Intelligent Computing Technology and Automation
Z. Hou (Ed.)

doi:10.3233/ATDE231254

A Cost Prediction Model for Construction Projects Based on Bayesian Networks

Ping XU[a,1], Chengming LIU[b]
[a] *Chongqing College Of Architecture And Technology, Chongqing, 401331, China*
[b] *Chongqing Jundao Yucheng Green Building Technology Co., LTD, Chongqing, 401520, China*

Abstract. Due to the complex calculation process of traditional methods in the entire process cost prediction of construction projects, it doe not consider the complex interaction relationship between multiple influencing factors during the cost formation process, so a construction project cost prediction model based on Bayesian networks is proposed. This method integrates project experience, historical information, and actual cost performance data sequences observed during project implementation, to determine the degree of influence and interrelationship of each influencing factor on the cost components. Then, an explanatory structural model and Bayesian network are used to establish a cost prediction model for construction projects, and the entire process cost of the construction project are calculated and output through analysis functions. Finally, the application effect of the model in an example is explored, and the results show that the various costs and total costs acquired by Bayesian network conditional probabilities are basically consistent with the budgeted costs of original plan. The model not only has feasibility and high accuracy, but also has a positive impact on the probability and degree of low project cost risk events.

Keywords. construction projects; cost; Bayesian networks; prediction model; index system

1. Introduction

For construction enterprises, cost management is one of the most important tasks, and cost prediction is one of the primary tasks of cost management. By accurately predicting the cost of construction projects, enterprises can better plan and control project budgets, ensuring the economic benefits and profitability of the project. During the project progress process, predicting the project completion cost needs to be based on the estimated project plan cost during the project planning phase, and comprehensively consider the dynamic impact of actual cost performance during the project implementation process on the project completion cost. The probability prediction of project completion cost during project implementation is essentially a dynamic prediction process. Currently, cost prediction of construction projects is a complex task that is influenced by multiple factors. Therefore, the comprehensive application of multiple methods and technologies, combined with professional knowledge and experience, can obtain more accurate and reliable cost prediction results [1]. Cost forecasting provides a foundation for enterprises to control and manage project costs. By comparing with actual costs, enterprises can promptly identify and solve cost overruns or savings, ensuring the economic benefits of the

[1] Corresponding Author: Ping XU; Chongqing College Of Architecture And Technology, Chongqing, 401331, China; kassr8828@163.com

project.

At present, there are various algorithms and methods available for building cost prediction, including regression analysis, neural networks, genetic algorithms, support vector machines, and so on. It should be noted that different algorithms and methods are applicable to different construction cost prediction problems and data features. When selecting and applying algorithms, it is necessary to consider factors such as data availability, model interpretability, computational complexity, and conduct a comprehensive evaluation based on professional knowledge and experience. In recent years, research has also shown that Bayesian networks can be effectively applied to construction cost prediction, providing accurate, reliable, and interpretive cost prediction results through advantages such as modeling uncertainty, capturing dependencies between variables, integrating domain knowledge, and handling data incompleteness [2].

This article first defines the concept of construction project cost, and uses a combination of quantitative and qualitative methods to construct an indicator system for the influencing factors of construction project cost. Then, the structural equation is combined with Bayesian networks to analyze and study the factors that affect the cost of construction projects, and determine the degree of influence and interrelationships of each influencing factor on the cost components. Predict the weight of construction project costs through causal reasoning in the model, and calculate the final key impact indicators through statistics such as posterior distribution expectations, median, and maximum posterior probability. Case analysis shows that the cost prediction method for construction projects based on Bayesian networks has a clear idea, and the results reflected are in good agreement with the actual situation, with high reliability,; The algorithm can combine its advantages in handling dynamic and uncertain information with the characteristics of implicit quality costs in construction projects, and can dynamically predict specific costs during project implementation in real-time, thereby obtaining more accurate prediction results through past project experience and related project history.

2. Principle Idea of Bayesian Reasoning

The basic idea of Bayesian inference is to continuously update probabilities and combine prior knowledge with observational data to obtain more accurate and reliable posterior probabilities. Bayesian networks have a vivid and intuitive expression form, which can not only combine probability and statistics, but also have low requirements for existing information. They can make inferential predictions under incomplete and uncertain information, and have good comprehensibility and logicality; Bayesian networks can effectively express and fuse multi-source information. Bayesian networks can incorporate various information related to fault diagnosis and maintenance decisions into the network structure, and process them uniformly according to the nodes, effectively fusing the relevant relationships of information. It can also combine existing typical engineering technical solutions with computer experimental simulations, and incorporate the experience of existing engineering data as subjective guidance information during the prediction process, which has guiding significance for the prediction results. Bayesian inference obtains a posterior probability based on the results of two antecedents: a prior probability; A likelihood function derived from a statistical model of observed data. Bayesian inference calculates a posterior probability based on Bayesian formulas.

$$P(H \mid E) = \frac{P(E \mid H) \cdot P(H)}{P(E)} \tag{1}$$

H represents any hypothesis that its probability may be influenced by data (hereinafter referred to the evidence). Usually, these assumptions compete with each other, and our task is to decide which one is most likely $P(H)$, a prior probability, is an estimate of the probability of hypothesis H before data E (i.e. currently obtained evidence) is observed. E refers to new data that has not been used to calculate prior probabilities. $P(H \mid E)$, a posterior probability, refers to the probability that H gives E, that is, the probability of updating after observing evidence E. The posterior probability is what we want to obtain: what is the probability of a hypothesis occurring under the current observed evidence $P(E \mid H)$ is the probability of observing evidence E under the assumption of H, and is called the likelihood function. As a function of E under fixed H, it reflects the compatibility between the current evidence and the given hypothesis. The likelihood function is a function of evidence E, and the posterior probability is a function of hypothesis H. $P(E)$ referred to as marginal likelihood function or model evidence. This factor is the same for all possible assumptions considered (it can be clearly seen that there is no H in the symbolic expression), so it does not affect the relative probability between each hypothesis.

Due to the logical relationship between the static investment of a construction project and its various sub project costs, that is, the investment of each sub project cost is influenced by the scale of the engineering construction investment, and the investment fluctuations of each sub project cost have a certain fluctuation on the final cost level. Therefore, this article takes the investment cost of each sub project cost as the parent node, and the final static investment total of the construction project as the child node, The Bayesian network used for prediction is constructed after establishing a Bayesian network model, fitting the probability distribution of each sub item cost, and obtaining their respective prior probability tables based on statistical sample data; Then analyze the existing research methods for measurement uncertainty, select appropriate static investment calculation methods, estimate them, and obtain a prior probability table; Finally, based on the historical data of existing construction projects and corresponding typical projects, calculate the joint conditional probability, laying a solid foundation for the final static investment prediction [3,4].

3. Cost Prediction for Construction Projects Based on Bayesian Networks

3.1 Overall Idea

When applying Bayesian networks to predict construction project costs, it is necessary to comprehensively consider multiple factors and conduct reasonable model validation and adjustment. The use of Bayesian networks to predict the implicit quality cost of construction projects mainly includes four steps. Firstly, determine the cost variables of the construction project that need to be predicted, such as total cost, material cost, labor cost, etc. At the same time, determine the influencing factors related to these cost variables, such as building size, geographical location, material prices, etc. Secondly, collect historical project data and relevant factor data. These data can include cost data of completed construction projects, material price indices, labor cost data…, and undergo necessary data cleaning and preprocessing. Then, based on the dependency relationship between variables, construct the structure of a Bayesian network. This can be guided by expert knowledge and experience, or data analysis methods can be used to learn the relationships between variables [5-8]. Verify and adjust the constructed Bayesian network model using historical data and evaluate the accuracy and reliability of the model by comparing it with actual cost data. Finally, use the constructed Bayesian network model to predict the cost of construction projects. Calculate the predicted cost result through probability reasoning based on the given input factors. The specific prediction process of this process is shown in figure 1.

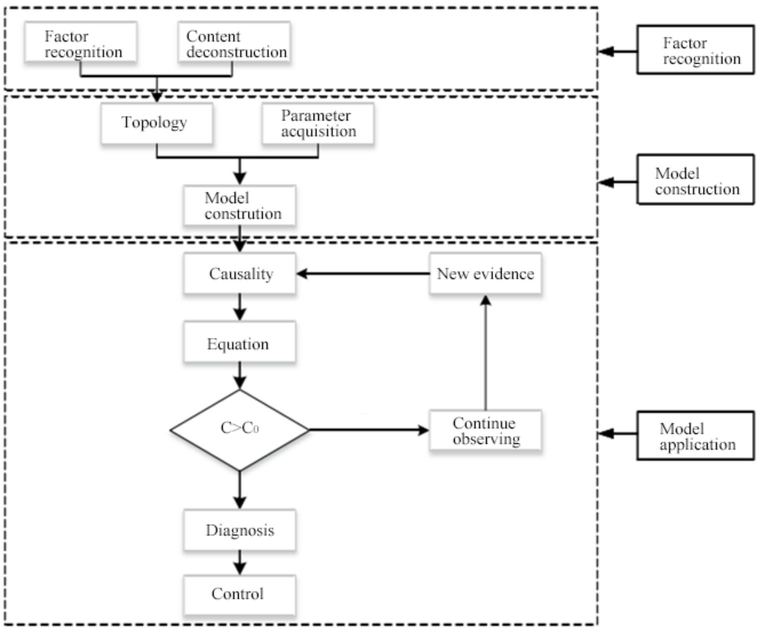

Figure 1. Cost prediction process for construction projects.

3.2 Index System Structure of Factors Influencing Construction Cost Projects

Firstly, we adopt research methods, expert methods, and interview methods to analyze and organize the risk factors that affect project costs, and the indicators of project cost risk influencing factors are shown in Table 1.

Table 1. Influencing factors of project cost prediction

1st level factors	2nd level factors
Production cost A1	Supply and demand relationshipA11, capital riskA12, material pricesA13
Transportation cost A2	Transportation distance A21, transportation route A22, transportation efficiency A23
Operation and maintenance cost A3	Labor costs A31, Energy costs A32
Management cost A4	Project management level A41, worker technical level A42

Based on the characteristics of the project, consider the project lifecycle cost as the sum of the costs of each stage, and set the level, degree, and probability calculation formula of the project cost influencing factors:

$$F = f(g, y) \tag{2}$$

where F is the level of cost risk events that occur in the project, g is the probability of cost risk events occurring in the project, and y is the degree of impact of cost risk events occurring in the project. When calculating specifically, only when the prediction model is effective can it be used for actual prediction. Import data that affects project cost risks into the model, use BIM software for calculations, and we can acquire corresponding level, probability, and impact degree data.

3.3 Construction and Computational Analysis of Bayesian Networks

Assuming the project is performed according to current cost performance index, the linear model slope $k(t)$ will be equal to the project completion cost C_{final}, that is

$$C_{final} = k(t) = BAC / CPI(t) \tag{3}$$

Since $k(t)$ may be modified by cost performance data, we can acquire the posterior probability of C_{final}. Then the prior probability of C_{final} can be modified when the performance data series D of actual cost are acquired

$$P(EAC \mid D) = p(k \mid D) = p(D \mid k)p(k) / p(D) \tag{4}$$

where $p(D)$ is the edge distribution probability of D; $p(D) = \int_0^{+\infty} p(D,k)dk = \int_0^{+\infty} p(D,k)p(k)dk$ is a constant; $p(D \mid k)$ is the probability of actual cost deviating from the planned value when the set cost is k

Finally, we input the predicted construction project data into the trained Bayesian network, and use the analysis function to calculate and analyze the entire process cost of the construction project. The calculation equation is depicted as follows:

$$C = \sum_{i=1} (X_i^P + S)w \tag{5}$$

where i denotes the node number of hidden layer of Bayesian network, that is, the i_{th} phase of construction work; S is the Correction coefficient for total project cost indicators; w is the index weight of total project cost .

4. Case Analysis

The experiment selected a high-rise commercial building project in a certain area as the experimental object, and a total of 6 cases are collected, with a building area of 1825.69 meters ^{2}- 1964.71 m^2. Use our design method and traditional methods to predict and analyze the cost of construction projects. Firstly, it is necessary to determine variables related to the cost of construction projects, including building area, building material cost, labor cost, equipment cost, geographical location, etc. Use the collected data to estimate the parameters in the Bayesian network model. Fit the probability distribution of each sub item cost. Then analyze the existing research methods for measurement uncertainty and select appropriate static cost calculation methods. The uncertainty factors to be analyzed are set to start changing from the values used in the certainty analysis, and the magnitude of each factor's change is the same. Calculate the probability of joint conditions for the final static investment prediction. By combining the results of Monte Carlo simulation, the probability distribution of construction project costs can be obtained. Statistical indicators such as average cost, standard deviation, and interval estimates of cost at different confidence levels can also be obtained. At a confidence level of 96%, the simulation result data can yield a static cost estimation model as shown in Figure 2:

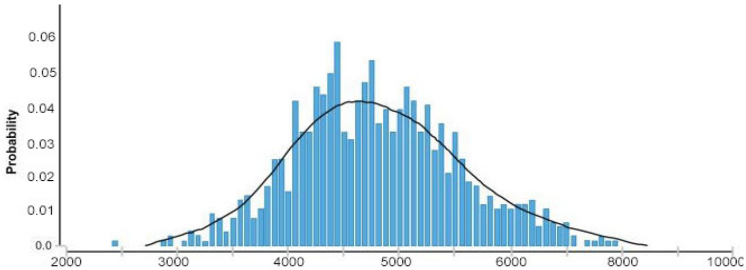

Figure 2. Probability distribution diagram of static cost prediction.

Figure 3 shows a model constructed using Bayesian network algorithm. Based on the cost management data collected through expert and research methods in the early stages of the project, it is imported. Building area is the root node, while building material costs, labor costs, and equipment costs are sub nodes starting from the root node. The arrows in a Bayesian network diagram indicate the direction of causal relationships. For example, building area may affect the cost of building materials, labor, and equipment, while other cost factors may also have mutually influencing relationships. This Bayesian network diagram can be used to establish a cost prediction model.

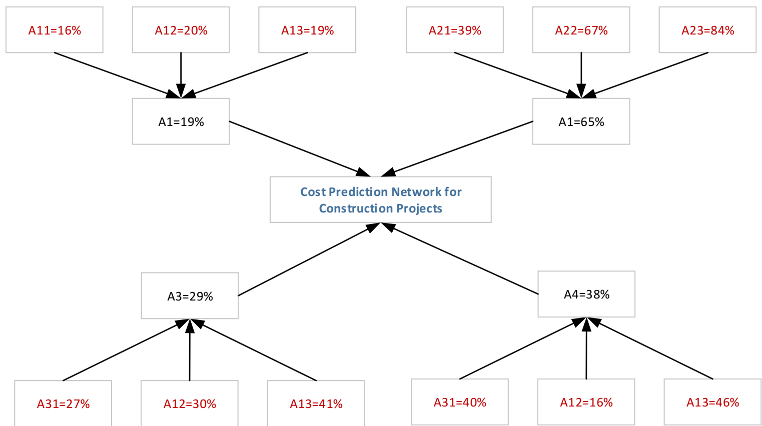

Figure 3. Bayesian cost prediction network.

By observing known variables (such as building area), the probability distribution of other variables (such as building material cost, labor cost, and equipment cost) can be inferred for cost prediction and analysis. From the results, the factors that are worth paying attention to are A23, A31, and A41, as their weight values are relatively large. Therefore, this network can clearly convey information that cost control personnel need to pay extra attention to.

The target error and initial learning speed of the input layer usually require experimentation and adjustment to determine the optimal values. You can try different parameter settings and evaluate and select the optimal parameter combination by monitoring the training process and performance indicators of the network. In the experiment, the target error of the Bayesian network is 0.0001, and the initial learning speed of the input layer is set to 0.01. Formula (5) is used to calculate the entire process cost of the construction project, and the built-in calculation function of YER is used to calculate the project cost. Based on the input data and cost classification. The comparison results of construction cost prediction are shown in Table 2, and the calculation results of deviation rate for the entire process cost prediction of construction engineering are shown in Table 3. It can be seen from this that traditional methods rely on experience and statistical rules, and they may not fully consider the complex relationships and uncertainties of various factors. Therefore, in some cases, the prediction results of traditional methods may have significant deviations; Bayesian networks can perform probability inference by inferring and predicting models to obtain the probability distribution and uncertainty range of costs, which can provide more comprehensive and accurate cost prediction results. Moreover, the cost prediction method for construction projects in this article has high prediction accuracy, and the prediction results are basically consistent with the actual situation.

Table 2. Cost prediction of construction projects

Project NO.	Actual cost(Million Yuan)	Predicted cost(Million Yuan)
1	1153.18	1152.36
2	1278.49	1258.51
3	1262.24	1300.17
4	1414.25	1407.38
5	1556.12	1556.67
6	1513.14	1499.09

Table 3. Deviation rate of cost prediction of two methods

Project NO.	Traditional method	Bayesian network
1	3.332	0.024
2	4.120	0.017
3	5.324	0.022
4	5.987	0.018
5	5.854	0.025
6	4.968	0.001

5. Conclusion

This article combines common construction cost calculation methods with Bayesian networks to analyze and study the influencing factors of construction project costs. Firstly, the relationship between various factors and project cost is analyzed, and an indicator system for construction project cost is also established. Then, a network model of the influencing factors of engineering cost is established using the interpretive structural modeling method and Bayesian network. Then the evaluation method and deviation standardization method are combined to standardize the prediction indicators of construction engineering cost, and the model is validated through engineering cases. The experimental results show that the deviation rate of the cost prediction method proposed in this article is smaller than that of traditional methods, and it has good feasibility and reliability, which has good application prospects in the field of construction project cost management.

References

[1] Luu, Van Truong, et al. Quantifying schedule risk in construction projects using Bayesian belief networks. International Journal of Project Management, 2009, 27(1):39-50.

[2] Kim, Sang Yon. Bayesian Model for Cost Estimation of Construction Projects. Journal of the Korea Institute of Building Construction. 2011, 11(1):91-99.

[3] Wang Xin, Geng Pengyun, Yin Bingbing Static investment prediction of power transmission and transformation projects based on Bayesian networks Management of Chinese Power Enterprises, 2015, 10:93-96.

[4] Gerassis S, Saavedra A, Garcia J F, et al. Risk analysis in tunnel construction with Bayesian networks using mutual information for safety policy decisions. WSEAS Transactions on Business and Economics, 2017(14):14.

[5] Jiang, Xueyu, et al. Engineered Injectable Cell-Laden Chitin/Chitosan Hydrogel with Adhesion and Biodegradability for Calvarial Defect Regeneration. ACS Applied Materials and Interfaces, 2023, 15(17):20761-20773.

[6] Piao, Yanmei , et al. Dynamic Fall Risk Assessment Framework for Construction Workers Based on Dynamic Bayesian Network and Computer Vision. Journal of construction engineering and management, 2021, 12:147-156.

[7] MATao. Brief Analysis of the Construction Project Cost Management and Control Analysis. Engineering technology, 2022, 4:80-83.

[8] Li D, Liu Z, Xiao P, et al. Intelligent rockburst prediction model with sample category balance using feedforward neural network and Bayesian optimization.Underground Space, 2022, 7(5):833-846.