Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
Nov-Dec 2021 Page No. 315-323
Article History: Received: 05 October 2021, Revised: 02 November 2021, Accepted: 25 November 2021, Publication: 31 December 2021

Optimal Adjustment of Investment Estimation Index for Single Project of Barracks

Wang Dehua, Hu Yuanxin*, Li Shengbo
Army Logistics Academy of PLA, Chongqing, China
*Corresponding Author.

Abstract:

The currently used investment estimation index for single project of barracks can no longer meet the application need under market changes. The problems present in the traditional estimation index compilation method are analyzed, and the estimation index optimization adjustment method is proposed based on big data technology. By analyzing the existing problems in single project investment estimation index for barracks, the investment estimation index compilation method is proposed based on the data mining C4.5 algorithm. At the same time, the influence of time factor on each index is comprehensively considered, and sensitivity analysis is used to determine the most sensitive factor of the index, so that the adjustment coefficient of the investment estimation index is calculated. Finally, the feasibility of the method is verified through analysis of examples, which has certain guiding significance for the optimization and adjustment of the investment estimation index of barracks project in the future.

***Keywords**: Big data, Investment estimation index, C4.5 algorithm.*

I. INTRODUCTION

"Investment Estimation Index for Single Project of Barracks" (GJB-2132B) provides an effective and feasible basis for the investment estimation of barracks and the preparation of project proposals, strengthens the macro-control of construction investment in barracks projects, greatly improves work efficiency, so that it is possible to accelerate the construction of modern barracks, adapts to the reform and market changes in the construction industry. The standard mainly includes base estimate index, regional price ratio coefficient, labor and main material consumption index, etc. However, Ding Wenxuan [1], Zhao Juan [2], Cheng Peng [3] et al. compared the collected project examples with the standard, and found that the actual project's square meter cost deviates much from the standard, and further improvement is needed in the application process.

As for the main existing methods for adjustment of investment estimation index, in the preparation of green building investment estimation, Hu Qingguo et al. [4] combined the characteristics of machine learning and genetic algorithm, constructed a green building investment estimation model to improve the accuracy of investment estimation application. Rang Xingyan et al. [5] established a BP neural network algorithm model through python machine learning, predicted the related costs of unknown indexes through limited data samples, and completed the prediction of investment estimation based on the main project features. Yang Fan [6] analyzed the main characteristic factors that affect the project cost, improved the BP neural network algorithm based on gray theory, approached the optimal value of the weight by continuously adjusting the value interval of the interval gray number, thus further improving the applicability and reliability of the estimation index. Wang Demei [7] used K-means clustering to screen samples, used the original feature index and the comprehensive index after least square regression to establish an SVN model and predict the project cost per square meter.

However, the above research methods fail to fundamentally solve the problems of long compilation time and slow update of investment estimation index. In view of the above problems, this paper proceeds from the perspective of data mining, collects market and historical engineering data, performs certain preprocessing, and uses the machine learning C4.5 algorithm for independent learning of the processed data files to generate a learning machine model and investigate the impact of market environment changes on investment estimation. Taking into account the influence of time factors, we conduct sensitivity analysis to calculate investment estimation index adjustment coefficients, thereby preparing investment estimation index for barracks under time changes.

II. C4.5 ALGORITHM MODEL BASED ON DATA MINING

2.1 Construction of C4.5 Algorithm Model

2.1.1 Introduction to C4.5 algorithm

The C4.5 algorithm is improved by J. Ross Quinlan on the basis of the decision tree core ID3 algorithm. It is a classification decision tree algorithm in machine learning algorithms [8]. The algorithm model mainly deals with data classification problems. That is, it uses data classification to discover the laws between data. Its basic principle is to evaluate project risks by forming a decision tree based on the known probability of occurrence under different situations, thereby determining feasibility of the plan. The steps are as follows:

(1) Divide the data set into training set and test set.

(2) Training stage: Check whether each sample in the data set belongs to the same category. If so, then a leaf node is formed. Otherwise, find the best features of the data set to divide and create intermediate nodes. When all the smallest subsets fall to the same category, leaf nodes are formed, and the decision tree is established. In the national military standard investment estimation index, the basic type, main structure, building height, etc. can be used as attribute values.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
Nov-Dec 2021 Page No. 315-323
Article History: Received: 05 October 2021, Revised: 02 November 2021, Accepted: 25 November 2021, Publication: 31 December 2021

(3) Test stage: Use the test data set to verify the established decision tree and evaluate the model accuracy.

2.1.2 Calculate the classification rule parameters to construct the classification tree

In the process of determining the split attributes and establishing the decision tree, the key is to clarify the "best feature of the data set", which is the basis for layer-by-layer division when the decision tree is formed.

Suppose the current sample information set D contains n types of samples. Where, the proportion of the k-th type of samples is pk, (k=1, 2, 3,..., n), then the information entropy of D is:

$$Ent(D) = -\sum_{k=1}^{n} p_k \log_2 p_k \tag{1}$$

Information entropy Ent (D) represents the information messiness degree. The greater the value, the greater the uncertainty.

When the C4.5 decision tree is forked, the attribute selection is carried out by the parameter "information gain rate", and the information gain rate is standardized through the "classification information value". For attribute A, the information gain rate is calculated by the following formula:

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo(A)} \tag{2}$$

Where, Gain (A) represents the information gain amount, namely

$$Gain(A) = Info(D) - Info\_A(D) \tag{3}$$

SplitInfo (A) represents the classification information value, and its formula is as follows:

$$SplitInfo(A) = -\sum_{j=1}^{n} \frac{|D_j|}{|D|} \times \log_2\left(\frac{|D_j|}{|D|}\right) \tag{4}$$

Where, D represents the number of samples in the data set; n represents the number of attribute values of attribute A in the data set; j represents the attribute value label of attribute A in the data set; D_j represents the number of samples whose value of attribute A in the attribute set is equal to the corresponding value of number j.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
Nov-Dec 2021 Page No. 315-323
Article History: Received: 05 October 2021, Revised: 02 November 2021, Accepted: 25 November 2021, Publication: 31 December 2021

Calculate the information gain rate of each attribute separately, and select the attribute with the larger information gain rate as the branch attribute.

2.1.3 Pruning optimization of classification tree

Since the C4.5 algorithm calculates the information gain rate of each attribute, the decision tree information obtained in this way is quite detailed and huge. At the same time, since the training samples covered by each node are absolutely classified, when the training samples are used in the decision tree classification, the error value is very low, but the wrong data in the training sample will also be learned by the decision tree and used as a component of the decision tree, leading to over-fitting and resulting in poor classification effect of new data. Therefore, while using the algorithm to construct the decision tree, pruning optimization should be carried out until the creation of molecules is stopped when branching is impossible. In order to prevent overfitting, the corresponding threshold should be set. When the information entropy is less than this threshold, the branching should be stopped, using it as a leaf node.

2.2 Estimation Index Adjustment

For the algorithm model derived from the completed project data, when calculating the latest investment estimation index, due to time changes, labor cost, material cost, and machinery cost will inevitably change. Therefore, it is necessary to adjust the investment estimation index under the influence of time. Sensitivity analysis can be made to determine the most sensitive factor of the investment estimation index, calculate the change of the sensitive factor relative to the estimation index, and use the ratio before and after the change as the adjustment coefficient of the investment estimation index, so that the adjusted investment estimation index has certain timeliness to cope with the impact of market environment changes on the estimation index.

Sensitivity analysis is to analyze the impact on economic index when the project's uncertain factors change, find out the project's sensitive factors, and determine the degree of sensitivity [9]. The main steps of sensitivity analysis are as follows:

(1) Select the uncertain factors to be analyzed. The engineering projects mainly incur: labor cost, material cost, and machinery usage cost.

(2) Determine the evaluation index of sensitivity analysis. By taking the construction and installation engineering cost as the evaluation index, we can intuitively see the changes in the labor, materials, and machinery usage cost due to the influence of time.

(3) Calculate changes in evaluation indexes due to changes in uncertain factors. Set the range of changes in labor, materials, and machinery usage cost, calculate the corresponding changes in the evaluation index for each amplitude change, form the corresponding quantitative relationship and express

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
Nov-Dec 2021 Page No. 315-323
Article History: Received: 05 October 2021, Revised: 02 November 2021, Accepted: 25 November 2021, Publication: 31 December 2021

it in a chart.

(4) Sort the sensitive factors according to the calculated sensitivity coefficient, and the ratio of construction and installation cost before and after the change is the adjustment coefficient of the investment estimation index. The calculation formula of the sensitivity coefficient is as follows:

$$\beta = \frac{\Delta A}{\Delta F} \tag{5}$$

Where, β is the sensitivity coefficient of the evaluation index A regarding the uncertainty factor F;

△ A is the change rate of the evaluation index A when the uncertainty factor F has the change rate of △ F;

△ F is the rate of change of the uncertainty factor F.

III. EXAMPLE ANALYSIS

3.1 Collect Data

This paper takes part of the residential buildings in Beijing in 2019-2020 as examples, collects the internal project instance data of the enterprises, mainly focuses on the influence of project construction site, construction time, structural characteristics, building floors, foundation form, roofing, interior decoration, doors and windows, exterior decoration, water supply and drainage, fire protection, heating, electrical engineering and other attributes on the construction cost per square meter.

3.2 Build a Classification Tree

The sample attribute set is {construction site, construction time, structural characteristics, building floors, foundation form, roofing, interior decoration, doors and windows, exterior decoration, water supply and drainage, fire protection, heating, electrical engineering), and the attribute set is determined by referring to "Investment Estimation Index for Single Project of Barracks" (GJB-2132B). The construction cost per square meter is classified. The average value is about 2275, the minimum value is 1768, and the maximum value is 2768. Taking the average value as the boundary, the category sets are {1701-2000, 2001-2300, 2301 -2600, 2601-2900}. Calculate the information gain rate of each attribute, select the attribute with the highest gain rate as the node, implement the C4.5 algorithm through Matlab, and build a classification decision tree as shown in Fig 1.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
Nov-Dec 2021 Page No. 315-323
Article History: Received: 05 October 2021, Revised: 02 November 2021, Accepted: 25 November 2021, Publication: 31 December 2021

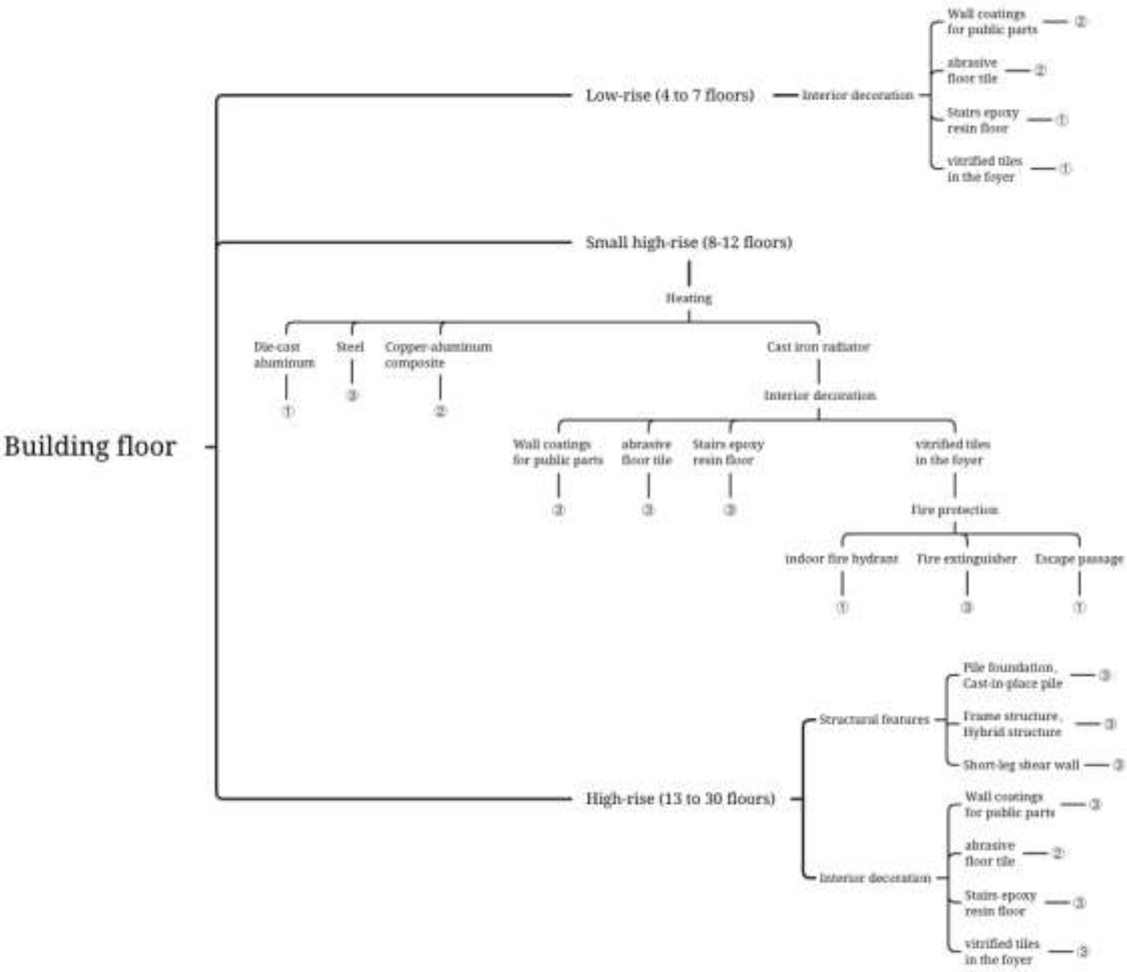

Fig 1: Classification decision tree

3.3 Verify Classification Rules

The constructed classification decision tree is verified through the test data set, and the attributes in the test set are substituted into the classification decision tree diagram of Figure 1 to calculate the final construction cost per square meter. Then, the construction cost per square meter derived from the model is compared with the actual cost, and the correct rate of the decision tree when tested on the test set is calculated. This paper collects 93 sets of data and conducts 9 experiments. The training set selects 1 to 9, 11 to 19,..., 81 to 89 data, and the remaining data is used as the test data set. The constructed classification tree has an accuracy of 97.78% according to the verification result, and it can be considered that the decision tree has good usability.

3.4 Adjustment Coefficient of Investment Estimation Index

To fully reflect the impact of time changes on labor, materials, and machinery usage cost, which

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
Nov-Dec 2021 Page No. 315-323
Article History: Received: 05 October 2021, Revised: 02 November 2021, Accepted: 25 November 2021, Publication: 31 December 2021

thereby leads to changes in construction and installation cost, construction and installation cost is used as the evaluation index, and the single factor sensitivity analysis is performed to calculate the maximum sensitivity index and measure the changes in construction and installation cost. The ratio before and after the change is used as the adjustment coefficient of the investment estimation index.

(1) Determine the sensitivity factor

Analyze labor, materials, and machinery usage cost as uncertain factors. Based on the analysis of project cost composition, the ratio of construction engineering labor cost, material cost, and machinery usage cost is roughly 2:7:1. The related cost for labor, materials, and machinery are set to 200,000, 700,000, and 100,000 yuan for subsequent sensitivity analysis.

(2) Determine the analysis index.

The construction and installation engineering cost is used as the analysis and evaluation index. The calculation is shown in Table I below (the correlation coefficient is compiled according to the pricing basis of Beijing construction projects):

TABLE I. Composition of construction and installation cost

Calculation method / Composition	Expense item	Calculation formula
Direct cost	Direct engineering cost	Labor cost + material cost + machinery usage cost
	extra construction cost	(Labor cost + material cost + machinery usage cost)×3.40%
Indirect cost	Stipulated cost	Labor cost×19.76%
	Management cost	(Labor cost + material cost + machinery usage cost)×9.62%
Profit		(direct cost + management cost)×7.00%
Tax		（direct cost + management cost +profit）×11.00%

It can be seen from Table 1 that the investment estimation index = construction and installation engineering cost = (labor cost + material cost + machinery usage cost) × 1.34 + labor cost × 0.22.

(3) Calculate the change result of the evaluation index caused by the different amplitude changes of the uncertainty factors.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
Nov-Dec 2021 Page No. 315-323
Article History: Received: 05 October 2021, Revised: 02 November 2021, Accepted: 25 November 2021, Publication: 31 December 2021

Due to time changes, labor cost, material cost, and machinery usage cost will all change to a certain extent. Suppose the amount of rated consumption remains unchanged, the fluctuations within 5% amplitude in labor, material, and machinery within a range of ±20% will affect the construction and installation cost. The impact is shown in Table II.

TABLE II. Sensitivity analysis table

Rate of change / Influencing factor	-20%	-15%	-10%	-5%	0%	5%	10%	15%	20%
Labor cost	132.16	133.72	135.28	136.84	138.4	139.96	141.52	143.08	144.64
Material cost	119.64	124.33	129.02	133.71	138.4	143.09	147.78	152.47	157.16
Machinery cost	135.72	136.39	137.06	137.73	138.4	139.07	139.74	140.41	141.08

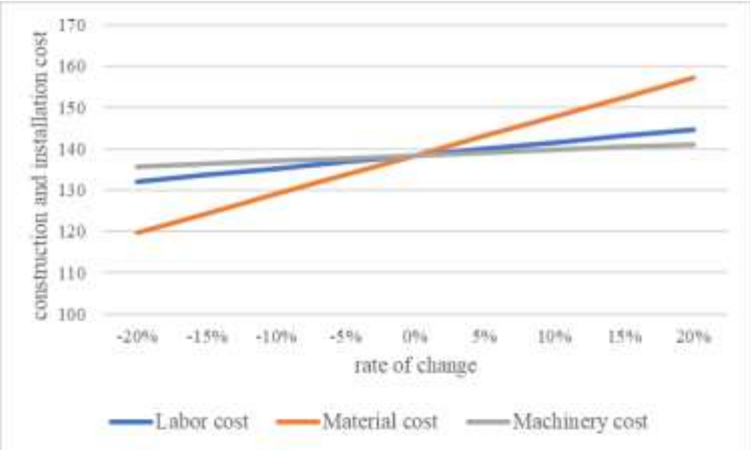

Fig 2: Sensitivity analysis diagram

(4) Analyze sensitivity factors and calculate the adjustment coefficient of investment estimation index.

It can be seen from the sensitivity analysis in Figure 2 that material cost is the most sensitive factor when considering the uncertainty factor change in the same amplitude due to the time factor. Changes in material cost in different periods can be used to adjust investment estimation index. For instance, in the calculation example, the construction cost per square meter of a construction project that meets certain characteristic attributes is 1,800 yuan/m^2. This index is based on the project example information in 2019 and 2020. When using this index to compile similar projects in other years, the proportion of changes in construction and installation engineering cost caused by material price changes is used as the adjustment coefficient of the estimation index.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
Nov-Dec 2021 Page No. 315-323
Article History: Received: 05 October 2021, Revised: 02 November 2021, Accepted: 25 November 2021, Publication: 31 December 2021

IV. CONCLUSION

(1) The compilation of the investment estimation index for single project of barracks makes the investment estimation control of the barracks project well-documented and greatly improves work efficiency. However, in application, the impact of time factors on investment estimation should also be fully considered. The currently used investment estimation index for single project of barracks is far from being applicable to the current project cost management. A data mining C4.5 algorithm is proposed to classify completed projects, a classification tree is constructed based on project example data, and classification rules are used for adjustment of investment estimation index to truly reflect the actual market conditions.

(2) Based on comprehensive consideration of the impact of time factors on investment estimation index, material cost is determined as the most sensitive factor through sensitivity analysis. According to the possible changes in future material prices, the construction and installation engineering cost under the influence of time is calculated, and the ratio before and after the change is used as the adjustment coefficient of investment estimation index. The investment estimation index determined by this method has certain timeliness and can effectively respond to market changes.

(3) This paper studies investment estimation index adjustment of the national military standard single projects based on big data. However, in the process of data collection, there are currently few open source data, and the amount of data collected for the corresponding military installation projects is also insufficient. In actual use, there may be certain limitations, which is also the direction demanding further investigation and improvement in the next step.

REFERENCES

[1] Ding WX, Zhao J, Wei DG, Li KX, Liu JX (2016) Investment estimation and problem analysis of National Military Standard single projects. China Storage & Transport 2016(11):119-121
[2] Zhao J (2017) Problems and countermeasures in the practical application of "Investment Estimation Index for Single Project of Barracks". Journal of Logistics University of PAPF 2017(04):72-73
[3] Cheng P, Wang YZ, Zhao JS (2018) Reasonable compilation of investment estimation index. Naval Engineering Technology 2018(01):71-72
[4] Hu QG, Tian XZ, He ZM (2020) Green building investment estimation method based on genetic algorithm to optimize extreme learning machine. Construction Economics 41(10):125-130
[5] Rang XY, Li HL, Zhao QX, Peng HG (2021) Research on Investment Estimation of Residential Projects Based on BP Neural Network. Engineering Cost Management 2021(01):18-24
[6] Yang F (2016) Research on residential project cost estimation model based on improved BP neural network. China University of Mining and Technology
[7] Wang DM, Chen H, Xiao ZH, Xia SL, Fan SQ, Cui CH, Zhang QH (2021) Residential project cost prediction based on data mining. Journal of Civil Engineering and Management 38(01): 75-182
[8] Yu L, Yang YN (2010) Research and application of C4.5 algorithm. Journal of Yunnan University (Natural Science Edition) 32(S2):136-140
[9] Liu XJ (2014) Engineering Economics. Beijing: China Construction Industry Press 2014:92-93