Prediction of TBM Operational Parameters Using an Integrated Data Mining Framework

X Wang1,2, M Zhu3 and Y Shen3

1Department of Civil and Environmental Engineering, University of Wisconsin-Madison, 1415 Engineering Drive, Madison, WI 53706, USA
2Department of Computer Science, University of Wisconsin-Madison, 1210 W. Dayton Street, Madison, WI 53706, USA
3Department of Geotechnical Engineering, College of Civil Engineering, Tongji University, 1239 Siping Road, Shanghai 200092, PR China

xwang2463@wisc.edu

Abstract. Tunnel Boring Machine (TBM) is widely used in tunnel construction because of its high safety, less effect on surroundings and rapid excavation speed. However, its performance is highly dependent on the adjustment of operational parameters during tunnelling. In many cases, the project delays even the accidents take place due to the mis-operation of inexperienced drivers. To improve the adaptability of a TBM in complex geological conditions, this study proposes an integrated data mining framework to perform near real-time prediction of two operational parameters including thrust and torque. The integrated framework provides a set of data processing methods including data cleaning, partition of full tunnelling cycles, feature extracting and model establishment. These data mining methods were applied to analyse the in-situ data of a water conveyance project. The results showed that the proposed framework performed well in predicting those two parameters with the determination coefficient R^2 all exceeding 0.9, which illustrated the feasibility of using the proposed framework to assist driving in TBM construction.

1. Introduction

Tunnel Boring Machine (TBM) has become an essential part in long infrastructural tunnelling since 19th century, because of its high driving performance [1]. Besides, the low extra-excavation and low manpower are assured along the highly mechanized processes of TBM [2]. Basically, the successful application of TBM highly depends on the correct selection of operational parameters, which are mainly dependent on the drivers’ experiences. Therefore, the tunnelling efficiency and cost are closely related to the drivers’ control of several critical operational parameters. In many cases, the project delays even the accidents take place due to the mis-operation of inexperienced drivers [3]. Lots of research have been conducted to manage the risks caused by the mis-operation of inexperienced drivers. They either relied on experimental studies [4,5] or prediction models based on in-situ operational data [6,7]. The experimental studies were typically focused on depicting the interaction between cutterhead and rocks to provide guidance for drivers. Recently, the employment of in-situ operational data to predict the best alternative values of the critical parameters during TBM construction is becoming prevalent. Case-by-case studies were proposed to study the interaction between geological conditions and TBM performance, such as the Field Penetration Index (FPI) [6] and Torque Penetration Index (TPI) [8].
Although the previous studies give insights on the optimization of the critical TBM parameters based on the data tunnelling cycles, their limitations are obvious. First, most studies ignore the process of data cleaning or roughly handle the raw data by experience. Second, most studies roughly apply the tunnelling cycles or excavation time to divide the monitoring data for the subsequent analysis and prediction. These limitations may induce unreliable prediction results. Although machine learning and deep learning methods performed well in some practical applications [9,10], it is important to ensure the effectiveness and correctness of the learning process.

The objective of this study is to recognize and learn different patterns of each tunnelling cycle, which are closely related to the interactions between rock mass and TBM. This study proposes an integrated data mining framework to perform near real-time prediction of operational parameters based on the historical sensor data near the foresight areas. The proposed framework integrates data cleaning, partition of tunnelling segments, feature selection, statistical variable selection and model establishment. Two significant TBM operational parameters including thrust and torque are selected to be trained and predicted. The results showed that the proposed framework performed well in predicting those two parameters with the determination coefficient R² all exceeding 0.9, which illustrated the feasibility of using the proposed framework to assist driving in TBM construction.

2. Related work

2.1. Experimental studies

Experimental studies have been conducted to analyse the interaction between the surrounding rock mass and the TBM performance to provide guidance for drivers. For example, Colorado School of Mine (CSM) [11] and Norwegian University of Science and Technology (NUST) [12] have carried out a series of experiments and proposed the most widely used models. The results indicate that the rock stress can be positive or negative for the boring depending on the stress level, and the magnitude and orientation of the stress anisotropy. Besides, rock fracturing is a geological factor that has the largest influence on the net advance rate and thereby the tunnelling cost.

2.2. Prediction models

Using in-situ operational data instead of traditional mechanical analysis to predict the best alternative values of the critical parameters during TBM construction is prevalent recently. Delisio et al. [6] implemented a mapping between FPI and two rock mass parameters via a multivariate regression analysis. Hassanpour et al. [7] provided empirical equations and charts for estimating the performance of TBM by statistically analysing the correlation of rock mass characteristics with FPI. Tóth et al. [5] proposed a TBM performance model fitted in with heterogeneous ground conditions and the penetration rate was selected to represent the behaviour of the TBM. Ghasemi et al. [3] utilized a fuzzy logic model to predict the penetration rate of TBM with four kinds of rock property parameters. Gao et al. [13] used the recurrent neural networks-based (RNN-based) predictors to deal with the real-time TBM in-situ operating data including the torque, thrust, velocity and pressure of chamber.

3. Proposed framework

In order to realize the learning process in near real-time situation, this study faces with three challenges: (1) the length of the historical data should be short; (2) the dimension of the training data should be small; (3) and the complexity of the learning algorithms should be minimized. In this section, a framework integrating four main steps is proposed to meet these requirements.

3.1. Data cleaning

The raw data of the advance rate in one typical tunnelling cycle are shown in Figure 1 as an example. A successful tunnelling cycle consists of three segments, namely, shutdown segment, ascending segment and stable segment. The first step of the framework is to clean the raw data. Redundant data including the shutdown segments need to be removed and the noise data existing in useful information need to be eliminated.
Figure 1. The variation of advance rate with time in a full tunnelling cycle.

To remove the shutdown segment, we should identify the starting point and terminal point of the shutdown segment in a full tunnelling cycle firstly. In this study, to compress the invalid information as much as possible, the thrust and torque are employed as indexes for removing shutdown segments. The procedure to remove the shutdown segments is conducted according to Equations (1) and (2). Let $F$ and $T$ denote the value of thrust and torque, separately. And $F(X) = [F, T]$ is the indicative function. The set of the data are considered to be the shutdown segments when $F(X)=0$, which are demanded to be removed from the database.

$$ F(X) = f(F) f(T) $$

$$ f(x) = \begin{cases} 0, x = 0 \\ 1, x \neq 0 \end{cases} $$

After removing the shutdown segments, the abnormal data should be deleted as well. Taking Figure 2 as an example, when the excavation time is around 410 s and 1320 s, the outliers appear. This practice may be caused by machine failures or human errors. Based on statistical calculation, the absolute value of skew in this database is less than 3, while the absolute value of kurtosis is less than 7. It indicates that the shown data obey approximate normal distribution [14]. Therefore, based on “3σ criterion”, the upper bound and the lower bound of the truncation line is 157.15 and -41.76 respectively in this excavation section. As shown in Figure 2, the red dotted line is the upper bound of the “3σ” truncation line which effectively identifies the abnormal points.

3.2. Partition of full tunneling cycles

The partition of ascending segment has barely appeared in other studies. However, in different ascending segments, the parameters of a TBM will show distinct characteristics and trends because of geological changes. It is easier for machine learning algorithms to capture data characteristics if the variables in a specific ascending segment are employed as inputs. Therefore, the division of ascending segment is conducive to forecasting the variables in the stable segment.

As shown in Figure 3, the ascending segment can be divided into two parts: empty pushing segment and trial excavation segment. There are two notable concave-convex changes on both sides of the trial excavation segment. The underlying reasons for such changes are attributed to rock-machine interactions during practical TBM construction. During tunnelling, the rotational speed of the cutter head is set firstly. At this time, the cutter head torque will rise and thus stabilize to a certain value. Subsequently, with the setting of advancing speed, the torque does not change greatly while the thrust
The cutterhead has not touched the rock wall during this period. The above procedure can be described as empty pushing segment. After the cutterhead contacts with the rock wall, the torque decreases a little bit and then rises sharply, and the curve of thrust will vibrate more fiercely than that in empty pushing segment. This period can be called as trial excavation segment. After the ascending segment, thrust, torque and advancing speed tend to be stable, and fluctuate in a certain range. The stage becomes the stable segment in a tunnelling cycle.

![Figure 2](image1.png)

**Figure 2.** The evolution of advancing speed in a certain tunnelling cycle.

![Figure 3](image2.png)

**Figure 3.** The evolution of the cutterhead torque, thrust and advance rate in a certain tunnelling cycle (Adapted from [15]).

Thus, the primary task is to detect these two change points (C-Ps) outlined in red rectangular boxes as shown in Figure 3. In this paper, the data of torque are employed to recognize the trial excavation segment. Firstly, cumulative sum (CUSUM) change-point detection method is used to identify the second C-P automatically. The idea of CUSUM method is elegant and simple. The cumulative residual sum is calculated recursively and then the maximum one of the cumulative sums deviation
from the reference state is identified and selected. Let \( \{x_1, x_2, \ldots, x_n\} \) be a time series with \( n \) data points. \( \bar{x} \) is the mean value of the time series. Then the cumulative sum of this sequence is defined as follows.

\[
\text{CUSUM} = |S_k| = \sum_{i=1}^{k} (x_i - \bar{x}) \quad (\text{for } k = 1, 2, \ldots, n)
\]

The point where CUSUM achieves the maximum is the change point. In this case, it can be recognized as the second C-P.

Once the second C-P is determined, the terminal point of trial excavation segment is thus identified. To recognize the first C-P and determine the starting point of trial excavation segment, the gradients between the second C-P and each data point ahead the second C-P are calculated. The smoothing method is utilized to remove the periodic effects. The point with the maximum value of smoothed gradient line can be regarded as the first change point. Therefore, the segment between the first and second C-Ps is determined as the trial excavation segment and the segments before and after trial excavation segment are the empty pushing segment and stable segment, respectively.

3.3. Feature extracting

Features are the inputs of the models and stay between data and models in the machine learning pipeline. In this case, all in-situ data are the features. Traditional feature engineering often pays attention to features that contribute more to the target vectors, while ignores the harm caused by redundancy among features [16]. Overabundant features may confuse the algorithms and lead to over-fitting or even worse. The training process might go awry because of overabundant features and generate unreliable prediction results. In order to reduce the dimension of input space and save the computational cost, the correlation between two features is measured according to Equation (4) and the threshold of the correlation coefficient is 0.90 [17]. Specifically, when the correlation coefficient between two features is greater than 0.9 or less than -0.9, one of the features should be removed to eliminate redundancy and improve efficiency.

\[
\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)} \sqrt{E(Y^2) - E^2(Y)}}
\]

where \( \rho_{X,Y} \) is the correlation coefficient of two features \( X \) and \( Y \), \( \text{cov}(X,Y) \) means the covariance between \( X \) and \( Y \), \( E \) and \( \sigma \) are the mean and variance of the corresponding feature, respectively.

3.4. Model establishment

The Support Vector Machine (SVM) is established as the regression model. The general formulas of SVM are proposed by Vladimir Vapnik and Corinna Cortes [18]. SVM is an effective regression algorithm. Based on the kernel trick, SVM can map the input data into high-dimensional feature space and thus solve the non-linear regression problem well. The popularly used kernel functions include polynomial kernel function, Gaussian kernel function and string kernel function. For given non-linear samples, SVM uses kernel function and soft margin or convex quadratic programming to obtain the decision function. More details can be found in Smola and Scholkopf’s work [19].

4. Application and validation

In this section, the proposed data mining framework was applied to analyse the in-situ data of a water conveyance project in Jilin Province, China. The details are presented below.

4.1. Project background

The main line of the whole project is composed by four parts, spanning more than 260 km [20]. The Tunnel #3 is the most difficult segment of construction in this project with complex geological conditions and frequent hazards. Tunnel #3 is approximately 24.3 km long and numbers of valuable data are recorded during tunnelling. Tunnel #3 is constructed by an open type of TBM. The tunnelling process lasted for 803 days including construction interruptions. The diameter of the disc cutterhead is
7.93 m. During tunnelling process, there are 199 kinds of operational parameters collected by data acquisition systems and geological prospecting, including cutter speed, cutter torque, total thrust of excavation, surrounding rock grades, etc.

4.2. Results
Through the procedures of data cleaning, the division of the whole time series data and feature extracting as described in Sections 3.1, 3.2 and 3.3, a total of 4802 full tunnelling cycles were obtained during 803 days’ construction as the dataset. For each full tunnelling cycle, 37 important features, such as gripper pressure, rotational speed, penetration and so on, are extracted from massive operational TBM data. The detailed input features are shown as Table 1. Trail excavation segments and stable segments are firstly extracted from full tunnelling cycles. Then the data of the first 30 s of trial excavation segment are selected, and the mean and variance of these 37 features are calculated inputs. Further, the mean values of the torque and thrust in stable segment are computed as outputs.

| Table 1. Extracted features. |
|-----------------------------|
| TBM parameters | TBM parameters | TBM parameters |
|-----------------|----------------|----------------|
| Gripper pressure | Motor current of gripper pump | Steel arch pump pressure |
| Motor torque of main drive | Motor current of stepping pump | Host belt pump pressure |
| Motor output frequency of main drive | Advance of thrust cylinder | CH4 concentration of cutterhead |
| Temperature of gear reducer | Advance of gripper cylinder | CO2 concentration of control room |
| Rotational speed of cutterhead | Aided system pressure | CH4 concentration of equipment Bridge |
| Torque | Left gripper angle of pitch | CH4 concentration of trialer tail |
| Tunneling time | Left gripper roll position | Drag cylinder pressure |
| Left gripper chamber pressure | Right gripper angle of pitch | Motor current of manipulator pump 1 |
| Right gripper chamber pressure | Right gripper roll position | Motor current of manipulator pump 2 |
| Pressure | Controlling pump pressure | Speed of host belt conveyor |
| Penetration | Controlling oil line pressure 2 | Speed of Bridge Belt Conveyor |
| Speed | Control oil circuit pressure 1 | Speed of Slag Conveyor |
| Displacement | Stepping pump pressure | Motor Current of Host Belt Pump |
| Motor current of push pump | | |

To train and test the classifier, we randomly select 90% tunneling cycles as a training set and the remaining 10% tunneling cycles are set as a test set. The training set is used for the training of the parameters in the predictors while the test set is employed to test the final performance of the predictors. Three evaluation indicators including mean absolute error (MAE), root-mean-square error (RMSE), and determination coefficient (R²) are used in this study. Table 2 indicates the prediction performance of SVM when employing the Gaussian kernel. The R² equals to 0.914 and 0.930 at predicting the thrust and torque, respectively. The MAE and RMSE for predicting the thrust are 199.934 kN and 276.359 kN, separately. The MAE and RMSE for the predictions of the torque are 848.386 kN·m and 1231.178 kN·m, respectively.
To evaluate the effectiveness of data-partitioning, the comparison between prediction results with and without data-partitioning is conducted. Table 3 shows the prediction performance of SVM without data-partitioning. By comparing Tables 2 and 3, we can find that the forecasting results are more satisfying when the trial excavation segment is extracted in all of the three evaluation indicators. This practice exhibits that the division of ascending segment is conducive for machine learning algorithms to capturing data characteristics and forecasting the variables in stable segment.

**Table 2. Performance of SVM for different operational parameters.**

| SVM  | Thrust | 0.914 | MAE         | 199.934 | RMSE  | 276.359 |
|------|--------|-------|-------------|---------|-------|---------|
| SVM  | Torque | 0.930 | 848.386     | 1231.178|

**Table 3. Performance of SVM for different operational parameters without data-partitioning.**

| SVM  | Thrust | 0.846 | MAE         | 263.725 | RMSE  | 361.030 |
|------|--------|-------|-------------|---------|-------|---------|
| SVM  | Torque | 0.860 | 1278.895    | 1701.016|

5. Conclusions and future work

TBM is widely used in tunnel construction because of its high driving performance. However, its performance is highly dependent on the adjustment of operational parameters during tunneling. To improve the adaptability of a TBM in complex geological conditions, an integrated data mining framework has been proposed to unveil the interactions between mutable geological conditions and TBM operational data. This paper divides the full tunneling cycle into four segments depending on their significant characteristics and highlights the role of trial excavation segment in the whole cycle. Data cleaning, feature extraction and statistical variable selections are conducted firstly to compress the big TBM-operational streaming data. Based on the analysis of historical sensor data, the thrust and torque in the stable excavation segment are forecasted via SVM. The results showed that the proposed framework performed well in predicting those two parameters with the determination coefficient R² all exceeding 0.9, which illustrated the feasibility of using the proposed framework to assist driving in TBM construction.

Future work will focus on two aspects. First, the related quantitative mechanical analysis should be further conducted. Second, it will investigate how to develop an automatic control system based on the integrated data mining framework.

6. References

1. Sun W, Shi M, Zhang C, Zhao J and Song X 2018 Dynamic load prediction of tunnel boring machine (TBM) based on heterogeneous in-situ data *Autom. Constr.* 92 23–34
2. Okubo S, Fukui K and Chen W 2003 Expert system for applicability of tunnel boring machines in Japan *Rock Mech. Rock Eng.* 36 305–22
3. Ghasemi E, Yagiz S and Ataei M 2014 Predicting penetration rate of hard rock tunnel boring machine using fuzzy logic *Bull. Eng. Geol. Environ.* 73 23–35
4. Farrokh E, Rostami J and Laughton C 2012 Study of various models for estimation of penetration rate of hard rock TBMs *Tunn. Undergr. Sp. Technol.* 30 110–23
5. Tóth Á, Gong Q and Zhao J 2013 Case studies of TBM tunneling performance in rock-soil interface mixed ground *Tunn. Undergr. Sp. Technol.* 38 140–50
6. Delisio A, Zhao J and Einstein H H 2013 Analysis and prediction of TBM performance in blocky rock conditions at the Lötschberg Base Tunnel *Tunn. Undergr. Sp. Technol.* 33 131–42
7. Hassanpour J, Rostami J, Khamehchiyan M and Bruland A 2009 Developing new equations for TBM performance prediction in carbonate-argillaceous rocks: A case history of Nowsood water conveyance tunnel *Geomach. Geoenngin.* 4 287–97
8. Zhao Y, Gong Q, Tian Z, Zhou S and Jiang H 2019 Torque fluctuation analysis and
penetration prediction of EPB TBM in rock–soil interface mixed ground Tunn. Undergr. Sp. Technol. 91 103002

[9] Zhu H, Wang X, Chen X and Zhang L 2020 Similarity search and performance prediction of shield tunnels in operation through time series data mining Autom. Constr. 114 103178

[10] DeVries P M R, Viégas F, Wattenberg M and Meade B J 2018 Deep learning of aftershock patterns following large earthquakes Nature 560 632–4

[11] Yagiz S 2008 Utilizing rock mass properties for predicting TBM performance in hard rock condition Tunn. Undergr. Sp. Technol. 23 326–39

[12] Bruland A 2000 Hard Rock Tunnel Boring Vol 5 of 10 - Geology and Site Investigation(1D-98) Dr. theses NTNU 199881

[13] Gao X, Shi M, Song X, Zhang C and Zhang H 2019 Recurrent neural networks for real-time prediction of TBM operating parameters Autom. Constr. 98 225–35

[14] Khine M S 2013 Application of structural equation modeling in educational research and practice

[15] Wang X, Zhu H, Zhu M, Zhang L and Ju J W 2021 An integrated parameter prediction framework for intelligent TBM excavation in hard rock Tunn. Undergr. Sp. Technol. Under Revi

[16] Yu L and Liu H 2004 Redundancy based feature selection for microarray data KDD-2004 - Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining pp 737–742

[17] Bai S, Li M, Kong R, Han S, Li H and Qin L 2019 Data mining approach to construction productivity prediction for cutter suction dredgers Autom. Constr. 105 102833

[18] Cortes C and Vapnik V 1995 Support-Vector Networks Mach. Learn. 20 273–97

[19] Smola A J and Schölkopf B 2004 A tutorial on support vector regression Stat. Comput. 14 199–222

[20] Li S, Nie L and Liu B 2018 The Practice of Forward Prospecting of Adverse Geology Applied to Hard Rock TBM Tunnel Construction: The Case of the Songhua River Water Conveyance Project in the Middle of Jilin Province Engineering 4 131–7