Automation in Construction 92 (2018) 23–34

Contents lists available at ScienceDirect

Automation in Construction

journal homepage: www.elsevier.com/locate/autcon

Dynamic load prediction of tunnel boring machine (TBM) based on heterogeneous in-situ data

Wei Sun[a], Maolin Shi[a], Chao Zhang[b], Junhong Zhao[b], Xueguan Song[a,*]

[a] School of Mechanical Engineering, Dalian University of Technology, No.2 Linggong Road, Ganjingzi District, Dalian, Liaoning 116024, China
[b] School of Mathematical Sciences, Dalian University of Technology, No.2 Linggong Road, Ganjingzi District, Dalian, Liaoning 116024, China

ARTICLE INFO

Keywords:
Load prediction
TBM
Heterogeneous in-situ data
Data-driven technique

ABSTRACT

Load prediction of tunnel boring machines (TBMs) is crucial for the design and safe operation of these complex engineering systems. However, to date, studies have mostly used only geological data, but the operation of TBMs also has an important effect on the load, especially its dynamic behavior. With the development of measurement techniques, large amounts of operation data are obtained during tunnel excavation. Mining these heterogeneous in-situ data, including geological data and operation data, is expected to improve the prediction accuracy and to realize dynamic predictions of the load. In this paper, a dynamic load prediction approach is proposed based on heterogeneous in-situ data and a data-driven technique. In this approach, the integration of heterogeneous in-situ data is conducted as follows: i) the geological data are extended to match the scale of the operation data using an interpolation method; ii) the categorical data and numerical data are fused through a proposed encoding method; and iii) the geological data are combined with the operation data according to the location of each operation datum. A data-driven technique, Random forest, is used to construct the prediction model based on the integrated heterogeneous in-situ data. The approach is applied to a collection of heterogeneous in-situ TBM data from a tunnel in China, and the results indicate that the approach can not only accurately predict the dynamic behaviour of the load but can also precisely estimate the statistical characteristics of the load. This work also highlights the applicability and potential of data-driven techniques in the design and analysis of other complex engineering systems similar to TBMs.

1. Introduction

Tunnel boring machines have been widely used for the tunnel excavation because of their relatively higher efficiency, safety and environmental friendliness compared to conventional blasting excavation. To ensure the proper design of TBMs and the safe operation of these systems during excavation, accurately predicting the load (generally referring to the thrust and torque of the cutterhead) is essential [1]. In recent decades, researchers have focused on using geological data to predict the load, and a number of load prediction methods have been proposed. These methods can be typically categorized into three classes: empirical methods, rock-soil mechanics methods, and numerical simulation methods. Krause [2] proposed an empirical model for TBM load prediction, and this model is widely used for load prediction in the design of TBMs [4–6]. Mikaeil [7] proposed different empirical equations for different types of rocks respectively and developed a fuzzy rock classification system. Yagiz [8] reviewed the studies of empirical methods and utilized a polynomial exponential regression to predict the load and performance of TBMs. In recent years, some researchers integrated empirical methods with scaled cutting experiments to improve the prediction accuracy. Gertsch [9] compared the effects of different geological parameters on the load using a rock failure experiment. Xue [10] analyzed the stress in the rock failure process caused by a cutter in a cutting experiment. Entacher [11] improved the cutting experiment and developed an empirical method to estimate the load based on the cutting results. In terms of rock-soil mechanics methods, Shi [12] and Wang [13] used rock-soil mechanics to calculate the load with the assumption that only one stratum is present in the excavation face. Zhang [14] advanced their work in order to take multiple strata into consideration. In terms of numerical simulation methods, Kasper [15] analyzed the stress distribution in the excavation face through a three-dimensional numerical simulation. Su [16] applied regression analysis to analyze the torque based on a numerical simulation of the cutting process. The above-mentioned methods provide insights into TBM load predictions and benefit the design and analysis of TBMs. However, as geological data are static data, these methods can be used to determine only certain aspects of the load, such as the mean and range, which are too coarse for the design and analysis of TBMs.

* Corresponding author.
E-mail address: sxg@dlut.edu.cn (X. Song).

https://doi.org/10.1016/j.autcon.2018.03.030
Received 22 October 2017; Received in revised form 2 February 2018; Accepted 20 March 2018

Past studies of TBMs have shown that the approximate range of the load is mainly determined by the geological conditions [2,4–18], but the dynamic behavior of the load is largely determined by the operation of the TBM [1,19]. Using both geological data and operation data synchronously is expected to not only improve the prediction accuracy but also realize dynamic load predictions. However, to the best of our knowledge, no such studies have yet been reported.

With the advancement and development of cyber-physical systems and measurement techniques, massive amounts of operation data on complex engineering systems, such as aircraft, chemical process systems, nuclear systems and wind power systems, are obtained during the engineering process, providing many opportunities for the practical application of data-driven techniques to aid in the design and analysis of these complex engineering systems [20–22]. Biyington used an artificial neural network (ANN) to predict the remaining life of key components in an aircraft based on in-situ data [23]. Yin used several data-driven techniques for chemical process monitoring and fault diagnosis (PM-FD) and provided some references to achieve successful PM-FD for large-scale industrial processes [24]. Zio used fuzzy similarity analysis to predict the failure scenarios and remaining useful life based on historical data from a nuclear system [25,26]. Zhang used support vector regression (SVR) to analyze the effects of ambient and wake turbulence on the power generation of wind turbines based on wind velocity and power data [27]. Data-driven techniques have been successfully used in construction as well. Kusiak used an ANN to predict the steam load in a new building based on historical data from other buildings [28]. Adoko used adaptive neuro-fuzzy inference systems to predict the rock burst intensity based on field measurement data [29]. Shirmohammadi applied adaptive neuro-fuzzy inference systems to predict the groundwater level based on historical geological data [30]. In recent years, TBMs have been widely used in numerous tunnel excavations, and considerable operation data have been recorded. Such heterogeneous in-situ data, including operation data and geological data, should be useful in predicting the dynamic load of the TBM. However, the heterogeneous in-situ data have two special characteristics that limit the application of data-driven techniques on them. The first is that the sizes of the geological data and the operation data are different, and the other is that the heterogeneous in-situ data include not only numerical data but also categorical data, which cannot be used directly in load predictions. Therefore, new methods need to be designed to efficiently integrate the heterogeneous in-situ data in order to predict the dynamic load of a TBM via a data-driven technique.

In this paper, a dynamic load prediction approach is proposed based on the heterogeneous in-situ data using a data-driven technique, and a collection of in-situ TBM data from a tunnel in China is used to validate the proposed approach. The remainder of this paper is organized as follows. Section 2 introduces the engineering project where the heterogeneous in-situ data come from. Section 3 presents the details of the proposed approach, including the integration of the heterogeneous in-situ data and the selected data-driven technique. The prediction results and discussion are provided in Section 4. The final section details the conclusions of this paper.

2. Project review

The tunnel studied in this paper is located in a metro line in China and has a length of 2000 m and a diameter of 6.4 m. A schematic illustration of the tunnel is provided in Fig. 1. The ground surface elevation ranges from 0.2 to 5.8 m, and the depth of the tunnel floor from the ground surface ranges from 11.8 to 25.4 m. From the ground surface to the tunnel floor, various geological layers, such as clay, sand and rock, are unevenly distributed. Some of the geological characteristics of these layers are described in Appendix A. To excavate the tunnel, an earth pressure balance (EPB) shield TBM was used. This system consists of a cutterhead, chamber, screw conveyor, tail skin and other auxiliary subsystems. The TBM has a diameter of 6.2 m and a total mass of over 500,000 kg, and the cutterhead features an opening percentage of 30% and 120 cutters.

3. Dynamic load prediction approach

A flow chart of the proposed dynamic load prediction approach is shown in Fig. 2. The details of each part are presented below.

3.1. Data description

The heterogeneous in-situ data used here include the operation data from the TBM measurement system and the geological data from the geological investigation report. The operation data are composed of 53 attributes (for details, see Appendix B) that were continuously measured with a frequency of 1 Hz along the entire length of the tunnel. The geological data were obtained from 98 sampling locations along the tunnel and include the layer classification (categorical data, shown in Appendix A), the geological spatial distribution (numerical data, shown in Fig. 1) and the mechanical parameters of each geological layer (numerical data, shown in Appendix A). To efficiently integrate the heterogeneous in-situ data, the geological data and operation data are processed as follows.

3.2. Data integration

3.2.1. Geological data processing

The geological data represent the geological information on the sampling locations, but the operation data represent the operational information along the length of the entire tunnel. For correlation with the operation data, the geological data need to be extended, and an interpolation method, kriging method (KRG), is used to estimate the geological conditions between the sampling locations. The details of the geological data extension are as follows [31]. By combining a global model and a localized departure, the KRG function can be formulated as follows:

$$y(x) = \sum_{j}^{p} \beta_j f_j(x) + z(x) \tag{1}$$

where x is the elevation of the boundary between adjacent layers at the sampling locations, as shown in Fig. 3; $y(x)$ is the elevation of the boundary in an unsampled location; $f_j(x)$ is a known approximation function, and β_j is its coefficient; and $z(x)$ represents a stochastic Gaussian process described by the following equations.

$$E(z(x)) = 0 \tag{2}$$

$$E(z(x_i)z(x_j)) = \sigma^2 R(\theta, x_i, x_j) \tag{3}$$

where σ^2 is the variance; $R(\theta, x_i, x_j)$ is the correlation function between x_i and x_j; and θ is the correlation parameter. Given m sampling locations and letting $(f(x) = (f_1(x), f_2(x), \ldots, f_p(x))^{\mathrm{T}}$, $\beta = (\beta_1, \beta_2, \ldots, \beta_p)^{\mathrm{T}}$, $Z = (z_1, z_2, \ldots, z_p)^{\mathrm{T}}$ and $F = (f(x_1), f(x_2), \ldots, f(x_m))^{\mathrm{T}}$, the elevation of the unsampled location Y can be reformulated as follows:

$$Y = F\beta + Z \tag{4}$$

Based on the linear regression model , $c(x) \in \mathbb{R}^m$, and the prediction error is

$$\hat{y}(x) - y(x) = c(x)^T Z - z(x) \tag{5}$$

The corresponding mean square error $\varphi(x)$ is as follows:

$$\varphi(x) = \sigma^2(1 + c(x)^T R c(x) - 2c(x)^T r_x) \tag{6}$$

where r_x is the correlation function between unsampled location x and