Machine Learning Approaches to Enable Resource Forecasting Process of Road Tunnels Construction

Arsalan Mahmoodzadeh\(^1\)\(^*\), Shima Rashidi\(^2\), Adil Hussein Mohammed\(^3\), Hunar Farid Hama Ali\(^4\), and Hawkar Hashim Ibrahim\(^5\)

\(^1\)Department of Civil Engineering, University of Halabja, Halabja, Kurdistan Region, Iraq
arsalan.mahmoodzadeh@uoh.edu.iq

\(^2\) Department of Computer Science, College of Science and Technology, University of Human Development, Sulaimaniyah, Kurdistan Region, Iraq

\(^3\) Department of Communication and Computer Engineering, Faculty of Engineering, Cihan University-Erbil, Kurdistan Region, Iraq

\(^4\) Department of Civil Engineering, University of Halabja, Halabja, Kurdistan Region, Iraq

\(^5\) Civil Engineering Department, College of Engineering, Salahaddin University-Erbil, 44002 Erbil, Kurdistan Region, Iraq

Abstract—Increasing demand for tunneling projects, increases attention to time and cost required for their construction. Most of parameters which are affecting on the time and cost of tunnel construction are unknown. The purpose of this paper is to provide a method to predict the construction time and cost of a road tunneling using linear regression (LR) method. In order to train the LR method, some datasets are obtained from the historical road tunnels. To verify the feasibility of the proposed method, it has been applied to a road tunnel. All of the forecasted results have been compared with the actual results obtained during the tunnel construction and the accuracy of the predictions has been investigated. According to three statistical evaluation criteria of root mean square error (RMSE), mean absolute percentage error (MAPE) and determination of the coefficient (\(R^2\)), a very high accuracy has been obtained in the prediction results. The RMSE, MAPE and \(R^2\) indices have been calculated as 0.0005 days, 0.9380637\% and 0.9874 for the construction time, respectively; and 7.1194 US\$,0.78891593\% and 0.9873 for the construction cost, respectively.

Index Terms—Underground space, Road tunnels, Tunnel ground conditions, Time and cost of tunnels construction, Gaussian Process; Linear regression.

I. INTRODUCTION

In the pre-construction step of tunnels, due to the unknown subsurface conditions and related uncertainties, engineers often have trouble choosing drilling method and support system. These uncertainties spontaneously affect the construction time and cost of tunnels (Mahmoodzadeh et al. 2019).

Today, the computer methods are very crucial in engineering matters, so that they can be used to reduce the uncertainties and the risk of the engineering problems without having to spend much time and cost (Kazemian et al. 2019).

Tunnels are among the most important underground projects that are used for a variety of applications. Excavation method and support systems of tunnels are designed based on geological or geotechnical properties (Aalianvari et al. 2018).

Due to the unknown properties along tunnel route, the tunnel projects are usually confronted with uncertainties about excavation method and support system, which these uncertainties lead to a lack of accurate estimation of the time and cost required for them. Today, conventional computer methods are presented to address such uncertainties and they are currently in a great development (Zhou et al. 2019). Among the computer methods, the Machine learning approaches have shown potential ability in various engineering issues (Mahmoodzadeh et al. 2021, Selvaraj et al. 2019, Sharma et al. 2022, Manikandan et al. 2020, Cao et al. 2022, Sivaram et al. 2020).

In this paper, the construction time and cost of a road tunnels are predicted. For this purpose, firstly with the experience from the previous tunnels, effective parameters on the time and cost of their construction are identified. In the next step, using the linear regression (LR) method, the tunnel construction time and cost are predicted. To this end, firstly a total of 350 datasets are collected from the previous road tunnel projects, each of which provided the status of the all parameters considered. For each dataset, the actual time and cost of constructing each meter of the corresponding tunnel are also presented. These 350 datasets are used to train the LR method to predict the time and cost required for the test tunnel. Also, the predicted status of the parameters considered along the test tunnel is applied as the test data in the LR method. Ultimately, the time and cost needed to construct the test tunnel are predicted at any location of its route. Also, after the test tunnel construction, the parameters estimated/predicted along its route and the predicted time and cost of its construction have been compared with the actual mode. Finally, the accuracy of the predictions is investigated by different statistical evaluation criteria.

II. LINEAR REGRESSION (LR)

LR can be defined as an algorithm of machine learning depends on supervised learning that carried out a task of regression.
Regression models a target forecasting value based on independent variables. The relationship between variables and forecasting can be found by utilizing this method. Based on the kind of relationship between dependent and independent variables they are considering and the number of independent variables being used different types of regression models can be applied.

If independent variable (x) is given, LR performs the task to predict a dependent variable value (y). So, this technique finds out a linear relationship between x (input) and y(output). Hence, the name is linear regression. In Fig. 1, X (input) is the work experience and Y (output) is the salary of a person. The regression line is the best fit line for the model.

![Fig. 1. An example of the LR method.](image)

Hypothesis function for the LR method is according to equation (1).

\[ y = \theta_1 + \theta_2 \cdot x \] (1)

While, x is the input training data (univariate – one input variable), y is the labels to data (supervised learning), \( \theta_1 \) is the intercept, and \( \theta_2 \) is the coefficient of x.

By finding the best \( \theta_1 \) and \( \theta_2 \) values, the model gets the best line to predict the value of y for a given value of x.

III. GENERAL PROPERTIES OF DATASETS

350 datasets gathered during the construction of historical road tunnels were used as the training datasets of the LR method. In the database, 16 effective parameters on the time and cost of road tunnels construction were considered as the input parameters. Also, 181 datasets obtained during the construction of test tunnel were used as the test datasets. An overview of the training datasets is provided in Table 1.

![Table 1: Range of different used parameters.](image)

IV. CONSTRUCTION TIME AND COST RESULTS VALIDATION

In Fig. 2 and Fig. 3, the predicted results of the construction time and cost are compared with the real mode along the test tunnel path. According to these figures, very little difference is observed in the predicted results compared to the real mode. In Fig. 4 and Fig. 5, the error values of the predicted construction time and construction cost compared to the real mode are shown. The RMSE values for the construction time and construction cost are less than 0.0005 days and 7.12 US$, respectively. These RMSE values indicate the high accuracy of the predicted results.

Table 2, summarizes the RMSE, MAPE and \( R^2 \) values calculated for the construction time and cost using the LR method.

![Fig. 2. The predicted construction time compared to the real mode of the test tunnel.](image)
In the road tunnel projects, it is important to predict the time and cost of their construction that can cause high risk and uncertainties in the initial planning. The construction time and cost prediction should be done based on the effective parameters. Also, according to the size of the training data available, useful prediction methods should be utilized. In this paper, 16 parameters that influence the time and cost of the road tunnels construction are considered.

Using the LR method to predict the time and cost of tunnel construction showed that this method has high ability to make predictions with large amount of the training data.

It should be noted that it is true that the LR method produced the good prediction results, but this precision could not be obtained if the 16 parameters were not selected correctly. That is, the selected parameters must influence the time and cost of the tunnel construction. The results presented in this study showed that these parameters were selected correctly.

**V. Conclusions**

In the road tunnel projects, it is important to predict the time and cost of their construction that can cause high risk and uncertainties in the initial planning. The construction time and cost prediction should be done based on the effective parameters. Also, according to the size of the training data available, useful prediction methods should be utilized. In this paper, 16 parameters that influence the time and cost of the road tunnels construction are considered.

Using the LR method to predict the time and cost of tunnel construction showed that this method has high ability to make predictions with large amount of the training data.

It should be noted that it is true that the LR method produced the good prediction results, but this precision could not be obtained if the 16 parameters were not selected correctly. That is, the selected parameters must influence the time and cost of the tunnel construction. The results presented in this study showed that these parameters were selected correctly.

**REFERENCES**


decentralized grid-connected smart energy system based on hybrid solar-geothermal resources; Techno-economic evaluation. *Sustainable Cities and Society* 76, 103425, doi.org/10.1016/j.scs.2021.103425