International Journal of New Developments in Engineering and Society
ISSN 2522-3488 Vol. 8, Issue 5: 42-47, DOI: 10.25236/IJNDES.2024.080507

# Research on Optimization of Engineering Cost Database Based on Big Data and Intelligent Technology

**Yiru Zhang**

*Consumer Electronics Technology, Amazon, New York, NY 10044, USA*
*yiruz315@163.com*

***Abstract:*** *With the vigorous development of China's market economy and the continuous growth of people's livelihood demand, industries such as transportation, construction, and electricity are facing dual challenges of tight funding and surging construction demand. In this context, the rapid development of information technology, especially the rise of big data and cloud computing, has brought unprecedented opportunities for engineering cost control. However, the application of cost big data is still in its early stages, limited by data processing and analysis capabilities. This study is therefore committed to exploring new methods of engineering cost control in the big data environment. This study builds an intelligent, information-based, efficient and accurate control system, defines the control connotation and requirements, and designs the system framework and principles. This study proposes an innovative control method based on big data mining, and constructs a project cost case database and prediction model. At the same time, artificial intelligence algorithms such as neural networks and support vector machines are introduced to improve prediction and control accuracy. The technical roadmap of this study comprehensively covers the entire process from theoretical construction to method innovation, to model design and case verification, aiming to provide timely and accurate control strategies for engineering construction, optimize resource allocation, reduce capital waste, and provide scientific and practical solutions for engineering cost control in the big data environment.*

***Keywords:*** *Engineering cost control, intelligent control methods, data mining and artificial intelligence, cost interval prediction, and intelligent matching of control strategies in the big data environment*

## 1. Introduction

Against the backdrop of the booming development of China's market economy, the growing demand for people's livelihoods has driven a sharp expansion of construction investment in industries such as transportation, construction, and electricity. Faced with the dual challenges of tight funding and huge construction demands, how to effectively control project costs, optimize resource allocation, and reduce investment waste has become an urgent issue in the field of engineering construction. Especially in the construction of power grid infrastructure, the huge investment scale of State Grid and related industries during the 14th Five Year Plan period highlights the importance of scientifically controlling engineering costs. However, the "three excesses" problem in Chinese construction projects has exposed the shortcomings of traditional engineering cost control methods in setting investment target values and process control.

With the rapid development of information technology construction, the rise of modern technologies such as big data and cloud computing has brought unprecedented opportunities for engineering cost control. Government policy support has promoted the accumulation and sharing of cost data, laying the foundation for building a comprehensive management system for engineering costs. However, despite the increasing abundance of engineering cost data, its potential for application has not been fully realized due to insufficient data processing and analysis capabilities. Therefore, introducing advanced technologies such as data mining and artificial intelligence to build engineering cost control systems in the big data environment has become an important way to address the challenges of complex engineering cost control.This study aims to explore the theoretical mechanisms and control methods of engineering cost control in the big data environment from the perspective of investors (owners). By introducing big data mining theory and modern artificial intelligence technology, the problems of unscientific determination of reasonable cost values and inadequate process control in

ISSN 2522-3488 Vol. 8, Issue 5: 42-47, DOI: 10.25236/IJNDES.2024.080507

traditional engineering cost control can be solved. Propose key technologies such as engineering cost case representation, intelligent prediction of cost intervals, and intelligent matching of cost control strategies to achieve intelligent and effective control of the entire process of engineering costs. This study not only enriches the theory of engineering cost control and improves management paradigms, but also promotes the practical application of artificial intelligence technology in the field of construction cost control, which has important theoretical significance and practical value. In summary, this study will combine the extensive application value of cost big data, the current application status, and the urgent need for engineering cost control to explore in depth the methods of engineering cost control in the big data environment, providing strong support for the scientific development and practical application of the engineering construction field.

## 2. Correlation Theory

### *2.1 Theory Related to Engineering Cost Management*

Reasonably determining the engineering cost is a key link in effectively controlling the overall cost of engineering cost management, providing important decision-making basis for investors in uncertain environments. However, traditional methods such as coefficient estimation, quota pricing, and general cost estimation lack dynamic pre control capabilities and effective processing capabilities for uncertain information, making it difficult to meet the needs of lean control in the current market-oriented reform of engineering cost. With the advent of the big data era, the accumulation of structured and unstructured cost data, as well as the implementation of the "five price" filing system, have provided new opportunities for the processing of engineering cost information. At first, scholars attempted to predict costs through statistical analysis methods such as multiple regression analysis, fuzzy mathematics, and time series analysis. However, these methods have limitations, such as difficulty in fully considering the complex interactions between factors, high requirements for data quality, and complex training parameters, resulting in unsatisfactory prediction results.

In recent years, artificial intelligence technology based on cost big data has become an effective way to solve this problem. It can not only effectively reduce the uncertainty of information, but also quickly and accurately determine the reasonable value of engineering cost. Various artificial intelligence algorithms based on big data[1], especially the combination models of adaptive structural radial basis function neural network (RBF)[2], ensemble backpropagation neural network (BPNN) and genetic algorithm (GA) models, as well as the innovative combination of BPNN and support vector machine (SVM), have gradually become the mainstream methods for engineering cost prediction. In addition, emerging technologies such as complex network theory have also been introduced into the field of cost prediction, demonstrating higher accuracy through preliminary predictions based on node similarity. At the same time, the application of mixed algorithms such as principal component analysis (PCA) and nonlinear autoregressive neural network (NARX) also provides new ideas for municipal engineering cost prediction. The development and application of these technologies not only improve the efficiency and accuracy of engineering cost prediction, but also provide strong support for the scientific and intelligent management of engineering cost

### *2.2 Theory of Uncontrolled Engineering Cost Control*

One important reason for the loss of control in engineering cost is that the project investment decision-making stage did not fully anticipate the implementation risks and cost control difficulties of the project, and lacked targeted control strategies. This issue has received widespread attention both domestically and internationally. In the in-depth exploration of engineering project management by foreign scholars such as Ye, several research papers have jointly revealed the multidimensional path of cost control and optimization through the artificial bee colony comprehensive analysis model (ABC). Liming W et al. proposed an environmental economic dispatch method for power systems based on multi-objective artificial bee colony algorithm, providing a new solution for balancing economic and environmental benefits[3]. Meanwhile, Leung M Y et al. constructed a value risk management model for construction projects, aiming to enhance the scientificity and economic benefits of project decision-making by comprehensively considering project value and potential risks[4]. Almeida R et al.'s research focuses on the impact of progress measurement uncertainty in earned value analysis, providing theoretical support for improving the accuracy and reliability of project performance evaluation[5]. In addition, LIU Dongping discussed the pricing method of engineering quantity list and emphasized the core role of accurate pricing in cost control of engineering projects[6]. These studies have jointly promoted the development of engineering project management from different perspectives, providing valuable references for achieving more efficient and accurate cost control and optimization.

At the same time, Chinese scholars continue to explore and improve cost control strategies in feasibility studies, contract management, value engineering, and limit design, but overall, they still focus on post control. Taking the cost system of the power grid as an example, State Grid Corporation of China continuously improves its cost management level by implementing benchmark indicator systems among peers, strengthening design budget reviews, and annual cost analysis. In addition, since the mid-1990s, the life cycle cost management (LCA) [7]and comprehensive cost management (TCM) theories of engineering projects have gradually been introduced into China, becoming a hot topic in cost process control research, promoting the development of cost management towards a more comprehensive and refined direction.

## 3. Research Method

### *3.1 Big Data Intelligent Cost Control*

With the continuous accumulation of cost information, the "5V" characteristic of engineering cost big data is becoming increasingly prominent, and traditional cost control models are no longer able to meet the complex needs of the big data environment. This research system has constructed a cost control framework for construction projects in the big data environment, providing theoretical mechanism support and management guidance paradigms. This framework clarifies the connotation, requirements, and supporting theories of big data cost control. Through three functional modules: data synthesis, intelligent control, and dynamic feedback, it achieves intelligent determination of reasonable cost values and precise control throughout the entire process. We emphasize the principles of data-driven, systems science, precision computing, and feature oriented approaches, utilizing data mining and artificial intelligence technologies to reduce information uncertainty and enhance the scientific and intelligent nature of cost control. In terms of cost case representation, a representation model construction scheme based on data preprocessing technology was designed, and a clear, unified, and accurate cost case information database was constructed. At the same time, an intelligent combination prediction model that integrates case-based reasoning and statistical learning theory is adopted to improve prediction accuracy and flexibility. And introduce knowledge reasoning technology and expert system construction thinking, construct an intelligent matching model for cost control strategies, and achieve precise identification and real-time dynamic control of engineering construction risks. In summary, the intelligent control scheme proposed in this study integrates three modules: cost case representation, intelligent combination prediction of cost intervals, and intelligent matching of cost control strategies, forming a closed-loop control system. It not only solves many problems in traditional cost control, but also significantly improves the efficiency and effectiveness of engineering cost management, opening up a new path for engineering cost management in the big data environment.

### *3.2 Data Preprocessing Model*

As the core cornerstone of cost intelligent control, the construction process of engineering cost cases profoundly affects the efficiency and accuracy of intelligent control models. In response to the characteristics of the wide range, insufficient accuracy, complex and variable features, and dynamic randomness of the original cost data sources, this study carefully designed a case representation model construction scheme based on data preprocessing technology. This solution systematically solves problems such as data loss and singular values. Through data cleaning and integration, it achieves seamless integration and standardized processing of heterogeneous data from multiple sources, ensuring consistent time granularity and caliber of data, and enhancing data comparability and practicality. Subsequently, by utilizing data reduction and transformation techniques, the feature attributes were effectively streamlined, dimensional differences were eliminated, and the output format was unified, further enhancing the standardization and comprehensibility of the data. Of particular importance is the introduction of advanced data dimensionality reduction techniques such as Principal Component Analysis (PCA) and Exploratory Factor Analysis (EFA), which not only significantly reduce data dimensionality and improve model computational efficiency, but also deepen the understanding of common factors through methods such as factor rotation, making them more closely aligned with the actual needs of cost big data. This series of carefully designed preprocessing steps not only significantly improves the quality and usability of raw data, but also builds a clear structured, moderately granular, and easily understandable engineering cost case information database, providing

ISSN 2522-3488 Vol. 8, Issue 5: 42-47, DOI: 10.25236/IJNDES.2024.080507

unbreakable data support for subsequent intelligent prediction and control strategy optimization. This database not only serves as the fundamental data source for engineering cost control and plays an important role in the big data environment, but also becomes an indispensable knowledge base for intelligent prediction models and control strategy intelligent matching models, effectively promoting the intelligent and refined transformation and development of engineering cost management

### *3.3 CBR-MKRVM-KDE model*

The case-based reasoning model for engineering cost case retrieval and matching plays a core role in engineering cost interval prediction, using a "4R" loop to achieve fast retrieval. This article uses Grey Relational Analysis (GRA) combined with CRITIC weighting method to calculate case similarity and improve accuracy. The study also constructed a prediction model using a hybrid kernel correlation vector machine (MKRVM) and kernel density estimation (KDE). MKRVM combines the advantages of Gaussian and polynomial kernels to handle nonlinear problems, while KDE analyzes the distribution of prediction errors and outputs prediction intervals. The model solves the optimal parameters through maximum likelihood estimation, and KDE constructs an error probability density function to determine the cost error interval. The article constructs an evaluation system, including the analysis of point prediction and interval prediction effects, and comprehensively evaluates the accuracy and reliability of predictions through indicators such as RMSE, R2, PICP, NMPIW, etc., to support engineering cost control.

## 4. Results and Discussion

### *4.1 Building a Knowledge Base*

Building a comprehensive knowledge base for engineering costs is the foundation of intelligent matching cost control strategies, integrating case feature parameters for precise definition, continuously updating the control strategy library, and establishing rigorous matching logic. In this study, data mining and expert analysis are used to optimize the feature parameters, reduce the knowledge redundancy, and improve the computational efficiency and matching accuracy. In the face of project diversity and data differences, a two-step clustering algorithm is used to mine data commonalities, providing support for the construction of the strategy library. By combining Fuzzy Petri Net and Normal Cloud Model, an intelligent matching algorithm is designed to achieve precise mapping and real-time feedback from cost case features to control strategies, providing intelligent and efficient support for engineering management. At the same time, it is necessary to clarify the regulatory positioning, improve the legal system and technical supervision, ensure the compliant application of AI in financial market supervision, and promote the healthy and stable development of the market.

### *4.2 NCM-FPN Intelligent Reasoning Model*

This chapter provides a detailed introduction to the construction process of an intelligent inference model for engineering cost control strategy matching based on cloud models (NCM) and fuzzy Petri nets (FPN)[8]. This model aims to accurately match the characteristics of cost cases with control strategies through intelligent means, thereby improving the efficiency and accuracy of engineering cost control. The following is a comprehensive overview of the model construction process:

Firstly, the model utilizes NCM to quantitatively express and match the feature set of cost cases. A cloud model reflecting qualitative concepts and quantitative features was constructed by introducing three parameters: expectation (Ex), entropy (En), and superentropy (He). This study uses a reverse cloud generator to transform expert experience or actual data into an objective cloud model to accurately quantify the characteristics of cost cases. In the feature matching stage, the forward cloud algorithm is used to calculate the similarity (i.e. membership degree) between the target case and the features of each case in the case library, providing basic data for subsequent inference.

Secondly, a knowledge representation and inference mechanism was designed based on FPN. FPN represents causal relationships of knowledge through production rules, combining fuzzy rules with graphical models to establish an intuitive and structured framework for fuzzy reasoning. In FPN, the propositions represented by the library, the transitions represent the rules, and the weights on the directed arc reflect the quantitative support between the premises and conclusions in the rules. By defining transition thresholds and mapping functions, control over rule triggering conditions has been achieved. In the reasoning process, based on the degree of authenticity of the input proposition (i.e. feature similarity), the degree of authenticity of the output proposition is calculated through matrix transformation and iterative algorithms to determine whether there is a corresponding control strategy.

To optimize the parameters of FPN and improve the matching accuracy and generalization ability of the model, a sample training method was adopted. By dividing the training set and validation set, the FPN is simulated and inferred multiple times using training samples, and parameters such as directed arc weight, transition threshold, and confidence are adjusted based on the inference results. When the matching accuracy reaches the preset standard, the model is considered stable and the final parameter values are output.In summary, this chapter has achieved intelligent mapping and response from cost case features to control strategies by constructing an intelligent inference model for engineering cost control strategy matching based on NCM-FPN. This model not only improves the accuracy and efficiency of engineering cost control, but also provides useful reference and inspiration for intelligent decision-making in similar fields.

### *4.3 Example Analysis*

In order to verify the effectiveness of cost control methods for construction projects in the big data environment, this chapter carefully selected two sets of historical datasets for power grid construction project costs, which cover engineering projects with complex construction conditions and variable factors affecting project costs. Taking the design budget control in the decision-making design phase as the core, this chapter conducted detailed case simulations in high-performance computing environments using advanced tools such as IBM SPSS Statistics 25, Modeler 18, and Matlab 2018a. A main network and distribution network cost case database containing high-quality case information was constructed by cleaning, integrating, and dimensionality reducing raw data. In terms of engineering cost interval prediction, the CBR-MKRVM-KDE model has been used to achieve high-precision cost interval prediction. Through comparison with various comparative algorithms, it has been verified that it can effectively reduce the width of the prediction interval and improve the reliability of prediction while maintaining high prediction accuracy. In addition, the intelligent matching model based on NCM-FPN has demonstrated excellent performance in the selection of engineering cost control strategies. Through optimized parameter configuration, intelligent matching of control strategies has been achieved, with a matching accuracy rate of up to 93.33%, significantly better than traditional methods such as BP neural networks. The research results of this chapter not only provide strong support for the reasonable determination and control decision-making of power grid engineering costs, but also demonstrate the enormous potential of big data and intelligent algorithms in engineering cost management.

## 5. Conclusion

This article has made significant progress in the field of data asset valuation, but there are still some shortcomings and broad space for further research in the future. Firstly, due to the limitation of data openness, this study is mainly based on data from Youyi Data Network, which limits the diversity and breadth of data sources. In the future, with the gradual opening of data trading platforms and the deepening of data sharing, it is necessary to further integrate data resources from multiple platforms to comprehensively verify and expand the universality of research conclusions.

Secondly, this study focuses on textual numerical data assets, while the value evaluation of unstructured data assets such as images, videos, and audio has not been deeply explored. With the explosive growth of multimedia data, these types of data assets are playing an increasingly important role in value creation and trading. Therefore, future research should strengthen the exploration of the value evaluation of unstructured data assets and compare the differences in value characteristics and evaluation methods of different types of data assets.Finally, in terms of modeling methods, this paper adopts machine learning models such as multiple linear regression, B-P neural network, and random forest, and finds that the random forest model performs the best in data asset value evaluation. However, the field of machine learning is constantly evolving, with more advanced methods and technologies emerging. Future research can attempt to combine more types of machine learning models or deep learning methods, such as convolutional neural networks (CNN), recurrent neural networks (RNN), etc., to further improve the accuracy and efficiency of data asset valuation. At the same time, it is also possible to explore fusion and integration strategies between different models to fully utilize their respective advantages and achieve a more comprehensive evaluation of data asset value.

ISSN 2522-3488 Vol. 8, Issue 5: 42-47, DOI: 10.25236/IJNDES.2024.080507

In summary, future research should focus on expanding data sources, deepening the exploration of unstructured data asset valuation, and introducing more advanced modeling methods and technologies to promote the sustainable development and innovation of the field of data asset valuation.

## References

*[1] Cao Y, Cao P, Chen H, Kochendorfer K M, Trotter A B, Galanter W L, & Iyer R K. Predicting ICU admissions for hospitalized COVID-19 patients with a factor graph-based model[J]. In Multimodal AI in healthcare: A paradigm shift in health intelligence 2022, 245-256.*

*[2] Khatir A, Capozucca R, Magagnini E, et al. Structural Health Monitoring for RC Beam Based on RBF Neural Network Using Experimental Modal Analysis[C]. International Conference of Steel and Composite for Engineering Structures. 2023. DOI:10.1007/978-3-031-24041-6_7.*

*[3] Liming W, Yingming L, Xiaodong Q W. Environmental economic dispatch method of power system based on multiobjective artificial bee colony algorithm[J]. Electrical engineering, 2024, 106(1):567-579.*

*[4] Leung M Y, Wei X, Ojo L D. Developing a Value-Risk Management Model for Construction Projects[J]. Journal of construction engineering and management, 2024, 150(1):1.1-1.13.*

*[5] Almeida R, Abrantes R, Mário Romo, et al. The Impact of Uncertainty in the Measurement of Progress in Earned Value Analysis[J]. Procedia Computer Science, 2021, 181:457-467. DOI:10. 1016/ j. procs.2021.01.191.*

*[6] Liu Dongping. Discussion on the Method of Valuation with Bill of Quantities [J]. Foreign Science and Technology Journal Database (Abstract Edition) Engineering Technology, 2022 (4): 103-107*

*[7] Dui H, Lu Y, Chen L. Importance-based system cost management and failure risk analysis for different phases in life cycle[J]. Reliability engineering & system safety, 2024, 242(Feb.):1.1-1.22.*

*[8] Bollen K A, Gates K M, Luo L, et al. A Model Implied Instrumental Variable Approach to Exploratory Factor Analysis (MIIV-EFA)[J]. Psychometrika, 2024, 89(2):687-716. DOI:10.1007/ s11336 -024-09949-6.*