Research on the Influencing Factors and Prediction Model of Highway Engineering Cost Based on BIM Theory

Ruiqi Jia

School of Architectural Engineering, Sichuan Water Conservancy Vocational College, Chongzhou, Sichuan, 611231, China

13631894319@163.com

Keywords: Highway Engineering, Influencing Factors of Cost, Cost Prediction Model

Abstract: In view of the problems of incomplete analysis of influencing factors, insufficient model accuracy and small sample size in the current project cost prediction, based on grounded theory, BIM theory and related theories of machine learning, the key influencing factors are identified through feature screening, a BIM integrated prediction model is established, and the model is used to optimize parameters to improve the prediction accuracy. Verified by actual engineering cases, this model can effectively improve the accuracy and reliability of highway engineering cost prediction and provide a scientific reference for investment decisions.

1. Introduction

The cost prediction of highway engineering is interwoven and influenced by multi-stage and multi-dimensional factors, such as dynamic risk quantification and cross-professional collaboration. It is necessary to study the integration of big data, the optimization of BIM models, and establish a dynamic prediction system covering the entire life cycle. Therefore, it is particularly necessary to conduct in-depth research on the influencing factors and prediction models of highway engineering cost.

Many scholars have devoted themselves to the research on the cost prediction of highway engineering. Tang Ming [1] et al. established a vector autoregressive model (VAR) using the comprehensive cost index of highway engineering and revealed the relationship between macroeconomic variables and the cost index of highway construction through analytical tools such as variance decomposition. Yu Tiancai [2] studied the influence of factors such as terrain and climate on the cost of expressway construction in mountainous and high-altitude areas, and pointed out that the high proportion of Bridges and tunnels and the high construction safety risk are the key factors leading to the high project cost. Que Yunlong and Wei Kang [3] proposed the influencing factors of the cost deviation of highway projects in border areas, covering natural factors such as high cold, high altitude, and strong seismic intensity, as well as special working conditions such as the route passing through protected areas and uninhabited areas. Zheng Xiaolei and Zhang Shilian [4], based on the training of BP neural network, screened out seven main characteristic factors affecting the cost of highway engineering, such as the mileage of the main line and the number of channels. Li Junda [5] et al. established a database using 22 cases of existing expressways and proposed a cost estimation model for highway projects based on case-based reasoning technology. However, at present, there are still many difficulties and challenges in related research. For example, the analysis of the influencing factors of highway engineering cost is not comprehensive enough, and the selection of characteristic factors is rather subjective. The accuracy of some cost prediction models needs to be improved. The models are not updated in a timely manner and are overly simplified. The data sources of some studies are limited to specific regions and do not fully consider the impact of regional differences on the cost.

In view of this, this paper intends to adopt the method combining grounded theory and machine learning to establish a BIM highway engineering cost estimation model, integrate multi-dimensional data and intelligent analysis technology, and achieve the full-process cost control from design to

  DOI: 10.25236/icfmhss.2025.079

construction. Cost estimation was carried out by using regression analysis and equivalent linear model, and time series analysis was constructed in combination with BIM big data mining technology to predict the fluctuation of material prices and the trend of construction costs. This research provides predictive improvement methods for the cost management of highway engineering.

2. Theoretical basis and conceptual definition

2.1. Difficulties and Problems in the cost prediction of highway engineering

Highway engineering cost refers to the estimation, accounting and control of all the engineering costs, other construction costs and contingency reserves required to complete the highway project in the process of highway design, construction and maintenance management, in accordance with the planning, design documents and construction contracts, and in compliance with the relevant national pricing methods and standards.

Prediction difficulties and problems: Weak data foundation, insufficient accumulation of historical cost data, especially the lack of continuous tracking data for dynamic factors (such as market price fluctuations). The data accuracy in the survey and design stage is insufficient. The traditional quota method is difficult to adapt to the cost changes brought about by new technologies and new processes. Complex nonlinear relationships (such as the correlation between policies and costs) lead to difficulties in fitting mathematical models.

Dynamic and uncertainty management: Unforeseen factors in long-cycle projects (such as extreme weather and supply chain disruptions) increase the risk of prediction bias. Frequent engineering changes lead to the need for dynamic correction of predictions, but the existing methods lack the ability of real-time adjustment. The absence of a collaborative mechanism, the separation of design, construction and cost management links, and the insufficient information sharing have exacerbated the prediction error. The application of digital tools such as BIM is insufficient, and the ability of collaborative cost prediction throughout the entire life cycle is weak.

The cost of highway engineering is interwoven and influenced by multiple stages and dimensions, and a dynamic prediction system covering the entire life cycle needs to be established. It is necessary to study big data integration, BIM model optimization, such as dynamic risk quantification, and cross-professional collaboration, etc..

2.2. Grounded theory

Grounded theory research method is a qualitative research approach that uses a systematic procedure to develop and generalize grounded theories based on a certain phenomenon. It has become a qualitative research method widely applied in fields such as social sciences, health research, education, and information technology.

Grounded theory particularly emphasizes enhancing theory from data, holding that only through in-depth analysis of data can a theoretical framework be gradually formed. This is a process of induction, continuously condensing the data from bottom to top. Unlike general grand theories, grounded theories do not conduct logical deductions based on the assumptions set by the researchers themselves in advance, but rather carry out inductive analysis starting from the data. A theory must be traceable to the original materials from which it originated and must be based on empirical facts. This is because grounded theorists believe that only theories derived from data have vitality. If the theory is consistent with the data, the theory will have practical uses and can be used to guide people's specific life practices.

Unlike traditional theoretical construction methods, grounded theory does not preposition any assumptions, avoiding the interference of subjective inference. It ensures that the theory is based on solid facts, thus being more valuable in application and guiding, and closer to practical operations. Therefore, grounded theory, as a tool for discovering the facts behind laws or values, is particularly suitable for studying the influencing factors of highway engineering costs, in order to construct a complete framework for analyzing influencing factors.

2.3. Random forest learning algorithm

Random Forest is an ensemble learning method, mainly used for classification, regression and other tasks. It improves the accuracy and stability of the model by constructing multiple decision trees and combining their results.

The cost of highways is affected by nonlinear factors such as geological conditions, material prices and construction periods. Random forests can effectively capture complex relationships through the integration of multiple decision trees. Project cost often involves dozens of features (such as construction area, terrain complexity, labor cost, etc.), and random forest reduces redundant interference through random feature extraction. Engineering data often contain outliers (such as sudden cost changes caused by extreme weather), and the Bootstrap sampling mechanism of random forest reduces the sensitivity to individual outlier samples.

Extract key features such as the bill of quantities (earthwork volume, thickness of pavement structure layer), geographical parameters (slope, soil type), and economic indicators (steel price index, labor cost). The interquartile range, like variance and standard deviation, represents the dispersion of various variables in statistical data. However, the interquartile range is more of a robust statistical filter for extreme cost data to prevent the model from being disturbed by unreasonable samples. The key cost drivers (such as asphalt usage and machinery rental costs) can be identified through the ranking of feature importance.

We introduce derived features, such as unit area cost and material price volatility, to enhance information density. The decision tree node splitting strategy is improved by combining ridge regression to enhance the ability to capture linear relationships.

Case reference: Based on the data of 30 historical sections, the prediction error (MAE) of random forest is 38% lower than that of linear regression, and the accuracy rate of key feature recognition reaches 92%. Compared with neural networks, random forests train faster in scenarios with small samples (< 100,000), have stronger interpretability, and are convenient for engineering audits. Random forests can provide high-precision and interpretable solutions for highway cost prediction, especially suitable for multi-variable and nonlinear engineering scenarios.

2.4. The advantages of BIM in predicting highway cost

Based on BIM 3D models, automatic generation of engineering quantity data, such as earthwork volume and material usage, significantly improves efficiency and reduces human errors compared to traditional manual or two-dimensional software calculations. Complex components are precisely processed. For irregular structures, such as curved terrain and pipeline intersections, solid subtraction calculations are carried out to avoid omissions or miscalculations and ensure the objectivity of the data. Compared with the two-dimensional design, the deviation rate of the automatically generated engineering quantity report is significantly reduced, and the engineering quantity calculation is efficient and accurate.

Dynamic response to design changes, parameter linkage update. After the model parameters are modified, the engineering quantity and cost data are automatically associated and updated to avoid duplicate calculation. In order to optimize the cost of conflict detection, we conduct early detection of pipeline collisions, elevation conflicts, etc., to reduce rework and changes during the construction phase and control the expenses that exceed the budget.

BIM provides dynamic cost prediction throughout the entire process, investment decision support, integrates historical project databases, quickly generates investment estimation indicators and economic parameters, and enhances the scientific nature of investment plans. Phased cost simulation outputs real-time cost data at each stage of decision-making, design and construction to assist in dynamic cost control, such as capital planning and material procurement planning. Multi-scheme comparison and optimization: By adjusting design parameters such as route slope and structural layout, real-time cost comparisons of different schemes are generated to select the most economically viable one.

The advantages of collaboration and data integration enable a unified platform to integrate professional models such as roads, Bridges, and electromechanical systems. Multi-disciplinary collaboration avoids cost errors caused by information silos. Data traceability, complete recording of model versions and change history, supports cost audits and full-process cost traceability.

In order to enhance the cost-effectiveness of the business and free up employees' energy for undertaking high-value tasks, automated quantity calculation liberates cost engineers from cumbersome calculation to risk assessment, inquiry optimization and other core tasks. By integrating with emerging technologies and combining AI and big data analysis, such as Glodon's "AI+ Data + platform" solution, the intelligence and refinement of cost prediction can be achieved.

BIM technology, through its three core capabilities of three-dimensional visualization modeling, automatic data integration, and dynamic cost simulation , has addressed the pain points in highway cost prediction such as large deviations in engineering quantities, lagging response to changes, and difficult coordination among multiple specialties, significantly enhancing the accuracy, timeliness, and full-process controllability of cost prediction . With the in-depth application of "BIM+AI" technology, its advantages will be further expanded. This paper is based on grounded theory and optimization algorithms to improve the prediction accuracy.

3. Analysis of Influencing Factors of Highway Engineering Cost

3.1. Identification and Analysis of External Influencing Factors Based on Grounded Theory

Grounded theory is a method of summarizing theories from raw data and is applicable to exploring the external influencing factors of highway engineering cost. This study collected relevant literature, policy documents, construction data, etc., and conducted an analysis using the three-level coding method of grounded theory.

Relevant literature and materials on highway cost prediction have been collected, including research papers on the influencing factors of highway cost, industry reports, etc., covering various aspects such as policies and regulations, technical standards, and market environment for highway construction. After the initial organization of these materials, the first-level coding begins, decomposing the data into several concepts. Secondary coding classifies the relevant concepts to form categories. The three-level coding further integrates the categories, determines the core categories, and constructs the theoretical framework of external influencing factors.

Through grounded theory analysis, it is identified that the external influencing factors of highway engineering cost mainly include the natural environment, economic environment, policies and regulations, etc. Natural environmental factors include topography, climate, geological conditions, etc. These factors directly affect the route design, construction difficulty and engineering volume of highways. Economic environmental factors include price levels, labor costs, capital supply, etc., which affect the direct and indirect costs of engineering construction. Policy and regulatory factors such as relevant policies issued by the government and industry standards have an impact on aspects such as the investment scale and construction standards of highway construction.

The analysis results based on grounded theory show that the external influencing factors of highway engineering cost are complex and diverse. The topographic conditions among the natural environmental factors are particularly crucial in the construction of mountainous roads. Complex terrain will increase the construction costs of structures such as Bridges and tunnels. Price fluctuations among economic environmental factors will directly affect the procurement costs of materials and equipment, and thereby influence the project cost. Changes in policies and regulations may bring about additional construction costs or alter the construction standards of the project, having a significant impact on the cost. These factors interact with each other and jointly influence the cost level of highway engineering.

Applying grounded theory to conduct a systematic analysis of the external influencing factors of highway engineering cost can provide comprehensive and scientific theoretical support for the subsequent construction of cost prediction models, and is conducive to grasping the influencing mechanism of highway engineering cost more accurately.

3.2. Analysis of Influencing Factors based on the characteristics of engineering projects

The cost of highway engineering is not only affected by the external environment, but also closely related to the characteristics of the project itself. The characteristics of different engineering projects directly affect the volume of work, construction difficulty and resource input, and thereby determine the total cost of highway engineering. The analysis of influencing factors based on the characteristics of engineering projects can provide a more accurate basis for cost prediction.

The cost of highway engineering includes construction and installation project fees, land use and demolition compensation fees, other construction project fees, contingency reserves, etc. Among them, the proportion of construction and installation engineering costs is the highest, mainly including the costs of subgrade engineering, pavement engineering, bridge and tunnel engineering, intersection engineering and facilities along the line, etc. Subgrade engineering involves earthwork and stonework, special subgrade treatment, etc. Pavement engineering includes pavement materials, thickness of structural layers, etc. Bridge and tunnel engineering includes key indicators such as the length of Bridges, the length of tunnels, and the ratio of Bridges to tunnels. The cross-engineering and facilities along the line cover the number of interchanges, passages, service areas and parking areas, etc.

The factors such as the route length, roadbed width and earthwork volume of subgrade engineering directly affect the cost. In pavement engineering, the type of pavement materials, the thickness of the structural layer, etc. have a significant impact on the cost. The characteristics of bridge and tunnel engineering, such as the length of Bridges, the length of tunnels, and the ratio of Bridges to tunnels, determine the construction costs of Bridges and tunnels, and the construction costs of Bridges and tunnels are usually relatively high. The number of interchanges, passageways, service areas and rest areas in the cross-project and along the line facilities will also affect the total cost of the highway project.

The framework of influencing factors based on the characteristics of engineering projects was verified through literature verification and questionnaire surveys. Literature verification indicates that characteristic factors such as subgrade width, bridge-tunnel ratio, and pavement material type have been mentioned in multiple studies. The results of the questionnaire survey show that characteristic factors such as terrain, climate, roadbed width, bridge-tunnel ratio, and pavement material type have significant influences on the cost of highway engineering.

The analysis of influencing factors based on the characteristics of engineering projects can systematically identify the key driving factors of highway engineering cost, providing a scientific basis for the subsequent construction of prediction models. By conducting an in-depth analysis of the characteristics of engineering projects, the influencing mechanism of highway engineering cost can be grasped more accurately, providing strong support for investment decisions and cost control.

3.3. Analysis of Vector Machine Prediction Framework

To ensure the accuracy and comprehensiveness of the influencing factors of highway engineering cost, it is necessary to verify and analyze them. Based on the completed highway engineering cases, extract the data of important influencing factors. To ensure the uniformity and comparability of the data, it is uniformly transformed into standardized numerical variables. It may involve the standardization processing of different types of data, such as encoding categorical variables as numerical variables, or normalizing continuous variables. After the above steps, the dataset will contain multiple numerical variables with consistency and operability, providing a foundation for subsequent model training and experiments.

The random forest algorithm is used to screen the importance of feature factors in the data set. During the feature screening process, by replacing a certain feature in the data, the importance of this feature for the prediction of project cost is measured according to the degree of change in model performance. We identified the key factors that have a significant impact on cost prediction and used them for the improved calculation of the vector machine. Through experiments, the feature set that can optimize the prediction performance is selected to ensure that the important factors are observed intensively during cost prediction.

The improved search algorithm is adopted to optimize and calculate the preferred important parameters of the improved vector machine algorithm, further improving the prediction accuracy of the model. The improved vector machine algorithm after parameter optimization will be used for regression calculation of highway engineering cost. After inputting the screened feature data, the model will predict the cost of the target project through the optimized parameters, thereby obtaining the predicted value that is closest to the actual situation.

4. Establishment of BIM Highway Engineering Prediction Model

4.1. The Relationship between Vector Machine Algorithm and BIM

The combination of improved vector machine algorithms (such as integrated support vector machines, kernel function optimization, etc.) and Building Information Modeling (BIM) drives data optimization, supports intelligent decision-making and collaborative management, and enhances the efficiency of highway engineering cost estimation.

BIM integrates multi-dimensional data such as geometric information, material properties, construction progress (4D), and cost (5D), providing a structured data foundation. Provide high-quality input sources for the vector machine algorithm. Vector machines enhance the analytical ability of BIM models. By integrating multiple weak classifiers, the predictive robustness for nonlinear issues such as construction risks and material consumption is enhanced. For the high-dimensional features in BIM (such as terrain complexity and component relationships), the kernel functions (such as Gaussian kernels) are improved to capture the nonlinear correlations among the features more accurately.

The construction parameters based on BIM (such as shield tunneling parameters and mechanical configuration) are input into the improved vector machine model, and the optimal construction plan is matched through intelligent algorithms to reduce the cost of manual trial and error. We employ the improved vector machine to conduct hierarchical early warnings for hazardous areas, such as landslide zones and structurally vulnerable points, in order to guide the path planning of unmanned aerial vehicles.Combined with the BIM-5D model, the dynamic optimization of material and labor costs is achieved.

The BIM model provides precise component coordinates and assembly logic for robots, while the improved vector machine optimizes the robot's action decisions (such as the path planning of the mechanical arm), forming a closed loop of "BIM data → algorithm decision-making → robot execution", enabling real-time obstacle avoidance and dynamic adjustment.

The integration of vector machine improvement algorithm and BIM is essentially a bidirectional enhancement of 'data precision' and 'decision intelligence' , driving the construction process towards automation and high precision. By combining ensemble models such as random forests and support vector machines, we have constructed the BIM intelligent decision center. This center is capable of adapting to edge computing devices (such as on-site inspection robots) and achieving low-latency responses.

4.2. The establishment method of BIM highway engineering cost prediction model

The BIM highway engineering cost prediction model established based on the latest industry practices focuses on three core links: terrain modeling accuracy, standardization of component libraries, and dynamic database updates, integrating multi-dimensional technologies and management processes.

By using unmanned aerial vehicle (UAV) oblique photography technology, we collect terrain data with centimeter-level accuracy and generate three-dimensional real-scene models, which serve as spatial benchmarks for cost prediction. Through the large-scale Civil engineering design software Civil 3D, we import data such as elevation points and contour lines, establish digital terrain surfaces, and set the maximum triangle side length (such as 100m) to eliminate outliers.

For engineering structure modeling : we convert CAD road centerlines to Civil 3D route objects, or directly design parametric plan-longitudinal curves. We invoke standardized components such as lanes, slopes, and drainage ditches to assemble cross-sections according to the step lengths of pile numbers. We assign time parameters to the components to form a 4D model and correlate it with the geological model to optimize the excavation plan.

Based on the component codes of the BIM model (such as concrete grade and steel bar type), the computer automatically calculates the material usage and sub-item quantities. We utilize AI image recognition technology to parse two-dimensional drawings and complete the quantities of work not covered by BIM. We associate enterprise quotas or government quota libraries and access real-time market price fluctuation data (such as steel, asphalt, etc.). We use machine learning algorithms to analyze historical engineering data and predict the changing trends of labor and mechanical rates. We utilize a dynamic cost database to generate cost comparison files for multiple schemes. We adjust the parameters of assembly components (such as road surface thickness) to quickly recalculate the costs. The computer automatically identifies interface conflicts in the model (such as pipeline collisions), reducing change costs.

Key technical support: We compress the model data volume through version control, geometric transformation, and BIM processing modules, etc., to enhance the loading efficiency of the web page. We have established a cloud-based BIM workbench to synchronously update design changes to the cost model (such as the longitudinal slope modification linked to the recalculation of earthwork volume). We use professional software (such as Svel Road Edition, Glodon) to review the compliance of the list, avoiding omissions and measurement deviations.

AI deep integration : based on expert opinions and grounded theory research on the influence laws of highway engineering cost, we use the random forest algorithm to screen and optimize indicators, and establish a characteristic indicator system for highway cost prediction. Support vector machines are adopted to establish prediction rules. The computer generates cost constraint conditions in the BIM model. Combined with the dynamic data of construction progress, the full-cycle cost prediction is corrected.

4.3. BIM Highway Engineering Cost prediction model

The BIM highway engineering cost estimation model integrates multi-dimensional data and intelligent analysis technology, achieving full-process cost control from design to construction, and greatly improving the accuracy and degree of computational automation. The traditional analogical empirical valuation method and parameter valuation method have low accuracy and weak scientificity. BIM software has advantages such as parameterization, visualization, collision checking and convenient drawing generation. It can be directly used for calculation and statistics, improving the speed and accuracy of real-time calculation of cost engineering quantities.

The model based on Building Information Modeling (BIM) integrates terrain data, such as digital elevation models and digital surface models, with design parameters, material prices and historical engineering data. We automatically extract the engineering quantity through the BIM model, reduce manual calculation errors, and support real-time updates of engineering change information.

We use regression analysis and equivalent linear models for cost estimation. By integrating BIM big data mining technology, we construct a time series that can predict the fluctuations in material prices and the trends of construction costs. We introduce the fuzzy correlation degree scheduling method, taking cost, quality and efficiency as constraint indicators, to achieve multi-factor joint estimation. We can identify potential risks during construction, such as fluctuations in material prices and geological conditions, through the risk assessment platform and generate response strategies. Combined with the resource allocation management module, the BIM model can provide an optimized selection solution for project management. We integrate the technical process and implementation steps, data collection and modeling, and import the highway design drawings into the BIM platform. BIM can automatically generate a three-dimensional model, associate the attributes of the engineering quantity, and also integrate historical cost data and market price information to establish a machine learning-driven predictive model.

The BIM platform can adjust the cost forecast based on the dynamic data of construction progress and monitor the budget deviation in real time. The BIM model generates intermediate measurement tables and outputs value reports to support the online approval process. BIM models can automatically generate cost documents that comply with standards, such as bill of quantities and budget, and display cost distribution and risk points through a visual interface.

BIM enhances the efficiency of cost budgeting, shortens the cost estimation time by over 80%, and keeps the accuracy error within 3%. BIM saves approximately 10% of the contract amount through conflict checking and reduces the number of changes not listed in the budget by over 40%. BIM full-cycle control, covering investment decision-making, design, construction to settlement stages, supports multi-stage data traceability and collaborative management. BIM resolves the lagging issue of traditional quota management and ADAPTS to the rapid iteration demands of new processes and technologies.

5. Application of Highway Cost Prediction Model

5.1. Project Overview Case

This article selects a new expressway project as a case. The total length of this project is 165 kilometers. It is constructed to the standard of a two-way four-lane expressway, with a roadbed width of 24.5 meters and an asphalt concrete pavement. The terrain of the project is mainly mountainous and flat, and the climate belongs to the humid and hot zone in the southeast. It is constructed under the traditional unit price contract model. The main features include a bridge-tunnel ratio of 13.8% and a earthwork volume of 272,000 cubic meters per kilometer, etc. This case covers the typical characteristics of highway engineering and is applicable to verifying the constructed cost prediction model.

This article collects data on 13 characteristic factors of 63 highways across the country from 2013 to 2022, including construction time, construction environment, highway grade, etc., as well as cost data per kilometer. To eliminate the dimensional influence, the extreme value method is adopted to normalize the data and map the data to the interval of [0,1]. The processed data is used for subsequent model training and prediction to ensure the stability and accuracy of the model.

5.2. Screening of characteristic factors

To optimize the input variables of the prediction model, the random forest algorithm is adopted to quantitatively analyze the importance of features, and the backward elimination strategy is combined to screen the optimal feature combination. Random forest calculates the contribution of each feature to the prediction accuracy through permutation importance. The number of decision trees is set to 500, and 4 features are randomly selected in each split. The importance of 13 candidate features is ranked (Figure 1). The results show that engineering characteristics such as the bridge-tunnel ratio, terrain, and roadbed width, as well as external factors such as the construction environment and climate, are of high importance, while characteristics such as design speed and construction time contribute less.

The influence of feature combination on the model accuracy is further verified through the backward elimination experiment: Starting from the full feature set (13), the features with the lowest importance are removed successively (the design speed is eliminated for the first time, and the construction time is eliminated for the final time), and k features are retained each time (k=13,12,...). We constructed a least squares support vector machine model to calculated the root mean square error (RMSE). As shown in Figure 2, RMSE shows a trend of first decreasing and then increasing with the decrease in the number of features. When k=11, RMSE reaches the minimum value of 141.69, indicating that at this time, the feature combination can not only retain the key information but also avoid the interference of redundant variables. The 11 optimal features finally determined include: bridge-tunnel ratio, topography, roadbed width, earthwork volume per kilometer, pavement material type, land acquisition indicators, construction environment, climate, highway grade, number of interchanges per kilometer, and construction mode. The design speed and construction time that were excluded were not included in the core feature set due to their high correlation with other features or weak impact on cost.

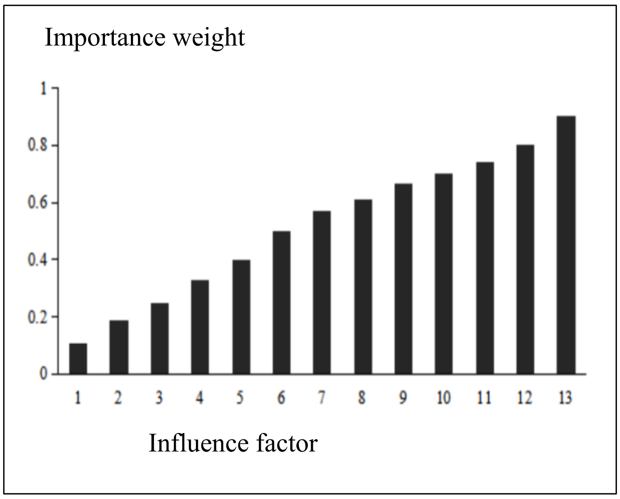

Figure 1 Ranking of the importance of characteristic factors

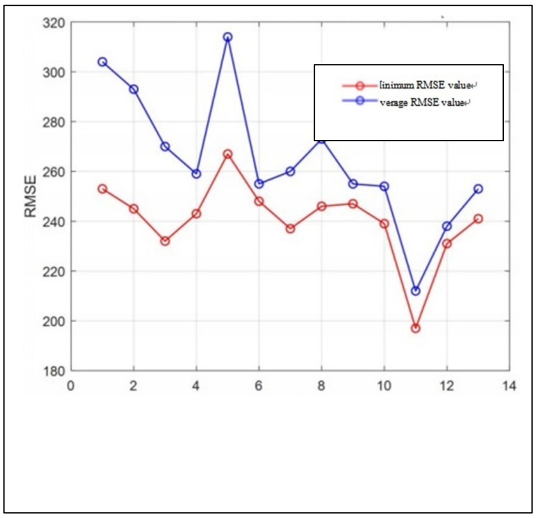

Figure 2 Minimum and average RMSE values under different numbers of features

The feature screening results show that the random forest can effectively identify the key factors. Combined with the backward elimination strategy, it can significantly improve the model efficiency, laying the foundation for the subsequent sparrow search algorithm to optimize the parameters of the least squares support vector machine.

Compared with the BP neural network and the traditional least squares support vector machine model, the prediction model based on least squares support vector machine performs the best in all indicators: R^2 increases by 33.8% compared with the BP neural network, and RMSE decreases by 52.2%. Compared with the unoptimized support vector machine model, R^2 increased by 26.4% and RMSE decreased by 40.7%. The results show that the random forest feature screening effectively reduces the input dimension, and the parameter optimization of the sparrow search algorithm significantly improves the generalization ability of the support vector machine, verifying the modeling advantages of multi-algorithm fusion for complex nonlinear relationships.

The model operation analysis shows that the least squares support vector machine has both high efficiency in feature screening and high accuracy in prediction in the small sample scenario, providing a reliable technical path for the cost estimation in the early stage of highway engineering.

5.3. Predictive model operation analysis

The Sparrow Search algorithm (SSA) is used to globally optimize the parameters (penalty factor γ and kernel width σ) of the least squares support vector machine. During the iteration process, the fitness curve (Figure 3) shows that the algorithm converdes nearly at the 10th iteration and finally obtains the optimal parameter combination (γ=85.6, σ=12.3). It indicates that SSA has an efficient global search ability under the condition of small samples and effectively avoids the blindness of the traditional trial-and-error method.

The 11 screened features were input into the RF-SSA-LSSVM model. The prediction results of the test set are shown in Figure 4. The predicted values fit closely with the true values, and the absolute error is concentrated within ±2 million yuan /km. The quantitative indicators show that the goodness of fit R^2=0.91, indicating that the model explains 91% of the cost fluctuations. The root mean square error RMSE=141.69, and the mean absolute error MAE=92, which were reduced by 41.2% and 16.9% respectively compared with the full-feature LSSVM model. In the case verification, the relative error of the ND expressway prediction was 4.15%, meeting the accuracy requirements of the decision-making stage (MRE<10%).

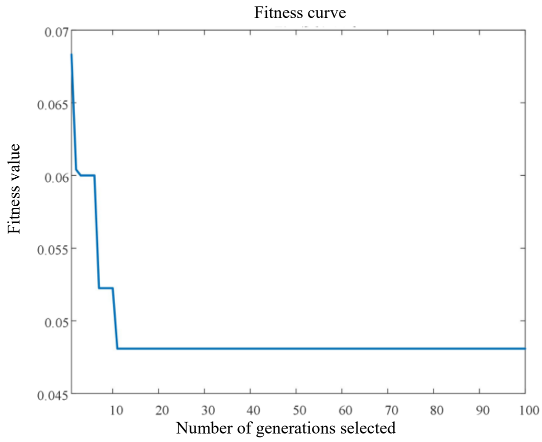

Figure 3 Fitness curve

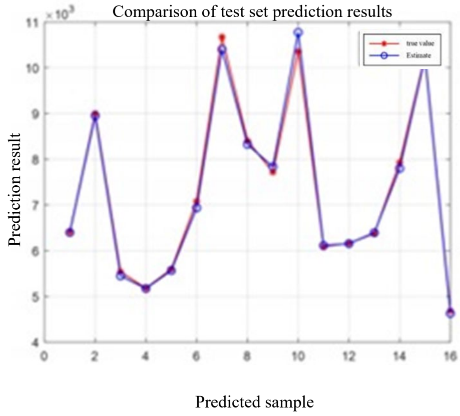

Figure 4 Model prediction effect diagram

Compared with the BP neural network and the traditional least squares support vector machine model, the prediction model based on least squares support vector machine performs the best in all indicators: R^2 increases by 33.8% compared with the BP neural network, and RMSE decreases by 52.2%. Compared with the unoptimized support vector machine model, R^2 increased by 26.4% and RMSE decreased by 40.7%. The results show that the random forest feature screening effectively reduces the input dimension, and the parameter optimization of the sparrow search algorithm significantly improves the generalization ability of the support vector machine, verifying the modeling advantages of multi-algorithm fusion for complex nonlinear relationships.

The model operation analysis shows that the least squares support vector machine has both high efficiency in feature screening and high accuracy in prediction in the small sample scenario, providing a reliable technical path for the cost estimation in the early stage of highway engineering.

6. Conclusion

The engineering quantity information of highway engineering cost is constructed by using grounded theory, the ontology of engineering quantity information of highway engineering cost is built by BIM, the ontology instance and specification instance are converted by BIM standard format, and the rules are set to realize the prediction of engineering quantity of highway engineering cost.

The problems in project cost prediction: the investigation of influencing factors is not comprehensive, the prediction is not accurate enough, and it is not automated enough. It is necessary to use the method of combining artificial intelligence and BIM to build an effective and highly accurate cost prediction model.

We should comprehensively analyze the internal and external influencing factors of highway engineering cost. It can be summarized into four main categories: natural environment, economic environment, decision-making information and organizational management. Taking regional factors into consideration, the overall framework of external influencing factors of highway engineering cost was systematically established.

The random forest algorithm is introduced into the highway engineering cost prediction model. The optimal characteristic factor set is obtained through quantitative analysis as the bridge-tunnel ratio, terrain, roadbed width, earthwork volume per kilometer, pavement material type, land acquisition index, construction environment, climate, highway grade, number of interchanges per kilometer, and construction mode, which are used as the input characteristic factors of the subsequent model.

We should make full use of the advantages of the least squares support vector machine in small sample prediction to establish a cost prediction model for highway engineering. Taking the characteristic factor data as the input and the unit mileage cost as the output, the computational efficiency and prediction accuracy of the model have been improved.

The least squares support vector machine based on the preferred feature factors of random forest was used for the reference of highway cost prediction. Through the simulation and training of actual engineering data, the accuracy of the intelligent BIM highway engineering cost prediction model was verified.

Acknowledgements

The author is extremely grateful to the following sponsors for their support of this research: the Soft Science Project of Sichuan Provincial Department of Science and Technology, "Research on Intelligent Cyber-Physical Numerical Control System Based on Physical Motion Laws", 25RCYJ0011.

References

[1] Tang Ming, Liu Weijun, Wang Le. Analysis of the Impact of Macroeconomic Variables on the Cost of Highway Engineering [J]. Construction Economics,2021, 42(03):111-114.

[2] Yu Tiancai. Discussion on the Influencing Factors of Expressway Project Cost in Mountainous and High-Altitude Areas [J]. Highway, 2016, 61 (01): 177-180.

[3] Que Yunlong, Wei Kang. Analysis of Influencing Factors of Cost Deviation of Border Highways [J]. Highway, 2023, 68 (08): 254-258.

[4] Zheng Xiaolei, Zhang Shilian. Research on Highway Engineering Cost Prediction Model Based on Main Characteristic Factors and BP-GEP Network [J]. Highway Engineering, 2018, 43 (01): 206-210.

[5] Li Junda, Li Yuanfu, Wang Guangkai. Highway Engineering Cost Estimation Model Based on CBR [J]. Highway Traffic Science and Technology, 2020, 37 (06): 44-49+67.