A whole process cost prediction system for construction projects based on improved support vector machines

Xueqing Zhang,
CPI Power Engineering Co. Ltd.,
Shanghai, 200233,
China

Jie Song,
CPI Power Engineering Co. Ltd.,
Shanghai, 200233,
China

Chaolin Zha,
clzha@fudan.edu.cn
Sichuan Dincen Electrical Engineering Co., Ltd.,
Chengdu, 610056,
China

Abstract—The current project cost system requires high data scale, small amount of data and large prediction deviation. In order to improve the prediction accuracy of the whole process cost of construction project, this paper designs a whole process project cost prediction system based on improved support vector machine. In the hardware part of the system, the control core adopts arm controller S3C6410 and introduces 4G communication module to analyze the actual engineering data with the support of hardware. In the software part, the whole process cost prediction index system of the construction project is established, the index is reduced by the principal component method, and the support vector machine is improved by particle swarm optimization algorithm to realize the whole process cost prediction of the project. The system function test results show that the average prediction deviation of the designed system is 4.11%, the average prediction deviation of the cost prediction system is 3.05%, and the average prediction deviation of the system is 1.57%.

I. INTRODUCTION

THE construction market is developing rapidly, and the reform of engineering costing mode is deepening. Engineering construction is a process, it is composed of a series of complex system engineering, and the main factor throughout the whole process is the construction price of the project. There are great uncertainties in the process of engineering construction, which will make the construction price change, such as the pre-project feasibility study report, which determines the optimal construction program, which in turn directly affects the choice of engineering design program. According to the design plan to make the bill of quantities, the

bidder to quote the unit comprehensive price to determine the construction unit, in the construction process of engineering changes, engineering claims and add variables to the completion settlement. All in all between the various systems, the factors that affect each other is the project cost, the forecast of the construction price of the project is also known as the valuation of the project, the so-called project valuation is the forecast of the expected price of a planned project ready for construction before its implementation [1]. Forecasting a project both quickly and accurately is a crucial task for the owner, the consultant commissioned by the owner, and the bidder. At each stage of construction, the owner has to forecast the full expected price of the proposed project, and accordingly determine the bid price, evaluate and fix the bids; for the bidder. The project cost prediction is the key to the success or failure of the bid and the profitability in the process of project implementation [2].

In foreign countries, the current research on engineering cost prediction is mainly focused on prediction models and applicability of the models. Principles such as neural networks and regression method estimation are used to predict engineering cost [3]–[5]. The principles of fuzzy mathematics, neural network, and gray modeling are introduced in China to predict engineering cost [6]–[10]. Although the above methods have good prediction effect on engineering cost, neural network, gray model, etc. have strong dependence on historical data, and the robustness and convergence of the algorithm can be guaranteed only when the training sample size is considerable, and their prediction effect is not satisfactory when construction enterprises have difficulty in obtaining sufficient amount of historical engineering data. In addition, some scholars have proposed some cost prediction methods. For example, Phung et al. developed a prediction system integrating low-cost sensors [11]; Kim used the additive risk factor method to design a cost and time risk assessment method for large-scale projects [12]; Keshavarz and Shoul studied the trade-off between time cost and quality of the project based on fuzzy decision-making, so as to complete the prediction of construction project cost [13]. At that time, there are still some pattern recognition problems in the above methods.

Support vector machine is proposed for the pattern recognition problem and can obtain the best generalization performance by maximizing the classification interval under the premise of zero empirical risk, which makes it have better learning ability under the condition of small sample learning. In addition, support vector machines transform complex learning problems into simple problems in a high-dimensional linear space to solve, enhancing the reliability and control of the algorithm. Since the project entity is composed of several sub-projects of each professional engineering, the cost changes of different sub-projects will have an impact on the overall construction project cost, and the cost prediction of the whole process of construction project from the whole process of construction can obtain the actual input closest to the completion of the project in a more comprehensive way. In order to improve the accuracy of cost prediction for the whole process of construction, this paper will improve the support vector machine and use the improved support vector machine to design the cost prediction system for the whole process of construction.

II. HARDWARE DESIGN OF WHOLE PROCESS COST PREDICTION SYSTEM FOR CONSTRUCTION PROJECTS BASED ON IMPROVED SUPPORT VECTOR MACHINE

A. Hardware Architecture Design of Engineering Cost Forecasting System

Fig. 1 below shows the block diagram of the hardware architecture of the engineering cost prediction system designed in this paper.
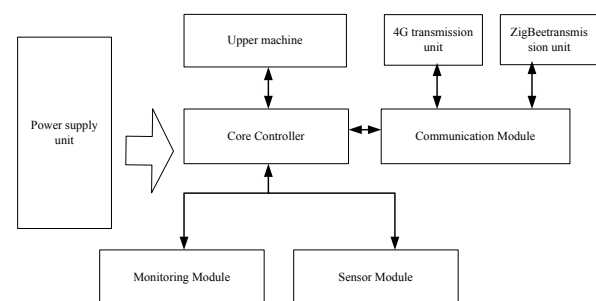

Fig. 1 Block diagram of hardware architecture of engineering cost prediction system

The ARM controller S3C6410 is selected in this design, which is a low-cost, low-power, high-performance microprocessor running at up to 667MHz, with a good display

interface and more communication interfaces, and high processing power to meet the application requirements of this design. The sensor module is used to collect the basic information such as temperature and humidity during the construction process. The monitoring module is composed of several cameras and handheld test instruments to collect the construction quality data in real time. In order to improve the data communication speed, this design mainly optimizes the design of communication module.

B. Optimal Design of Communication Module

The internal communication of the construction quality monitoring module and the communication between this module and the sensor module is done through the RS485 interface. In order to avoid RS485 interface chip breakdown and damage, the design selects a high level ESD protection chip SP3485, which has a high load carrying capacity to ensure both transmission distance and transmission reliability, and has a low power consumption.

In this topic, wireless data transmission is not only to transmit common sensor signals, but also to transmit video data. The video data transmission volume is large, usually after compression, to ensure stable data transmission, the occupancy of network traffic reaches several megabytes to more than ten megabytes, so we have to choose a suitable 4G transmission method to complete the wireless transmission of video data. This design selects the USR-G781 4G LTE DTU communication module. The G781 is an industrial 4G router with powerful DTU function, providing users with an integrated solution of industrial 4G router and DTU.

Sensor monitoring data acquisition communication is implemented using ZigBee communication chip. The WLT2408NZ module is planned to be selected in this design. The module has a maximum output power of 8 DBM, and the transmission distance can reach more than 200 m under unobstructed conditions when the external 5dbi antenna is connected, and the working frequency band is 2. 380 GHz to 2500 GHz, and there are 9 extended frequency bands in addition to the 16 channels of standard ZigBee, which can effectively avoid the interference of other 2.4 G signals such as Wi-Fi and Bluetooth [14].

Supported by the hardware architecture of the prediction system designed above, the functional design objectives when predicting the whole process cost of construction projects are analyzed, and the software part of the prediction system is designed and studied using improved support vector machine algorithms.

III. SOFTWARE DESIGN OF WHOLE PROCESS COST PREDICTION SYSTEM FOR CONSTRUCTION PROJECTS BASED ON IMPROVED SUPPORT VECTOR MACHINE

A. Construction of whole process cost prediction index system for construction projects

Through the analysis of specific cases, it is clear that in the actual project, the component project cost accounts for the largest proportion of the total cost. Therefore, this cost is taken as the main object of analysis, and the four parts of construction features, structural foundation, decoration and installation works are taken as the entry point for refinement of cost prediction index. In addition, the measure item costs, other item costs, fees and taxes in the total project cost are recorded according to the rates specified in each place, and their amount is closely related to the component project costs and also influenced by the external environment in which the project is located, so for such indirect costs, the main influencing factors will be identified from the perspective of project characteristics [15].

In summary, the four indicators related to construction characteristics, structural foundation, installation engineering and other project characteristics are taken as the first-level indicators. At the same time, according to the main characteristics of specific engineering cases and the principles of index selection, the first-level indicators are further divided, and 24 second-level indicators can be initially obtained, so as to reflect the common characteristics of construction projects in a comprehensive and objective way.

For this reason, combining with the basic principles of construction cost prediction index system construction, the expert scoring method is chosen to remove the redundant indexes which have less influence on the project cost prediction effect. Questionnaires were issued to 100 experts from technical departments of construction units, construction units and design units respectively. The five-level rating method was adopted: "very important", "very important",

"important", "general", "unimportant", respectively. "unimportant", corresponding to the quantitative values 1, 2, 3, 4, and 5, respectively, and the importance of a number of obtained characteristic indicators was assessed according to the practical experience of experts [16].

In order to conduct quantitative analysis of expert opinions, the expert opinion concentration degree $\tilde{E}_i$ and the expert opinion dispersion degree σ_i are used to reflect the expert rating, which are defined as follows [17] :

$$\tilde{E}_i = \frac{1}{p}\sum_{j=1}^{5} E_j n_{ij} \qquad (1)$$

$$\sigma_i = \sqrt{\frac{1}{p-1}}\sqrt{\sum_{j=1}^{5} n_{ij}\left(E_j - \tilde{E}_i\right)^2} \qquad (2)$$

In the above formula, E_j refers to the value of the j th degree of importance; n_{ij} represents the number of experts who agree that the i th indicator is the j level importance, p represents the number of experts,, and σ_i refers to the degree of dispersion of expert ratings on the importance of the index i . In general, when $\tilde{E}_i \leq 3$ and $\sigma_i \leq 1$, the index item is considered to be important or above. According to the concentration degree and dispersion degree of experts' opinions on each prediction index, the index which has less influence on the project cost prediction is determined and deleted. Therefore, the whole process construction engineering cost prediction index system as shown in Table 1 is established [18].

Table 1. Cost prediction index system for the whole process of construction projects

Indicator System	Primary indicators	Secondary indicators
Construction project whole process cost prediction index system	Indicators related to architectural features x 1	Ground floor area x 11
		Ground floor area x 12
		Number of building floors x 13
		Building floor height x 14
	Structural foundation-related indicators x2	Pile type x21
		Building structure form x22
		Seismic rating x23
	Installation-related indicators x3	Fire protection system type x31
		Elevator and other equipment installation completeness x32
	Other x4	Construction Project Management x41
		Construction operating environment x42
		Construction plan duration x 43
		Duration construction plan volume x 44
		Unit construction volume cost x 45

For construction projects, there are many factors affecting their project costs, and although the representative factors have been initially screened out, there are still direct or indirect relationships between the indicators in terms of quantity, and some factors can be expressed by some existing factors, resulting in a large overlap of sample information.

In order to make the generalization of construction cost prediction based on support vector machine better, this paper proposes to use the principal component analysis method combined with the prediction model, and the integrated variables obtained after the principal component analysis are used as the input vectors of the support vector machine to

eliminate the high correlation among the selected feature indicators, so as to overcome the dimensional disaster faced by the support vector machine to a certain extent and improve the prediction accuracy of the model.

B. Cost forecast index reduction

As a multivariate comprehensive analysis method, principal component analysis can effectively realize the dimensionality reduction of indicators and has been widely applied in many fields. Its basic calculation process is as follows [19]:

(1) Select the initial sample data

Assuming that the sample size is n and each sample has p indicators to represent its various characteristics, the following $n \times p$ matrix is constituted:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (3)$$

(2) Standardized processing of initial sample data

In order to eliminate the influence of large numerical differences and dimensional differences among indicators, Z-score standardization method was adopted and the attribute values of each indicator were processed in accordance with the following formula:

$$x_{ij}^{*} = \frac{x_{ij} - \overline{x}_j}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_{ij} - \overline{x}_j\right)^2}} \quad (4)$$

$$\overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$$

In the above formula, $\overline{x}_j$ is the average value of index data; x_{ij}^{*} is the data processed after Z-score standardization. After the above process, the standard matrix is generated. The correlation coefficient matrix $X^{*}_{n \times p} = \left(x_{ij}^{*}\right)_{n \times p}$ was established, and the eigenvalues R and eigenvectors λ were calculated. Determine the number of principal components according to the following formula:

$$q_i = \frac{\lambda_m}{\sum_{i=1}^{m} \lambda_i} \times 100\% \quad (5)$$

In the above formula, q_i is the variance contribution rate of the m th principal component. Then the cumulative variance contribution rate of the first m principal components is:

$$Q = \sum_{i=1}^{m} q_i \quad (6)$$

To determine the principal component, an upper limit of the cumulative contribution rate should be defined in advance. It is generally believed that when the cumulative contribution rate is determined, the first m principal components corresponding to the eigenvalue reflect most of the information of the original variables, which can replace p indexes as the new input set of the prediction model to complete the index reduction process of the prediction index system. Combined with the prediction index after the above treatment and the improved support vector machine, the whole process cost of the project is predicted.

C. Using improved support vector machine to predict project cost

In this paper, particle swarm optimization algorithm is used to improve the support vector machine, and the improved support vector machine is used to predict the cost of the whole process of construction engineering.

In order to use PSO algorithm to optimize the super parameters of SVM, two important aspects need to be considered in the fusion process of the two algorithms: one is the representation of the super parameters; the other is the definition of the particle fitness function.

For SVM, since the form of kernel function selected in this paper is RBF kernel function, the hyperparameters that need to be optimized are regularization parameter C and kernel parameter η. When the PSO algorithm is used for optimization, each particle is required to represent the possible solution of the optimization problem, that is, the combination of the hyperparameters. Therefore, in the optimization process of SVM, a two-dimensional vector is used to represent the combination of the two hyperparameters C and η, that is, the position of the i th particle can be expressed as

$$XP_i = \left(C_i, \eta_i\right).$$

Fitness is a measure to evaluate the position of particles, and also indirectly describes the generalization performance of

SVM model. In this paper, the mean square error of the prediction results was selected as the performance index of the prediction model, which was mathematically defined as [20]:

$$f = \frac{1}{n}\sum_{j=1}^{n}(y_i - \hat{y}_i)^2 \tag{7}$$

In the above formula, y_i represents the actual output value of the j th sample; $\hat{y}_i$ represents the predicted value of support vector machine. f can be viewed as a composition of C and η. In the fitness evaluation, the smaller the error is, the better the particle position will be. Then the optimization problem can be described as:

$$\min_{C,\eta} f = \frac{1}{n}\sum_{j=1}^{n}(y_i - \hat{y}_i)^2$$
$$s.t. C \in [C_{\min}, C_{\max}], \eta \in [\eta_{\min}, \eta_{\max}] \tag{8}$$

In the above formula, $C_{\min}$ and $C_{\max}$ represent the minimum and maximum values of C; $\eta_{\min}$ and $\eta_{\max}$ represent the minimum and maximum values of η. In the modeling process of SVM, the adaptive value of equation (7) is minimized by searching within the value range of two parameters with the help of PSO, and then the solution vector (C, η) is the optimal super parameter to be found by the improved support vector machine.

After using particle swarm optimization algorithm to improve the support vector machine, the whole process of the project cost prediction is realized according to the following steps:

(1) Index dimension reduction was carried out through sample clustering and principal component analysis to complete the preprocessing of historical data;

(2) To determine the training data set and the prediction data set;

(3) Set the range of parameters C and η., and use the PSO algorithm to initialize the population composed of m particles

(4) The fitness of the particles was calculated and compared with the training data set. The individual optimal value $P_{best}(i)$ and the global optimal value G_{best} were set to update the velocity and position of each particle.

(5) Iterate until the end condition (minimum adaptation value) is satisfied.

(6) Output the optimal hyperparameters and assign them to the support vector machine.

(7) Train the model with the training set and input the prediction sample data for model prediction to obtain the optimal results.

(8) Input the prediction sample data set to the improved support vector machine with determined parameters, and the final result is calculated as the predicted value of the project cost.

The above process realizes the prediction of construction cost using improved support vector machine, and combines the software part of the designed system with the hardware part to complete the design of the whole process cost prediction system of construction project based on improved support vector machine.

IV. SYSTEM TESTING

A. Test Description

After the system hardware and software design is completed, the hardware module and the software part of the system need to be tested by testing the functional implementation of the system. This test is divided into two parts, testing the hardware and software parts of the improved support vector machine based construction project cost prediction system designed in this paper. The main focus of the test is on the software part of the cost prediction function.

Hardware testing is mainly to separately power on each module of the hardware, detect the signal processing and transmission of the module, and verify the adaptation between the system hardware and software by a loading system software part. Software testing is the use of test tools to test the functionality and performance of the product according to the test plan and process, write different test tools as needed, design and maintain the test system, and analyze and evaluate the possible problems of the test plan. After executing test cases, faults need to be tracked to ensure that the developed product meets the requirements.

The software part of the system was tested by comparing the designed prediction system with a support vector machine based prediction system and a genetic algorithm based prediction system with each other (document [19], document

[20] system). The prediction functions of each system were loaded into the same configuration of computers to avoid interference with the experiment. The system test data came from a building construction project at a location that had entered the completion period, and the construction plan of the project and the construction parameters of each sub-project were pre-processed separately to generate the test data set. The whole process cost of the project before construction and the actual construction cost were used as the reference values for this system test. Three cost prediction systems were used to predict the construction cost of a random sub-project and the construction cost of the whole project, and to compare the deviation between the predicted cost and the planned and actual values, so as to evaluate the system's function.

B. Test Data Analysis

Each module of the system hardware designed in this paper was tested independently. The test results show that the hardware module of the system designed in this paper is intact, and after loading the software part of the system, it can make cost prediction, and it is well adapted to the software part of the system, so it can make cost prediction for the whole process of construction of building projects.

Three cost forecasting systems were used to forecast the cost of construction projects based on the data sets, and the forecast comparison results are shown in Fig. 2 and Fig. 3 below.
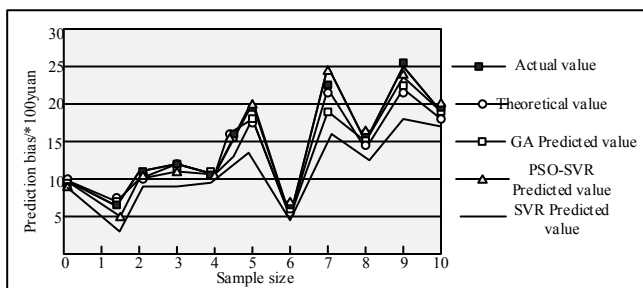

Fig. 2 Comparison of the deviation between the predicted and actual values of the system during the construction of sub-projects
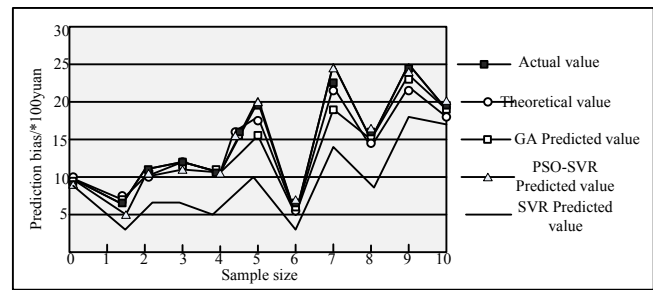

Fig. 3 Comparison of the deviation between the predicted and actual values of construction cost of the whole process of the system

When analyzing Fig. 2, it can be seen that the deviation between this system and the genetic algorithm-based prediction system and the actual cost is smaller when predicting the construction cost of subprojects, but this system is relatively closer to the actual cost value. Analysis of Fig. 3 shows that when analyzing the whole process cost of building construction, the predicted cost value of this system is closest to the actual value, while the deviation between the genetic algorithm-based cost prediction system and the actual cost value is obvious. This indicates that the error of the system based on the prediction of the whole process cost of construction is larger. The average prediction deviation of the support vector machine based cost prediction system is 4.11%, the average prediction deviation of the cost prediction system based on 3.05%, and the average prediction deviation of the system in this paper is 1.57%, which means that the average prediction accuracy of the system in this paper is improved by about 48.5% and the cost prediction deviation is smaller.

V. DISCUSSION

Through the analysis of the above test data, it is proved that the whole process cost prediction system of construction project based on improved support vector machine designed in this paper is effective and feasible for the whole process cost prediction of construction project. The reason why the system designed in this paper has the above effect is closely related to the following points.

(1) In the hardware part of the system, the control core adopts arm controller S3C6410 and introduces 4G communication module to improve the communication speed of the system. With the support of hardware, the actual

engineering data are analyzed, the whole process cost prediction index system of construction project is established, and the index is reduced by principal component method.

(2) In order to improve the effect of cost prediction, this paper also uses particle swarm optimization algorithm to improve the support vector machine, and realizes the cost prediction of the whole process of the project.

VI. CONCLUSION

Construction project cost forecast is an important economic document in the preliminary work of the proposed project and an important basis to prove the economic rationality of the proposed project. The project cost must be reasonably controlled, and the traditional post control must be changed to pre prediction. This paper designs a whole process cost prediction system of construction project based on improved support vector machine. The advantage is that in the hardware part of the system, the control core adopts arm controller S3C6410 and introduces 4G communication module to analyze the actual engineering data with the support of hardware. In the software part, the whole process cost prediction index system of the construction project is established, the index is reduced by the principal component method, and the support vector machine is improved by the particle swarm optimization algorithm, so as to improve the effect of the whole process cost prediction of the project. The research of this system enriches the project cost prediction method, can provide necessary support and help to improve the whole process management accuracy of construction project, and has important theoretical and practical significance. However, because the differences of engineering projects are not considered in the research process, the prediction effect may be affected, which needs to be further analyzed and improved in the next research process.

REFERENCES

[1] S. T. Hashemi, O. M. Ebadati, and H. Kaur, "Cost estimation and prediction in construction projects: A systematic review on machine learning techniques," SN Appl. Sci., vol. 2, no. 10, pp. 1-27, 2020.

[2] S. Sharma, S. Ahmed, M. Naseem, W. S. Alnumay, and G. H. Cho, "A survey on applications of artificial intelligence for Pre-Parametric Project Cost and Soil Shear-Strength estimation in construction and geotechnical engineering," Sensors, vol. 21, no. 2, p. 463, 2021.

[3] P. Kačmáry, D. Malindžák, and J. Spišák, "The design of forecasting system used for prediction of Electro-Motion spare parts demands as an improving tool for an enterprise management," Manag. Syst. Prod. Eng., vol. 27, no. 4, pp. 242-249, 2019.

[4] M. H. Kim and E. B. Lee, "A forecast model for the level of engineering maturity impact on contractor's procurement and construction costs for offshore EPC megaprojects," Energies, vol. 12, no. 12, p. 2295, 2019.

[5] Y. Elfahham, "Estimation and prediction of construction cost index using neural networks, time series, and regression," Alex. Eng. J., vol. 58, no. 2, pp. 499-506, 2019.

[6] L. Zhao, W. Zhang, and W. S. Wang, "Construction cost prediction based on genetic algorithm and BIM," Int. J. Pattern Recogn., vol. 34, no. 7, p. 2059026, 2020.

[7] C. C. Xu, Y. Wang, K. Ye, S. M. Zhang, M. H. Zou, and C. Chen, "Research on transmission line project cost forecast method based on BP neural network," IOP Conf. Series: Mater. Sci. Eng., vol. 688, no. 5, p. 055074, 2019.

[8] Y. Zhang and S. T. Fang, "RSVRs based on feature extraction: A novel method for prediction of construction projects' costs," Korean Society of Civil Engineers, vol. 23, no. 4, pp. 1436-1441, 2019.

[9] H. M. Liu, C. H. Jiang, Y. Liu, M. Hertogh, and X. Lyu, "Optimism bias evaluation and decision-making risk forecast on bridge project cost based on reference class forecasting: evidence from China," Sustainability, vol. 10, no. 11, pp. 1-29, 2018.

[10] M. Fan and A. Sharma, "Design and implementation of construction cost prediction model based on SVM and LSSVM in industries 4.0," Int. J. Intell. Comput., vol. 14, no. 2, pp. 145-157, 2021.

[11] N. K. Phung, N. Q. Long, N. V. Tin, and D. T. T. Le, "Development of a PM (2.5) forecasting system integrating Low-cost sensors for Ho Chi Minh city, Vietnam," Aerosol Air Qual. Res., vol. 20, no. 6, pp. 1454-1468, 2020.

[12] B. C. Kim, "Dependence modeling for large-scale project cost and time risk assessment: additive risk factor approaches," IEEE T. Eng. Manage., to be published.

[13] E. Keshavarz and A. Shoul, "Project time-cost-quality trade-off problem: A novel approach based on fuzzy decision making," Int. J. Uncertain. Fuzz., vol. 28, no. 4, pp. 545-567, 2020.

[14] M. Christantoni and D. Damigos, "Individual contributions, provision point mechanisms and project cost information effects on contingent values: findings from a field validity test," Sci. Total Environ., vol. 624, pp. 628-637, 2018.

[15] S. M. R. Alavipour and D. Arditi, "Impact of contractor`s optimized financing cost on project bid price," Int. J. Proj. Manag., vol. 36, no. 5, pp. 808-818, 2018.

[16] Hirijanto, I. W. Mundra, and A. Utomo, "Cost prediction model based on system dynamics in water resource project," MATEC Web of Conferences, vol. 258, p. 02027, 2019.

[17] M. Cai and X. Li, "Multi-strategy particle swarm optimization based on adaptive selection," Comput. Simul., vol. 38, no. 3, pp. 239-244, 2021.

[18] Y. Alsaawy, A. Alkhodre, M. Benaida, R. A. Khan, "A comparative study of multiple regression analysis and back propagation neural network approaches on predicting financial strength of banks: An Indian perspective," WSEAS Transactions on Business and Economics, Vol. 17, pp. 627-637, 2020.

[19] Y. C. Deng, Y. Xu, D. D. Xu, X. Jia and X. C. Tian, "GNSS time series prediction based on support vector machine," Global Positioning System, no. 2, pp. 70-75, 2019 .

[20] Q. Su, H. J. Zhou and L. Z. Zou, "Research on cost optimization of molten steel "deoxidation alloying" based on genetic algorithm," Henan Sci. Technol., no. 16, pp. 32-34, 2019.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

Xueqing Zhang put forward the research ideas, and formulated the overall research goals

Jie Song organized and executed the experiments.

Chaolin Zha carried out the simulation and the optimization.