a lot of manual involvement and cannot satisfy real-time analytical quality criteria. This implies, in its place of real-time monitoring of TBM position, these studies can only be used for analyses before or during a mission. Figure 1 explains about proposed framework with process flow.
Furthermore, in real time, the tunnel building documents many parameters. The choice of parameter is based, however, primarily on familiarity and a lack of comprehension of the significance of these parameters. It likewise arises that similar parameters are not taken into account, thus reducing the model correctness. Furthermore, TBM training parameters affect the efficiency of the tunnel. At present, however only borehole sampling can obtain training data which makes the knowledge about formation among sampling points imprecise. Since DM's output relies on data consistency, this inexact data limits DM's accuracy.

These studies are often planned mainly to solve with a single approach a certain predictable problem. In literature it is unusual to see a structural analysis of TBM parameters and their relationships. There is still no standard synthesis of various analytical approaches for the resolution of complex problems. Therefore, the effect of different criteria dependent on real-time TBM tracking data needs to be considered in new approaches. In AEC programmes, several research literatures display the utility of DM. The manual involvement in the pre-process and post-process steps always considered that DM is a half-automated system. Reduced manual interference can minimize DM productivity and lead to mining outcomes becoming unpredictable. Some scientists have suggested ways to swap guide work in the existing DM phase by using another DM technique; in addition this is the central concept of hybrid DM methods.

As pre-process for identifying usable sub datasets for the principal DM procedure, the traditional Hybrid DM approach uses one before additional DM procedures. These sub datasets have higher data quality and more visible functionality, with better results predicted. Indeed, multiple studies suggest that the hybrid DM method has benefited. Unattended algorithms will first be carried out including classifications and forecasts, as will association laws and cluster analyses, followed by supervised algorithms. Early analysis is assumed to better explain data dynamics, define data features, and indicate recommendations for detailed analysis.

This study therefore suggests a hybrid DM method to solve the above-mentioned problems and guarantee the safety of TBM building. Multiple shortcomings can be solved with the implementation of the hybrid DM approach: (1) approaches of high labor involvement cannot meet real-term data collection performance criteria. The hybrid DM method substitute for manual work with automatic DM algorithms in data pre-processing and post processing that increases performance and permits real-time data analysis. (2) Experimental parameter collection will minimize the exactness of the effects of the study. This mechanism is replaced by the suggested hybrid DM approach with a pre-
process algorithm of Association rule discovery. Instead of interactions, parameters can be chosen to get a decent outcome based on rules of association. (3) DMs impact can be influenced by incorrect training results. A classification algorithm is designed to calibrate formation details in the hybrid DM context to solve this problem. (4) The bulk of experiments are directed at addressing human problems in particular.

The implementation of the hybrid DM algorithm meets several protection managements goals. These goals are not independent. It should be remembered. The results of the preceding analytical activities will help the following DM processes. The suggested solution gains from the integration of many DM approaches and achieves increased data processing precision and performance. In the production of the hybrid DM system, the characteristics of TBM are also recognized. These functions are used to strengthen the general DM mechanism through prior experience. The following section explains the structure of the hybrid DM process. Three goals were accomplished by the proposed approach in diverse areas of TBM protection management. A centralized database was first developed which incorporated data monitoring and training knowledge.

The analyses of TBM surveillance data were undertaken by three DM approaches with various security management objectives. The DM general process was followed by each step, including data selection, statistical analysis, core implementation of algorithm, and analysis of the results. The Association Legislation where first discovered to improve people's comprehension of TBM organizational rules. Multiple parameter data control was entered in the model besides parameter association instructions were removed as production. In the meantime, previous information of TBM limitations has been assumed to maximize the reliability of calculations. The mining consequences show the correlation between TBM parameters besides might also be used to warn early of irregular changes.

If modifications to TBM parameters are incompatible with association rules, there can be inconsistencies and early warnings should be made. For choosing parameters for the following method, the extracted rules can also be used. The dynamic relationship between land and TBM makes it impossible to thoroughly understand the factors influencing response variable. Classification algorithms display theoretically related TBM limitations that can be used to pick DM models contribution parameters. For e.g., the complicated parameters only need to be evaluated in connotation rules that include the development when deciding the contribution parameters of the development organization model.

The training classification review was then undertaken to resolve the issue of incorrect training. Classification model input parameters can be calculated by apriori algorithm to generate a better performance. The qualified organization model could be used to deliver information for persons that is important for TBM security management in the decision-making process. When calibrated formation varies from the findings of borehole samples, operators can carry out diligent checks to ensure protection and take suitable building strategy. The standardized creation will also deliver improved information for ROP and the performance of the ROP forecast model is predictable to recover.

Finally, an ANN to execute the ROP prediction was chosen in real time. The design was also chosen according to the assignment rules, and the measured data for the formation were used to increase the precision. During the modelling process, the monitoring information characteristics were analyzed besides a special system configuration was developed that took into account two neighboring follow-up documents. The ROP may be used as a benchmark for the definite data monitoring to aid through building conclusions.

4. Results and discussions

The protocol of interaction law was performed first. A total of 210 lawful mining database data points, each of which had 29 attribute objects to be mined, were obtained as feedback. It is separated into 2 classes of previous expertise for improving the mining algorithm, based on the types of TBM results. The primary category of information is holistic, directionally free information, while the middle section is directionally and spatially related data. There is no connection between the two parameter
sets owing to spatial symmetry. A measure and an error mechanism assess the support and trust threshold.

Association rules are increasing, suggesting that the rules derived are more common, but the number of rules is diminishing. Increased trust will also increase the average trust and consistency of Association Regulations, but reduce the number of Association Regulations. In the solution suggested, the threshold for support and trust is 0.17 and 0.8 to equalize laws, subjectivity and correctness. The programme, which indicated the dynamic relationship between TBM parameters, found a entire of 102 strong suggestion instructions following calculations. The extracted rules of association will give citizens a deeper understanding of building regulations and promote safety management.

Some traditional mining findings the first "total thrust of ROP" rule for the association revealed that ROP was typically wide when the shove was high, which was a check of people's experiences. The inverse statute 'ROP for the entire thrust,' however is not a good rule of association as assurance is less than the minimum. This is because other variables, such as forming and grouting, also influence ROP. The conclusion goes against the intuition of people and corrects misconceptions. The third rule "Cut Torque — Total Torque" indicates two not explicitly connected ties. TBM is always in hard form when a few torques’s is high, so the law is rational. And operators must raise the thrust to sustain penetration in hard training. The rule demonstrates rules which people have not understood. Figure 2 discussed about comparative result with numbers of rule and supports

These organization guidelines validate the expertise of individuals, fix mistakes and reveal new regulations, and can be used to guarantee stability in the sense of a rapid evaluation of on-site anomalies. Certain complex relationship laws can provide more details and can be used for further quantitative simulation. The ROP was connected to nine parameters at the same time, as seen in the third law. In reality, the candidate parameters are to be entered for the ROP prediction in the neural network model.

5. Conclusion
In this article, a highly reliable and successful TBM tracking data mining technique is suggested to achieve effective and precise DM data mining during tunnelling. Three DM approaches are merged to enhance the mining procedure besides to excerpt valuable designs to facilitate security organization: (1) Association regulation exploration is used for drawing parameter relationships and the rules derived for the selection of a parameter for subsequent algorithms are applied. (2) Organization examination may be used to approximate present development besides advanced preparation is used for improving ROP prediction data consistency. (3) Versions of ANN are designed to verify the present ROP on the basis of classified testing and data tracking in real time. The method suggested is tested by the data from a tunnel project in China in real time and the outcome shows that it is successful in security management. 102 association rules are derived by an updated Apriori procedure to give people a improved thoughtful of TBM. These guidelines show building regulation, which can be used as gained
knowledge for early detection of anomalies. The formation of a better CART model is graded in order to perfect the formation among exploration boreholes. Sophisticated formation increases geological data precision besides can help operatives follow effective training techniques to ensure protection. Two ANN models are industrialized in order to test ROPs in real time. In an irregular case, the evaluated ROP may be used as a measure for real meaning. The approach suggested offers a viable way to analyze TBM tracking data in real time. The findings of the study will provide credible support for the decision of the project and develop construction methods. The suggested DM hybrid approach is highly reliable and effective and satisfies real tunnel projects' requirements. The findings of the study will assist in the decision making and security administration of tunnel building in a consistent manner.

References

[1]. Benardos, A. G., & Kaliampakos, D. C. (2004). Modelling TBM performance with artificial neural networks. Tunnelling and Underground Space Technology, 19(6), 597-605.
[2]. JAVAD, G., & NARGES, T. (2010). Application of artificial neural networks to the prediction of tunnel boring machine penetration rate. Mining Science and Technology (China), 20(5), 727-733. Mikael, R., Naghadehi, M. Z., & Sereshki, F. (2009). Multifactorial fuzzy approach to the penetrability classification of TBM in hard rock conditions. Tunnelling and Underground Space Technology, 24(5), 500-505.
[3]. Grima, M. A., Bruines, P. A., & Verhoef, P. N. W. (2000). Modeling tunnel boring machine performance by neuro-fuzzy methods. Tunnelling and underground space technology, 15(3), 259-269.
[4]. Mahdevari, S., Shahriar, K., Yagiz, S., & Shirazi, M. A. (2014). A support vector regression model for predicting tunnel boring machine penetration rates. International Journal of Rock Mechanics and Mining Sciences, 72, 214-229.
[5]. Naghadehi, M. Z., Samaei, M., Ranjbarnia, M., & Nourani, V. (2018). State-of-the-art predictive modeling of TBM performance in changing geological conditions through gene expression programming. Measurement, 126, 46-57.
[6]. Elbaz, K., Shen, S. L., Sun, W. J., Yin, Z. Y., & Zhou, A. (2020). Prediction model of shield performance during tunneling via incorporating improved particle swarm optimization into ANFIS. IEEE Access, 8, 39659-39671.
[7]. Acaroglu, O. (2011). Prediction of thrust and torque requirements of TBMs with fuzzy logic models. Tunnelling and Underground Space Technology, 26(2), 267-275.
[8]. Festa, D., Broere, W., & Bosch, J. W. (2012). An investigation into the forces acting on a TBM during driving–Mining the TBM logged data. Tunnelling and Underground Space Technology, 32, 143-157.
[9]. Sun, W., Shi, M., Zhang, C., Zhao, J., & Song, X. (2018). Dynamic load prediction of tunnel boring machine (TBM) based on heterogeneous in-situ data. Automation in Construction, 92, 23-34.
[10]. Broere, W., & Festa, D. (2017). Correlation between the kinematics of a Tunnel Boring Machine and the observed soil displacements. Tunnelling and Underground Space Technology, 70, 125-147.
[11]. Zhao, J., Shi, M., Hu, G., Song, X., Zhang, C., Tao, D., & Wu, W. (2019). A data-driven framework for tunnel geological-type prediction based on TBM operating data. IEEE Access, 7, 66703-66713.
[12]. Zhang, Q., Liu, Z., & Tan, J. (2019). Prediction of geological conditions for a tunnel boring machine using big operational data. Automation in Construction, 100, 73-83.
[13]. Shi, M., Zhang, L., Sun, W., & Song, X. (2019). A fuzzy c-means algorithm guided by attribute correlations and its application in the big data analysis of tunnel boring machine. Knowledge-Based Systems, 182, 104859.
[14]. Shi, M., Zhang, T., Zhang, L., Sun, W., & Song, X. (2020). A fuzzy c-means algorithm based on the relationship among attributes of data and its application in tunnel boring machine. *Knowledge-Based Systems, 191*, 105229.

[15]. Salimi, A., Rostami, J., Moormann, C., & Hassanpour, J. (2018). Examining feasibility of developing a rock mass classification for hard rock TBM application using non-linear regression, regression tree and generic programming. *Geotechnical and Geological Engineering, 36*(2), 1145-1159.

[16]. Anand, S. S., & Hughes, J. G. (1998, April). Hybrid data mining systems: the next generation. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 13-24). Springer, Berlin, Heidelberg.