Research on Intelligent Investment Prediction Model of Building Based on Support Vector Machine

Yuan-ling Ma, Run-lin Li, and Xiao Ma(✉)

CCTEG Chongqing Engineering Co., Ltd., Chongqing 400042, China
mayuanling2367@163.com, mxiao2546@163.com

Abstract. In view of the imperfection of intelligent construction cost specification, the complexity of cost influencing factors and the lack of historical cost data, the expert system and support vector machine theory are combined to achieve knowledge acquisition and data integration. By using the expert system module, the regression calculation, the establishment of project cost prediction model and the model test of parameter setting and optimization are realized. In addition, the investment prediction speed of the model is faster. Finally, through the empirical data analysis, the accuracy and effectiveness of the model are verified, which provides the economic indicators and reference materials for the design stage of intelligent building projects.

Keywords: Building intelligence · Expert system · Support vector machine · Project cost prediction

1 Introduction

The forecast of building intelligent project investment is the basis to control the project cost of the whole process, the important index used in the project fund reserve management, the project system construction and the project economic benefit analysis. Predicting project investment with high efficiency and accuracy is of great significance for improving project work efficiency, strengthening project cost management and promoting project economic benefits. However, there are very few researches on investment prediction of intelligent building engineering in China, and most project information management is still in a traditional way, which makes it difficult to realize project information exchange [1]. Traditional engineering investment forecasting generally uses simple function regression model, least square method, quota calculation, exponential smoothing method, fuzzy mathematics, gray forecasting and other methods. The above traditional investment prediction methods have simple mathematical principle and fast speed on prediction. But for the investment estimation of large and complex intelligent building projects, their prediction quality is uneven, and the accuracy, effectiveness and practicability are difficult to guarantee [2].

Recently, machine learning technology has been gradually applied in the field of engineering cost predictions. For data mining and model prediction, this technology has higher reliability and accuracy than traditional ones. With the deepening of the

S. Liu and L. Xia (Eds.): ADHIP 2020, LNICST 348, pp. 15–24, 2021.
https://doi.org/10.1007/978-3-030-67874-6_2

research on machine learning technology, support vector machine, a method developed from hinge loss function based on the principle of minimizing structural risk, has been widely used in pattern recognition, text classification, data mining and regression analysis [3].

Due to the late starting time of intelligent building engineering, historical data of project design, construction and cost is relatively less. Also, the construction cost of intelligent building project varies with the difference of construction content, region and type, which brings out the characters like, small number of data sample, multiple impact factors and nonlinear regression of data in intelligent building system. Nevertheless, support vector machine algorithm works well in solving this type of problems.

2 Materials and Methods

2.1 Support Vector Regression Algorithm

The input sample x is firstly mapped to the high-dimensional feature space through nonlinear function $\phi(x)$, and a linear model is established in this feature space to calculate the regression function, as shown in the following formula, where w is weight vector and b is threshold [4].

$$f(x, w) = w \times \phi(x) + b \tag{1}$$

For a given training data set $(y_1, x_1), (y_2, x_2), \cdots (y_i, x_i)$, insensitivity loss function ε is adopted, and the corresponding support vector machine is called ε- support vector machine, then its constrained optimization problem can be expressed by the following formula [5].

$$\begin{aligned} &\min_{w} \frac{1}{2} \times w^2 \; + \; \mathrm{C} \sum\nolimits_{l}^{\mathrm{i}} (\xi_l + \xi_l^*),\ l = 1, 2, \cdots n \\ &\text{s.t.} \\ &y_l - w \cdot \phi(x) - b \le \varepsilon + \xi_l^* \\ &w \cdot \phi(x) + b - y_l \le \varepsilon + \xi_l \\ &\xi_l^*, \xi_l \ge 0 \end{aligned} \tag{2}$$

Where, C is the penalty coefficient, and ξ_l, ξ_l^* is the relaxation variable, the optimization problem in formula (2) is converted to dual problem to solve formula (1) after Lagrange function imported:

$$f(x) = \sum_{\mathrm{i}=1}^{n_{sv}} (a_i - a_i^*) K(x_i, x) + b \tag{3}$$

Lagrange multipliers are $a_i, a_i^*(i = 1, 2, \cdots l)$, a small part of them are not 0, which correspond to support vectors in sample, n_{sv} is the number of support vector, $K(x_l, x)$ is Radial Gaussian kernel function, λ is kernel parameter.

$$K(x_l, x) = \exp(-\lambda \times x_l - x^2) \tag{4}$$

Kernel parameter λ is used to control the degree of sample division; Penalty coefficient C is used to get command of empirical risk and confidence range of SVR model; Insensitive loss function ε controls the width of insensitive area which is used in regression function acting on sample data [6, 7].

2.2 Application Steps of Support Vector Regression (*SVR*)

The prediction model of building intelligent engineering investment can be regarded as a nonlinear regression function problem: in the i^{st} year, the index value of project cost influencing factor is the independent variable $X_i = \{x_{i1}, x_{i2}, \cdots x_{in}\}$; in the j^{st} year, the prediction value of project cost is assumed to be y_j. The input is X_i, the output is Y_j; the relation between X_i and Y_j is assumed to be a nonlinear function mapping $F(x)$, which makes $y_i = F(x_{i1}, x_{i2} \cdots x_{in})$, where input is $X_i = \{x_{i1}, x_{i2}, \cdots x_{in}\}$ and output is y_j. Then, SVR is used to solve the regression equation $f(x) = \sum_{l=1}^{n_{sv}} (a_l - a_l^*)K(x_l, x) + b$, to obtain the predicted value of project investment 错误!未找到引用源。. The application ideas are as follows:

(1) Determine the engineering cost impact factor, obtain the original cost data, extract the index value of the engineering cost impact factor, and preprocess the data with Maximum and minimum normalization method [8].
(2) Choosing a proper type of *SVR* function, reasonable kernel function and kernel parameter, the optimal parameter (C, λ) is obtained.
(3) The optimal parameter (C, λ) is substituted into the prediction model, and the training set is extracted from the sample data. The kernel function $K(x_l, x) = \exp(-\lambda \times x_l - x^2)$ and kernel parameter λ are substituted into formula (3) to train the sample data training set [9].
(4) Use the trained prediction model to verify its accuracy through the test set.

3 Construction of Expert System in Investment Prediction Model

The investment prediction expert system integrating expert knowledge, engineering data and prediction algorithm (as shown in Fig. 1) can be roughly composed of human-computer interaction interface, knowledge acquisition module, database, knowledge base, inference machine and explanation organization [10].

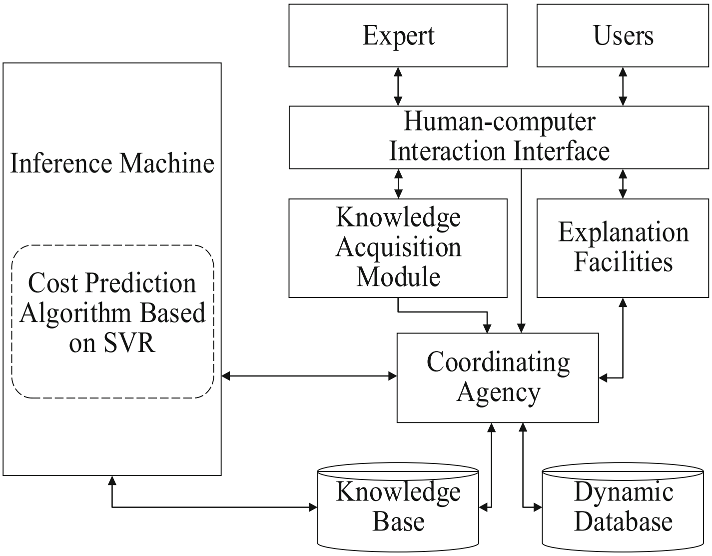

Fig. 1. Expert system architecture

In the process of system operation, users realize knowledge acquisition, data logging and prediction requests through human-computer interaction interface. The interactive system enables the inference machine to answer the request. This system USES the prediction algorithm based on support vector machine regression (*SVR*) to replace the traditional inference machine, which has higher prediction efficiency, better generalization ability and accuracy [11, 12]. The knowledge base and database are called in the prediction process [13–15]; Prediction results are output to the user through the interpretation mechanism and interactive interface.

4 Application of Investment Prediction Model

4.1 The Determination of Project Cost Prediction Index

In order to select the prediction index of building intelligent engineering cost accurately, this paper consulted a lot of literature and summarized dozens of prediction indexes. By means of questionnaire survey, combined with the opinions of 30 experts in intelligent building industry and the experience in design and construction, 5 factors that have the greatest influence on the cost are finally selected as prediction indicators: x_1 Construction Area, x_2 Building Type, x_3 Area of Structure, x_4 Construction Cost and x_5 Construction Demand. Both the quantization of qualitative index and magnitude difference between quantitative index needs data preprocessing.

The construction areas and different types of buildings are respectively quantified as natural numbers. The quotation index is determined by the construction time. According to the integration result of experts' opinions in the intelligent building industry and details of *Construction Project Pricing Quotation of Chongqing in 2018*, the labor cost and equipment cost index are calculated according to the weights of 0.15 and 0.85 respectively. The quantification of the prediction index of "construction demand" will affect the accuracy of the prediction result due to the variety of construction systems of building intelligence specialty. According to the *Intelligent Building Design Standard (GB50314-2015)*, for different types of buildings, there are three system construction schemes: "must be built", "should be built" and "can be built". The construction demand is simplified into: basic type (including items that must be built), expanded type (including items that must be built and should be built), and high-end type (including items that must be built, should be built and can be built) to facilitate data preprocessing.

4.2 Acquisition and Preprocessing of Sample Data

This paper selects 90 sets of cost data from the completed intelligent building projects in a certain area in recent years, covering different years, regions, building types, construction requirements and building area. According to the quantitative standard of prediction indicators (shown as Table 1), it is stored in the dynamic database as the original data of the investment prediction model in this paper (shown as Table 2). Data will be invoked while the prediction model is operating.

Table 1. Quantitative standard of engineering characteristics

Serial number	Predictive indicators	Value of feature vector	Quantized value
1	Construction area	Downtown	1
		Suburb	2
		Country	3
2	Building type	Residence	1
		Office	2
		School	3
		Hospital	4
		Business	5
		Hotel	6
		Factory	7
		Tourism	8
		Conference&Exhibition	9
3	Area of structure	The actual number	
4	Construction cost index	2016	0.9463
		2017	0.9781
		2018	1.0000
		2019	1.0204
5	Construction demand	Basic	1
		Expanded	2
		High-end	3

Table 2. Original sample data

Serial number	Construction area	Building type	Area of Structure (m^2)	Construction time (Year)	Construction demand (Type)	Cost of unilateral (RMB/m^2)
1	Downtown	Business	57000	2018	Basic	172
2	Downtown	Business	60000	2018	Expanded	224
3	Suburb	Business	89400	2017	High-end	292
4	Downtown	School	45000	2019	High-end	330
5	Downtown	School	80000	2018	Expanded	225
7	Suburb	Hotel	60109	2017	Expanded	353
8	Country	Hospital	64000	2016	Basic	273
⋮	⋮	⋮	⋮	⋮	⋮	⋮
86	Downtown	Office	100338	2019	Basic	221
87	Downtown	Office	97600	2018	Basic	215
88	Suburb	Factory	280000	2017	Basic	60
89	Suburb	School	69034	2018	High-end	287
90	Country	School	27800	2016	Basic	175

Pre-processing is required after the input data is quantized, in order to avoid the input data appearing in the saturated region of the program function, improve the influence of the data convergence speed and data magnitude difference in the program on the prediction result, and reduce the prediction error. Data normalization (normalization) is a common method for data preprocessing. The maximum and minimum value method can flexibly specify the value interval after normalization, eliminate the weight difference between different attributes, and normalize the sample data to [0, 1]. The calculation method is maximum and minimum normalization $X' = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}}$, where $i = 1, 2, \cdots n$. $X_{\max}, X_{\min}$ are the maximum and minimum values of the data in a prediction indicator.

4.3 Interface and Function Construction of Expert System

This paper uses software development environment to build an investment prediction platform based on the expert system.

(1) Achieve data acquisition through interface input or file import; Through interface input or automatic knowledge acquisition and update; Realize the visualization of database and knowledge base.
(2) Based on the good prediction results of the *SVR* prediction model, it is embedded into the expert system as a kind of inference machine and combined with the expert knowledge to make the project cost prediction.
(3) After clicking the extension button on the right, the interface displays important parameters of the model, comparison of test results and other information.

(4) New project cost information prediction is input from the interface, and finally the project cost estimation book is generated automatically with the import of expert knowledge.

4.4 Selection of SVR Parameters

(1) SVR function type selection: In support vector machine regression model, *nu*-SVR and ε-SVR are commonly used functions. In *nu*-SVR method, the number of support vectors have to be determined in advance, and the optimal model should be obtained by slowly adjusting the values of (optimization parameters). In ε-SVR method, it's necessary to determine the value of ε in advance, that is the value of loss function (limit error), which can determine the bandwidth of the fitting function, calculation error effectively, and the prediction model function has a good performance. As a result, this paper use ε-SVR.
(2) Selection of Kernel Functions: In SVR model, linear kernel function is not mapped to the high-dimensional space for linear regression problem; the complexity of model is increased by polynomial kernel function due to its parameter being large in number; The radial Gaussian kernel function performs well in solving nonlinear regression problems, owing to its rather small characteristic dimension, which makes it highly desirable when the number of samples is relatively small; Sigmoid kernel function generates neural network, its generalization ability is relatively weak, and some parameters are invalid. Eventually, this paper use radial Gaussian kernel function (RBF).
(3) Parameter Finding and Optimization: According to the content of chapter 2.1, parameter (C, λ) plays a job of vital importance in performance of prediction model. In this model, we use usual k- fold cross validation method to calculate each set of (C, λ) within a specified range to obtain the best solution. There are 3 kinds of optimization methods of (C, λ) which are frequently-used. Firstly, Genetic neural network algorithm (GA) is complex and weak in generalization. Secondly, Particle Swarm Optimization algorithm (PSO) is a heuristic algorithm. It can find the global optimal solution without traversing all points in the grid, but it is easy to fall into the local optimal solution. Last but not the least, Grid-search algorithm is to traverse all possibilities (C, λ) through the range of attempts and add a two-layer loop before the cross-validation program; the complexity of it is not high, but the accuracy is improved obviously. Therefore, grid search algorithm is used to optimize the (C, λ) parameters.

4.5 Model Training and Testing

For the determination of limit error parameters, 80 groups of data are used as training samples to establish a prediction model, then the value of ε with the best prediction performance will be selected. In this process, coefficient of determination is $R^2 = 1 - \frac{SSE}{SST}$, SSE is the sum of squares of the difference between the fitting value and the mean value of the original data, the sum of squares of the original data and its mean. R^2 is to determine the degree of fitting optimization of the model, the closer to 1, the more the model fits the training data; When the value of R^2 is far from 1, the model fitting degree is poor (Table 3).

Table 3. The model information corresponding to the change value of ε

Limit error ε	Number of training samples	Number of support vectors	Mean square error *MSE*	Coefficient of determination R^2
0.01	30	17	0.01402	0.96514
0.05	30	20	0.00817	0.97833
0.1	30	25	0.00685	0.93322
0.2	30	26	0.01887	0.94745
0.5	30	25	0.00463	0.98680

As shown in table, the model has high fitting degree and the best performance when the value of $\varepsilon = 0.5$. After 80 sets of data imported into the expert system as training samples, the optimal value of (C, λ) are obtained through Grid-search algorithm and cross validation, where $C - 16$, $\lambda = 0.3125$ and $n_{SV} = 25$ (n_{SV} is the number of support vectors); the remaining 10 groups of data are tested in the model, and the comparison and deviation between the predicted results of intelligent building engineering investment and the actual values are shown in Fig. 2.

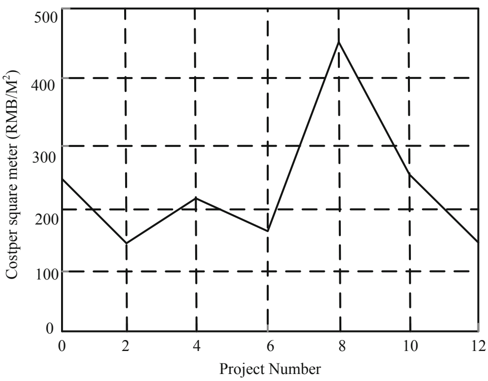

Fig. 2. *SVR* prediction model result comparison chart

4.6 Analysis of Prediction Results

The prediction model of intelligent building investment established in this paper is mainly aimed at the cost prediction in the design stage of intelligent engineering projects. Therefore, when the relative error between the predicted value and the actual value of the prediction model is less than or equal to 10%, the calculation of the model can be viewed as accurate and reliable one.

According to the test results, the regression model based on the expert system has a prediction error of less than 10% for the cost of intelligent building engineering, high accuracy and relatively strong generalization performance. For individual samples, such as item No. 1, the reason for the large error lies in the insufficient similarity between the training sample and original sample. With the further acquisition of data and knowledge by the expert system, the prediction error will be further reduced. This model has a good applicability for intelligent building engineering where the prediction index is in the middle dimension and the influence relationship between each index is rather complex.

5 Conclusion

The effective combination of expert system and support vector machine regression model can make up for the shortcomings of traditional estimation methods, such as weak data, weak knowledge acquisition ability, insufficient generalization ability and low prediction accuracy. The parameter setting of the model plays a key role in the final prediction results. Through reasonable selection, selection and optimization of kernel function, limit error and penalty factor, the fitting degree and prediction accuracy of the model are improved.

In view of the current situation that the cost standard of intelligent building specialty is not perfect, the cost influencing factors are complex, and the cost historical data are few, the regression investment prediction model based on expert system can reasonably predict the project cost in the design stage, and obtain the construction cost of new construction and reconstruction project. It can provide assistance and reference for the owners, investors and design units, and the prediction results can be used as the basis for project decision-making. At the same time, the cost prediction of this paper enriches the project management method of intelligent building engineering, saves manpower and material resources, and brings certain economic benefits for the project.

References

1. Zhou, Q., Ding, S., Feng, Y., et al.: The elevator brake intelligent monitoring and fault early warning system based on SVM. China Special Equip. Safety **34**(5), 22–27 (2018)
2. Li, Q.: Discussion on the integration method of weak building system in intelligent buildings. Modern Sci. Instrum. **12**(5), 80–84 (2018)
3. Yang, T., Yue, C.: Research on hydrometeor classification of convective weather based on SVM by dual linear polarization radar. Torrential Rain Disas. **38**(4), 297–302 (2019)
4. Kong, J., Cao, X., Xiao, F.: Research on cost prediction model of power transmission and transformation engineering based on support vector machine. Modern Electron. Technol. **41** (04), 27–130 (2018)
5. Goren, L.: Ten strategies for building emotional intelligence and preventing burnout. Family Pract. Manage. **25**(1), 11–14 (2018)
6. Lee, J.-Y.J., Miller, J.A., Basu, S., et al.: Building predictive in vitro pulmonary toxicity assays using high-throughput imaging and artificial intelligence. Arch. Toxicol. **92**(6), 2055–2075 (2018)

7. Wang, N., Wang, F., Yin, Y., et al.: Research on cost prediction of substation engineering based on support vector machine. Constr. Econ. **37**(05), 48–52 (2016)
8. Jin, Z., Hu, J., Jin, H., et al.: Analysis of traditional Chinese medicine prescriptions based on support vector machine and analytic hierarchy process. China J. Chin. Materia Med. **43**(13), 25–32 (2018)
9. Rios, D., Haake, J.: Getting off the ground floor of the smart building revolution. Appl. Des. **67**(3), 28–29 (2019)
10. Goldschmidt, I.: Intelligent building integration. Eng. Syst. **35**(2), 23–26 (2018)
11. Zheng, P., Shuai, L., Arun, S., Khan, M.: Visual attention feature (VAF): a novel strategy for visual tracking based on cloud platform in intelligent surveillance systems. J. Parallel Distrib. Comput. **120**, 182–194 (2018)
12. Car, J., Sheikh, A., Wicks, P., et al.: Beyond the hype of big data and artificial intelligence: building foundations for knowledge and wisdom. BMC Med. **17**(1), 36–42 (2019)
13. Bhandari, M., Reddiboina, M., et al.: Building artificial intelligence-based personalized predictive models. BJU Int. **12**(12), 25–31 (2019)
14. Fu, W., Liu, S., Srivastava, G.: Optimization of big data scheduling in social networks. Entropy **21**(9), 902 (2019)
15. Liu, S., Bai, W., Zeng, N., et al.: A fast fractal based compression for MRI images. IEEE Access **7**, 62412–62420 (2019)