Research on Substation Engineering Cost Prediction Technology Based on Variable Bayes Deep Learning Optimization

Xiao Zeng[1, a], Xiaoling Peng[1, b], Haiyan Tong[1, c], Xiu Xu[2, d*]

[1]State Grid Hubei Electric Power Co., Ltd, Wuhan, Hubei Province, 430000, China
[2]Hubei Sanqingtai Technology Co., Ltd, Wuhan, Hubei Province, 430000, China

[a]e-mail:70725510@qq.com, [b]e-mail:21012284@qq.com , [c]e-mail:349323045@qq.com, [d*]e-mail:juzen123@126.com

Abstract. The cost of power grid projects is a multivariable and highly nonlinear problem, characterized by complex uncertainties and significant variability. As investment scales expand, the interplay of technical, environmental, and operational factors further complicates cost prediction. Accurate modeling requires a combination of advanced statistical methods and machine learning techniques to address these challenges. This paper integrates variational Bayes deep learning with engineering-specific data preprocessing, including normalization and principal component analysis (PCA), to ensure precise predictions and robust decision-making. The factors affecting the cost of power grid engineering are complicated. When the engineering situation is complex and changeable, it is difficult to obtain the reliable prediction results of a single project through the experience estimation of technical personnel. Therefore, this paper proposes a substation engineering cost prediction technology based on variational Bayes deep learning. This method is obviously better than other models in terms of prediction skills and prediction reliability, and can provide effective uncertainty estimation and prediction results.

Keywords: substation engineering, variational Bayes, cost prediction technology

1 Introduction

As the mainstay of national economic development, power grid enterprises play an important role in economic construction. Therefore, the cost of power grid engineering is not only related to the economic benefits of enterprises, but also related to the standardized and orderly development of the engineering construction market. Therefore, the scholars have conducted the following research on the cost control of power grid engineering:

Literature [1] adiscusses the application of hybrid models in construction project cost prediction, combining regression analysis, principal component analysis, and

S. M. Ghadiri et al. (eds.), *Proceedings of the 2025 4th International Conference on Engineering Management and Information Science (EMIS 2025)*, Advances in Computer Science Research 121,
https://doi.org/10.2991/978-94-6463-736-6_14

artificial neural networks (ANN) to improve prediction accuracy. Literature [2] presents a cost prediction model for road construction using machine learning techniques like Ridge, LASSO, K-NN, and Random Forest. Literature [3] combines BP neural networks with construction cost prediction, showing significant accuracy improvements. Literature [4] highlights using machine learning, especially ANN, to improve cost predictions in construction projects. Literature [5] analyzes unit costs of substation and line projects, revealing that their cost distributions approximate normal distributions.

Literature [6] proposes a cost prediction model for substations based on the Sparrow Search Algorithm optimized BP neural network, significantly improving prediction accuracy. Literature [7] develops a Big Data analytics system to enhance cost estimation accuracy and decision-making in power transmission projects. Literature [8] introduces the MK-TESM method for cost prediction in power transmission projects, achieving high accuracy with a mean absolute error rate within 10%. Literature [9] summarizes the influencing factors of substation and overhead line costs, extracts cost control indicators, and constructs a prediction model using neural networks and support vector machines.

In conclusion, current cost prediction for power grid projects mainly relies on statistical analysis and some traditional forecasting methods. Due to the complexity of factors influencing project costs, these methods often result in insufficient prediction accuracy, making it difficult to effectively guide cost control in power grid enterprises. This paper adopts a variational Bayes deep learning approach, which effectively combines the powerful representational capability of deep learning with the uncertainty handling advantages of Bayes methods. It overcomes the limitations of traditional methods in dealing with complex and uncertain data. This approach not only improves prediction accuracy but also provides a reliable estimation of the prediction's uncertainty, offering more precise and dependable cost control solutions for power grid enterprises.

2 Train of Thought

This paper proposes the prediction method of substation engineering cost based on variational Bayes deep learning, considering the complex uncertainty of the project cost prediction, using the deep neural network, capturing the influence of cognitive uncertainty on the model structure and potential parameters, estimthe uncertainty of the model output, and finally evaluating the effect of the prediction model. The construction idea of the technical method is shown in Figure 1 below:

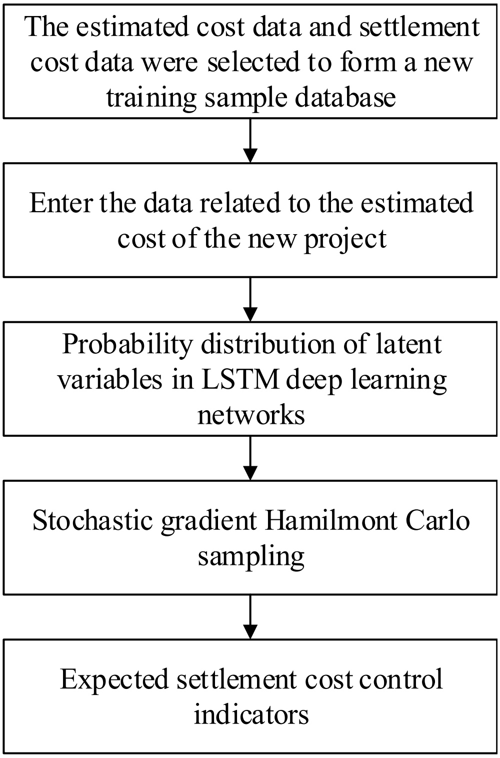

Fig. 1. Research idea diagram.

As can be seen from Figure 1, The technical approach begins with the preparation of a comprehensive project cost database, which includes detailed historical data on project costs, settlement records, and associated influencing factors. Data preprocessing involves standardization and dimensionality reduction to refine the quality and relevance of input variables. To handle uncertainties and nonlinear relationships in the data, the method employs an advanced deep learning model. This model utilizes structured data inputs, such as project specifications, environmental factors, and operational complexities, to capture the inherent variability in cost dynamics. Through iterative training and refinement, the model achieves robust predictions tailored to the unique challenges of substation engineering projects. Sampling was performed using a stochastic gradient Hamiltonian Monte Carlo probability prediction method, using the expected value as the predicted value of the probability. The implicit probability distribution of the cost control index prediction model is used to minimize the difference between the hidden variable probability distribution and the real distribution.

3 Model Building

3.1 Principles of the Bayesian Learning Algorithm

Bayesian learning is an important direction of machine learning research. It is based on Bayes' theorem, which assumes that the variables of the model follow a certain probability distribution, and reason according to this probability distribution and the observed data, so as to make the optimal decision. The basic principles are provided as follows:

D h ς Given a set of training instances, how do you infer the model that produces these data? We consider that the model is determined by the objective function, i. e., by the input variable mapping to the output variable, and the parameters of the data distribution, i. e. Θ h (ς,Θ) H $h = \arg\max\{P(h(\varsigma,\Theta)|D), h \in H\}$ For all possible model collections, then select. Since any category of objective function can learn a form function by fitting the training data, or directly specify the form, we are usually more concerned with the parameters. In other words, given a set of training instances, we are often able to infer the model by estimating the parameters, and the above formula can be repeated as. Θ D h $h = \arg\max\{P(\Theta|D), h(\Theta) \in H_\varsigma\}$

Taking the parameter value with the largest posterior probability as the estimate, the Bayesian formula provides us with a way to calculate the maximum possible value directly. It is defined as follows:

$$P(\Theta|D) = \frac{P(\theta)P(D|\theta)}{P(D)} = \frac{P(\theta)P(D|\theta)}{\oint_\theta P(\theta)P(D|\theta)d\theta} \alpha P(\theta)P(D|\theta) \tag{1}$$

$P(\theta)$ Where, is the initial probability assumed before the training observation data, also known as the prior probability. θ θ $P(D|\theta)$ D θ Denotes the probability of holds given the training set, commonly called the posterior probability. θ $P(D)$ Is given as the normalized constant.

D m In most cases, we need to find the assumptions that are the maximum given a set of training instances. In Bayesian methods, this assumption with the maximum likelihood is called the maximum posterior probability, or the maximum posterior hypothesis, denoted as. The question then changes to: θ_{MAP}

$$\theta_{MAP} = \arg\max\{P(\theta)P(D|\theta)\} \tag{2}$$

3.2 Principle of the Variational Bayes Algorithm

Variational Bayes is a kind of technique used for approximate computational complex (intractable) integration in the field of machine learning, which is mainly used to solve complex statistical models. It involves three variables: observation variable, hidden variable and unknown parameter. In the Bayesian approach, hidden variables and unknown parameters are collectively referred to as unobservable variables.

h According to the Bayesian formula, the problem of the inferred model is now transformed into a set of training instances to solve the posterior distribution of unobservable variables. D z

$$P_\theta(z|D) = \frac{P_\theta(z)P_\theta(z|D)}{P_\theta(D)} \tag{3}$$

θ $P_\theta(D)$ $P_\theta(z|D)$ It is often a region in high dimensional space, the calculation of probability is often very complex, and the accurate solution of probability becomes a challenge. Consider to approximate the posterior probability with an easily understood edge likelihood, where the error allows. Usually called the recognition mode, to deapproximate the true posterior probability. $P_\theta(D)$ $P_\theta(z|D)$ $P_\theta(z|D)$

$$q_\phi(z|D) \approx P_\theta(z|D) \tag{4}$$

ϕ Where, the symbol represents the parameters of the recognition model, also called the variational parameters. The identification model can be an arbitrary digraph model.

$$q_\phi(z|D) = q_\phi(z_1,\ldots,z_M|D) = \prod_{j=1}^{M} q_\phi(z_j|P_a(z_j),D) \tag{5}$$

$|P_a(z_j)$ Fanal node variables for variational parameters of the directed graph model. z_j $q_\phi(z|D)$ The model can usually be implemented by deep neural networks, and the variational parameters correspond to the weights and bias of the neural network.in compliance with: ϕ

$$(\mu, \log\sigma) = EncoderNeuralNet_\phi(D) \tag{6}$$

$$q_\phi(z|D) = \mathrm{N}(z; \mu, diag(\sigma)) \tag{7}$$

4 Empirical Analysis

The empirical analysis is based on a comprehensive dataset from a provincial power grid company, comprising 460 substation projects spanning a decade. Of these, 449projects were used for training the model, while 11 were set aside for validation.

The dataset includes diverse attributes such as substation capacity, equipment procurement costs, and regional environmental factors. In model evaluation, to ensure a more comprehensive evaluation of model performance, the root mean square error and coefficient of determination are introduced as important evaluation indicators. test the substation engineering test sample set cost prediction results as shown in the following table:

Table 1. Statistical table of the test sample model prediction results.

Sample number	actual value	Forecast results	RMSE	R^2
1	59.8	63.3	0.945	0.991
2	120.0	117.4	0.753	0.988
3	142.0	139.1	0.672	0.994
4	11.5	10.2	0.319	0.985
5	50.5	42.2	1.345	0.973
6	4.3	3.9	0.319	0.993
7	20.1	19.8	0.299	0.997
8	33.8	29.6	1.065	0.982
9	115.1	109.9	0.724	0.990
10	7.4	6.9	0.316	0.987
11	36.3	34.4	0.319	0.992

Note: The actual value is derived from a provincial power grid company.

Based on the visualization principle, the test set samples and the prediction results are plotted as shown in Figure 2 below.

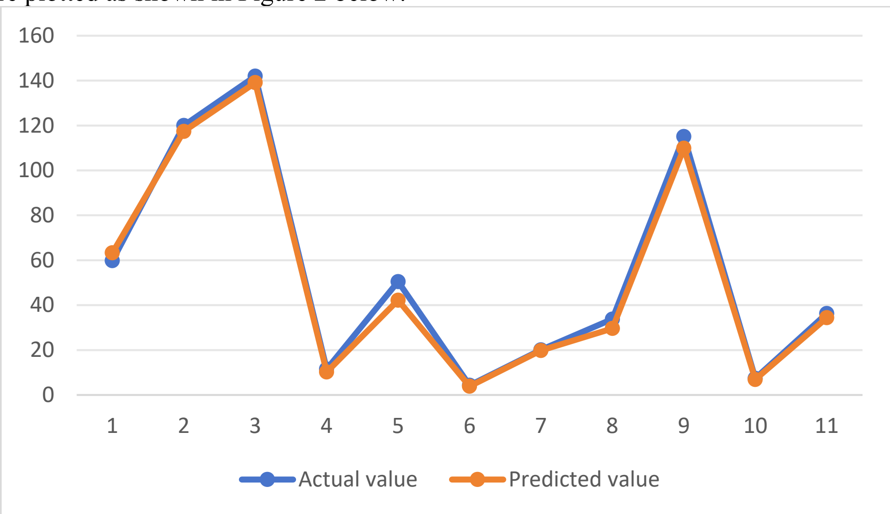

Fig. 2. Analysis diagram of the actual and predicted values.

As can be seen from Table 1 and Figure 2, each prediction point is close to the original actual value data point. Based on the evaluation metrics of RMSE and R^2, the model demonstrates small prediction errors and is capable of effectively explaining the variability in the data. This indicates that the model fits the actual data well and possesses strong predictive power. Overall, the variational Bayes deep learning method proposed in this paper performs excellently in handling cost prediction for power grid engineering projects, providing reliable cost control tools for power grid enterprises. It helps optimize investment decisions and improve project management efficiency.

5 Conclusion

This paper proposes a cost prediction method for power grid engineering projects based on variational Bayes deep learning. This approach effectively overcomes the challenge of achieving high prediction accuracy due to the complexity of cost-influencing factors. Experimental results demonstrate that the model exhibits high accuracy and reliability in cost prediction for power grid engineering projects, providing more precise cost control solutions for power grid enterprises. However, due to the nature of the model itself, there are areas for improvement. First, the scale and diversity of the dataset may impact the model's generalizability, especially when there are potential biases or deficiencies in the dataset. Second, although this study considers several common influencing factors, power grid engineering costs are affected by numerous external factors in practical applications. Future research can further explore how to incorporate these factors into the model to improve its adaptability and accuracy. The power grid company can integrate the predictive results from this model to conduct a scientific review of the cost control status of construction projects, comparing the predicted costs with the actual incurred costs. By performing this comparison, projects deviating from the budget can be promptly identified, which contributes to effective control of power grid construction costs and supports precise investment decisions.

References

1. Wang Y, Zuo J, Pan M, et al. Cost prediction of building projects using the novel hybrid RA-ANN model[J]. Engineering, Construction and Architectural Management, 2024, 31(6): 2563-2582.
2. Abed Y G, Hasan T M, Zehawi R N. Cost Prediction for Roads Construction using Machine Learning Models[J]. International journal of electrical and computer engineering systems, 2022, 13(10): 927-936.
3. Jia H. Project cost prediction for building complexes based on grey bp neural network model[J]. Computing, Performance and Communication Systems, 2022, 6(2): 1-9.
4. Matel E, Vahdatikhaki F, Hosseinyalamdary S, et al. An artificial neural network approach for cost estimation of engineering services[J]. International journal of construction management, 2022, 22(7): 1274-1287.
5. Li Yaming, Yao Gang, Han Wei. Power grid project cost analysis and investment forecast based on probability statistics [J]. Electric Power survey and design, 2017,(06):76-80.

6. Xiaomin X, Luyao P, Zhengsen J, et al. Research on Substation Project Cost Prediction Based on Sparrow Search Algorithm Optimized BP Neural Network [J]. Sustainability, 2021, 13(24):13746-13746.
7. Delgado D M J ,Oyedele L ,Bilal M , et al.Big Data Analytics System for Costing Power Transmission Projects[J].Journal of Construction Engineering and Management, 2020, 146(1):05019017-05019017.
8. Ding Zhengzhong, Peng Luwei. (2021). A Cost Prediction Method for Power Transmission Projects Based on the MK-TESM Method. Journal of Shenyang University of Technology, 43(02), 126-131.
9. Li M, Qixin W, Shuang L, et al.Research on Cost Prediction of Power Transmission and Transformation Project Based on Combination Prediction Model [J].IOP Conference Series: Earth and Environmental Science,2021,769(4):