Prediction of Highway Tunnel Cost by Least Squares Support Vector Machine Based on Particle Swarm Optimization

Jing Liao(✉) and Guanghua Li

College of Environment and Civil Engineering, Chengdu University of Technology, Chengdu, Sichuan, China
liaojing0422@163.com, liguanghua13@cdut.cn

Abstract. In the early stages of a project, the owner is faced with the challenge of deciding whether to invest in the project or not. The road tunnel project involves a large amount of investment, and in order to help the owner to make a selection decision, it is crucial to achieve accurate and efficient road tunnel cost prediction in the early stages. In this paper, the main factors affecting the cost of road tunnels are identified through literature research, and the least squares support vector machine is optimized using a particle swarm optimization algorithm. The results show that the least squares support vector machine model based on the particle swarm optimization algorithm performs well in road tunnel cost prediction, with high fit and low error, and meets the accuracy and practicality requirements of the pre-project.

Keywords: highway tunnel · particle swarm · least squares support vector machine · cost prediction

1 Introduction

In the last decade, China's highway tunnelling projects have seen rapid development in terms of quantity and quality [4, 12], but they also face many economic and technical problems, mainly in terms of the complexity of the project, the long construction period required and the large scale of investment [2, 7]. In order to promote the construction of road tunnels, it is important to make accurate predictions on the cost of road tunnels in the early stages of the project, which not only helps investors to realize multiple options and make decisions, but also may have a binding effect on the project cost control during the construction process [10].

With the advent of the age of intelligence, machine learning models are favoured by an increasing number of scholars. Currently, neural networks and support vector machines (SVM) are the mainstream machine learning models that are widely used in cost prediction [1]. For example, some scholars have used the advantage of BP neural network with good descriptiveness for complex non-linear problems to make fast and accurate cost prediction for power transmission line projects [5]; other scholars have

J. Yang et al. (Eds.): ICMSEM 2022, AHE 12, pp. 406–414, 2023.
https://doi.org/10.2991/978-94-6463-038-1_37

used BP neural network and RBF neural network to predict the cost of construction projects respectively [15], but BP neural network is more complicated to build and requires a large amount of sample data to support. However, BP neural network is more complicated to construct and requires a large amount of sample data to support, which is lower than RBF neural network in terms of prediction speed and accuracy [6]. For support vector machines, on the other hand, they have features such as avoiding overfitting and being friendly to a small number of samples, for example, some scholars use rough set theory-support vector machine models to predict road tunnel cost [13]; Others have used support vector machines (SVM) and least squares support vector machines (LSSVM) for engineering cost prediction, all with good results. [8, 9].

In addition, Besides, there are also scholars who use various optimization algorithms, such as artificial bee colony algorithm (ABC), sparrow algorithm (SSA), genetic algorithm (GA), etc., to optimize neural networks and support vector machines (SVM) to improve the performance of all aspects of the models and make them better serve the field of cost prediction. For example, some scholars proposed a BP neural network based on the sparrow optimization algorithm to improve the convergence speed of the basic BP neural network and finally the rural road cost prediction [14]. Other scholars used genetic optimization algorithm to optimize the BP neural network to avoid the BP neural network falling into local optimum, and finally established the distribution network engineering cost prediction model with high prediction accuracy [16].

Particle swarm optimization algorithms have the advantage of being simple and easy to implement compared to other optimization algorithms [9]. The least squares support vector machine model (LSSVM) is an optimized extension of the basic support vector machine model (SVM), which greatly improves the speed and accuracy of the operation [3]. This paper first identifies the main factors affecting road tunnels through literature research and other methods, substitutes the obtained sample data into the LSSVM model optimized by the PSO algorithm for training, and analyses and compares the prediction results of the GA-LSSVM model with the PSO-LSSVM model to arrive at the model with the best performance in predicting the cost of road tunnel projects.

2 Methodology

2.1 Particle Swarm Optimization Algorithm

The Particle Swarm Optimization (PSO) algorithm was proposed in 1995 by Eberhart and Kennedy in the USA. Initially inspired by the foraging behaviour of bird groups, it uses the cooperative sharing mechanism between groups to search and iterate through the problem space to obtain the optimal solution to the problem. The algorithm has a refined structure and few adjustment parameters, and has been widely used in various fields in recent years [17]. In the particle swarm optimization algorithm, it is assumed that a randomly initialized group of uniformly distributed particles $x = \{x_1, x_2,\ldots, x_n\}$, the kth particle position $x_k = \{x_{k1}, x_{k2},\ldots, x_{kD}\}$, with velocity $v_k = \{v_{k1}, v_{k2},\ldots, v_{kD}\}$. Each particle has its own fitness value and its velocity and position are updated according to the individual extremes (pbest) and the population extremes (gbest) until they reach a set value. The individual extreme value (pbest) is the optimal solution for the particle itself, while the population extreme value (gbest) is the optimal solution for the whole population of particles. All particles are updated iteratively according to the following equation:

$$v_k(t+1) = \omega v_k(t) + c_1 r_1 (pbest_k(t) - x_i(t)) \\ + c_2 r_2 (gbest - x_k(t)) \tag{1}$$

$$xk(t+1) = xk(t) + vk(t+1) \tag{2}$$

where: ω is the inertia weight coefficient with the principal of balanced global and local search; C_1, C_2 are the learning factors; $r_1, r_2 \in [0,1]$, are random numbers.

2.2 Least Squares Support Vector Machine Regression Algorithm

The least squares support vector machine regression algorithm (LSSVM), proposed by Suykens and Vandewalle in 1999, is an improved extension of the traditional support vector machine. It continues the advantages of the kernel function and small sample size of the traditional support vector machine, and has received widespread attention in classification recognition problems as well as regression prediction problems. In this algorithm, the sample set is assumed to be $N = \{(x_\mathrm{i}, y_\mathrm{i}), \mathrm{i} = 1,2,\ldots, \mathrm{k}\}$, where $x_\mathrm{i} \in \mathrm{Rn}$ is the input vector and y_i is the output value corresponding to x_i. And then mapped to the high-dimensional feature space by the non-linear mapping $\varphi(x_\mathrm{i})$, at this point the regression function is:

$$f(x) = \omega \cdot \varphi(x) + b \tag{3}$$

where: ω and b are the weight vector and offset of the function, respectively. Solving in the high-dimensional feature space yields $\xi_\mathrm{i}{}^2$ as the loss function, at which point the inequality constraint is transformed into an equation constraint, so that the optimization problem for the LSSVM evolves into the following equation:

$$\min \frac{1}{2}\|\omega\|^2 + C\frac{1}{2}\sum_{i=1}^{k}\xi_i^2 \tag{4}$$

$$s.t. y_i[\omega \cdot \varphi(x) + b] = 1 - \xi_i \tag{5}$$

where: C is the penalty coefficient, which has the function of regulating the complexity and error of the model; ξ_i is the slack variable.

The Lagrange function is solved according to Eqs. (4) and (5) by establishing the following:

$$L(\omega, b, \xi, a) = \frac{1}{2}\|\omega\|^2 + C\frac{1}{2}\sum_{i=1}^{k}\xi_i^2 - \sum_{i=1}^{k} a_i\{y_i[\omega \cdot \varphi(x) + b] - 1 + \xi_i\} \tag{6}$$

where: $a \in$ R is a Lagrange multiplier, and the partial derivatives of L with respect to ω, b, ξ, a respectively through the KKT (Karush-Kuhn-Tucker) condition, as follows:

$$\begin{cases} \frac{\partial L}{\partial \omega} = 0 \to \omega - \sum_{i=1}^{k} a_i y_i \varphi(x_i) = 0 \\ \frac{\partial L}{\partial b} = 0 \to \omega - \sum_{i=1}^{k} a_i y_i = 0 \\ \frac{\partial L}{\partial \xi} = 0 \to C\xi_i - a_i = 0 \\ \frac{\partial L}{\partial a_i} = 0 \to y_i[\varphi(x_i) + b] - 1 + \xi_i = 0 \end{cases} \tag{7}$$

Eliminating ω and ξ in Eq. (7) gives the following matrix:

$$\begin{pmatrix} 0 & Z^T \\ Z & A + C^{-1}I \end{pmatrix} \begin{pmatrix} b \\ a \end{pmatrix} = \begin{pmatrix} 0 \\ y \end{pmatrix} \tag{8}$$

where: $Z = [1,1,\ldots,1]^{\mathrm{T}}$, $y = [y_1,y_2,\ldots,y_{\mathrm{k}}]$, $a = [a_1,a_2,\ldots,a_{\mathrm{k}}]$, $A = [\varphi(x_1)^{\mathrm{T}}y_1,\varphi(x_2)^{\mathrm{T}}y_2,\ldots,\varphi(x_{\mathrm{i}})^{\mathrm{T}}y_{\mathrm{i}}]$.

Introducing the Gaussian radial basis (RBF) kernel function K $(x_{\mathrm{i},}\ x_{\mathrm{j}}) = \exp(-|x_{\mathrm{i}}-x_{\mathrm{j}}\ |/\sigma\hat{}2)$ and combining the Mercer condition for parameters a and b, the decision regression function of the LSSVM model is finally determined as follows:

$$f(x) = \sum_{i=1}^{k} a_i K(x_i, x_j) + b \tag{9}$$

In this paper, the particle swarm optimisation (PSO) algorithm is used to optimise two important factors of the least squares support vector machine (LSSVM), namely the parameter values of the kernel function and the parameter values of the penalty factor C. The main factors affecting the cost of a road tunnel are determined using a literature survey. A literature survey was then used to identify the main factors affecting the cost of road tunnels, which were used as input variables and substituted into the least squares support vector machine model optimised by the particle swarm algorithm to obtain predictions with the cost per linear metre as the output variable. Finally, the relative error δ and the mean absolute percentage error e_{MAPE} were used as evaluation indicators for the two models to compare and analyse the results, resulting in the optimal road tunnel cost prediction model. The specific process is shown in Fig. 1.

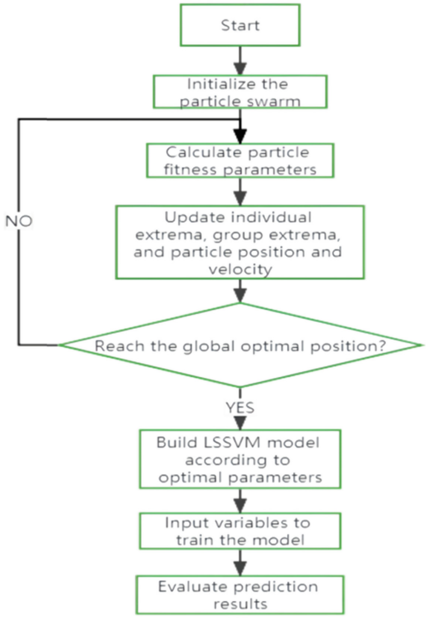

Fig. 1. PSO-LSSVM prediction flowchart

3 Model Application and Analysis

3.1 Analysis of Influencing Factors

According to relevant literature research, the main factors affecting the cost of highway tunnels are determined:

Tunnel length: The length of a road tunnel has an important impact on the scale of the entire project, which is divided into short, medium, long and extra-long tunnels according to existing regulations. The different lengths of tunnels have a significant impact on the form of work carried out during the construction process.

Rock grade: The cost of different rock grades is very different and is an important factor affecting the cost of the whole road tunnel project [11].

Average cross-sectional area of the tunnel: the choice of construction method, the determination of the facility plan and the final cost are all influenced by the cross-sectional area of the tunnel project, therefore the average cross-sectional area is taken as one of the main factors.

3.2 Cost Forecast

A total of 18 road tunnel project cost data samples P_i (i = 1, 2,…,18) were collected and some of the data are shown in Table 1, some of which are shown in Table 1. To verify the reliability and accuracy of the PSO-LSSVM-based road tunnel project cost prediction model, the first 16 samples were used as the training set and the last 2 samples were used as the test set. In order to verify the effectiveness of the model for project cost

Table 1. Road tunnel project cost data

Serial No.	Length of tunnel(km)	Percentage of enclosing rock grade length(%)			Average cross-sectional area(m^2)	Cost per linear meter(10k/m)[a]
		III	IV	V		
P_1	1.25	-	0.67	0.33	93.73	4.51
P_2	2.23	-	0.72	0.28	93.08	4.52
P_3	1.43	-	0.74	0.26	92.75	4.39
…	…	…	…	…	…	…
P_{14}	1.603	-	0.62	0.38	44.90	2.19
P_{15}	2.105	-	0.80	0.20	93.31	4.31
P_{16}	2.126	-	0.77	0.23	93.27	4.25

a.(10k/m) is RMB10,000 per meter.

Table 2. Test set sample data

Test sample	Length of tunnel (km)	Percentage of enclosing rock grade length(%)			Average cross-sectional area(m^2)
		III	IV	V	
P_{17}	2.230	-	0.75	0.25	93.14
P_{18}	2.086	-	0.63	0.37	44.54

application, the GA-LSSVM model was introduced for analysis and comparison with the PSO-LSSVM model, and the optimized parameters were substituted into the model respectively, and the input variables were started to train the cost prediction model. The final training results of the two models and the prediction results of the case test set are shown in Figs. 2 and 3, representing the fitting effect of the predicted and actual values of the training set.

As can be seen from Fig. 2, the PSO-LSSVM model fits the training set very well, with a coefficient of determination R^2 as high as 0.998, and the test samples are inverse normalized to obtain the predicted cost per linear meter of tunnel, multiplied by the tunnel length to obtain the total tunnel cost. It can be seen that the model is suitable for predicting the cost of road tunnels and has a strong predictive capability, which can effectively improve the accuracy of the cost prediction of road tunnel projects.

3.3 Model Comparison

To demonstrate the accuracy of the PSO-LSSVM model, a comparative analysis was performed using the GA-LSSVM model with the PSO-LSSVM model. Finally, the prediction performance of several models was evaluated by calculating the relative error

Table 3. Actual and projected tunnel cost values

Test sample	True value	Predicted value	
		GA-LSSVM	PSO-LSSVM
P_{17}	9650.32	9291.72	9674.00
P_{18}	4300.04	5381.95	4507.09

Note: Both true and forecast values are in RMB Ten thousand

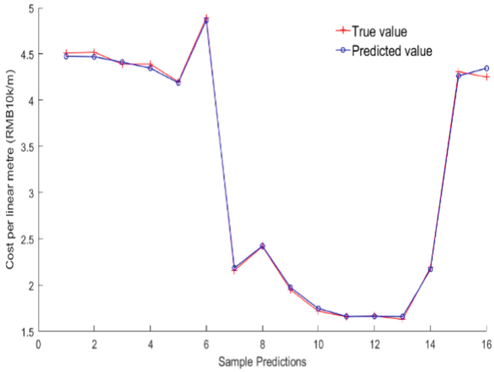

Fig. 2. PSO-LSSVM model training set fit graph

δ and the mean absolute percentage error e_{MAPE} of the test et prediction results are shown in Table 4, with the following equations:

$$\delta = \frac{y_{i\_t} - y_{i-p}}{y_{i\_t}} \tag{10}$$

$$e_{MAPE} = \frac{1}{n}\left|\frac{y_{i\_p} - y_{i-t}}{y_{i\_t}}\right| \tag{11}$$

where: yi_t samples the true value of cost and yi_p samples the predicted value of cost.

As can be seen from Table 4, the PSO-LSSVM model is smaller than the GA-LSSVM model in terms of both relative error δ and mean absolute percentage error e_{MAPE}, and the running time of the GA-LSSVM model in MATLAB2016a software is 20.11s, while the PSO-LSSVM running time is 3.33 s, indicating that the PSO-LSSVM model not only has It shows that the LSSVM model optimized by PSO not only has higher practicality and cost prediction capability, but also is better in terms of running speed, and is more suitable for the prediction of pre-project engineering cost of highway tunnel projects.

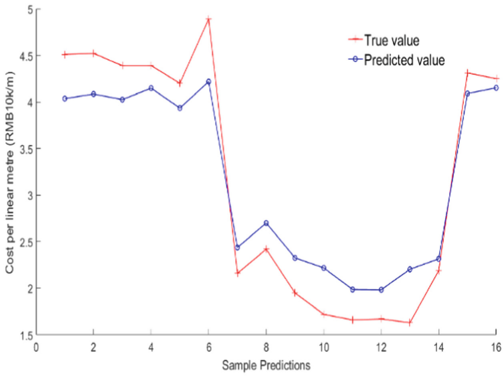

Fig. 3. GA-LSSVM model training set fit graph

Table 4. Comparison of PSO-LSSVM and GA-LSSVM model performance predictions

Test sample	Relative error(%)		Mean absolute percentage error(%)	
	GA-LSSVM	PSO-LSSVM	GA-LSSVM	PSO-LSSVM
P_{17}	3.7	0.25	10.08	2.50
P_{18}	25.16	4.82		

4 Conclusion

In this paper, the main factors affecting the cost of highway tunnels are determined through relevant literature research, and based on 18 project cost data, the LSSVM model is optimized using the PSO algorithm in MATLAB2016a. Finally, a least squares support vector machine learning prediction model based on the particle swarm optimization algorithm is established, and the prediction results are compared and analyzed with the GA-LSSVM model. The results show that the PSO-LSSVM model with $R^2 = 0.998$ and $e_{MAPE} = 2.50\%$ has a better fit and less error than the GA-LSSVM model, and performs excellent in terms of running speed. It is important to achieve reasonable investment decision and avoid causing waste of capital in the pre-project stage, and also provides an effective way for cost control.

References

1. Ai D and Yang JS. (2019). A machine learning approach for cost prediction analysis in environmental governance engineering. J. Neural Computing and Applications, 31(12): 8195-8203.
2. Gao S, Li R. (2019). Analysis and discussion on cost index of highway tunnel project. J. Construction Economics. 40(03): 57-63. DOI: https://doi.org/10.14181/j.cnki.1002-851x.201903057.
3. Huang M, Wu L, Yao Y. (2015). Cost prediction of mountain highway tunnels based on support vector machine - particle swarm algorithm. J. Highway, 60(07):285-288.

4. Hong KR, Feng HH. (2020). Development trend and reflection of highway tunnels in China in the past 10 years. J. China Journal of Highways. 33(12): 62-76. DOI: https://doi.org/10.19721/j.cnki.1001-7372.2020.12.005.
5. Ling YP, Yan PF, Han CZ, Yang CG. (2012). BP neural network-based cost prediction model for power transmission line project. J. China Electric Power. 45(10):95-99.
6. Liu J, Ye Q. (2013). Project cost prediction model of Xiamen city using BP and RBF neural network. J. Journal of Huaqiao University (Natural Science Edition). 34(05):576-580.
7. Li MC. (2019). Research on cost index prediction of highway tunnel project based on similar engineering discriminations. D. Chongqing JiaoTong University. doi: https://doi.org/10.27671/d.cnki.gcjtc.2019.000686.
8. Liu GN, Wang Y. (2021). Cost prediction method of road engineering construction stage based on LS-SVM. J. Construction supervision and inspection and cost. 14(Z1):56–60+67.
9. Liu YJ, Nie ZQ, Wang B, Xu.YH. (2021). Construction cost prediction of hydraulic tunnels based on PSO-SVM. J. People's Yellow River. 43(09):160-164.
10. Li TX, Li SS. (2021). Cost prediction of hydraulic tunnels based on particle swarm-wavelet neural network. J. Water Technology and Economy. 27(09):39-42.
11. Li PF. (2012). Influence of the classification of rock envelope level on the cost of highway tunnels. J. Heilongjiang Transportation Science and Technology. 35(11): 11-12. DOI: https://doi.org/10.16402/j.cnki.issn1008-3383.2012.11.017.
12. Niu JF. (2018). The development status and prospect of road tunnels in China. J. Housing and Real Estate. (12):229.
13. Peng C, Li QS, Liu HZ. (2016). Total cost prediction of tunnels in high altitude areas - based on rough set-support vector machine model. J. Heilongjiang Transportation Science and Technology. 39(08): 131-133. DOI: https://doi.org/10.16402/j.cnki.issn1008-3383.2016.08.084.
14. Wang SX, Zeng M. (2021). Research on cost prediction of rural roads based on SSA optimized BP neural network. J. Engineering Economics. 31(08): 25-29. DOI: https://doi.org/10.19298/j.cnki.1672-2442.202108025.
15. Wang S. (2018). Research on construction project cost prediction based on particle swarm optimization least squares support vector machine. D. Qingdao University of Technology,
16. Yang K, Yu B, Xiao YL, He YP, Wang FX. (2019). Cost prediction of distribution network projects based on GA-BP neural network. J. Automation Instrumentation, 40(07):91–93+99. DOI: https://doi.org/10.16086/j.cnki.issn1000-0380.2018080020.
17. Zhao NG, Deng JS. (2015). A review of particle swarm optimization algorithms. J. Science and Technology Innovation Herald. 12(26): 216-217. DOI: https://doi.org/10.16660/j.cnki.1674-098x.2015.26.114.