



Research on the Statistical Analysis Technology of Power Grid Engineering Project Category Cost Based on Big Data Analysis Technology

Hongyi Chen^a, Huiying Wu^{b,*}, Minquan Ye^c, Biao Chen^d

State Grid Fujian Economic Research Institute, Fuzhou, Fujian, 350000, China

^achenhongyi0708@163.com, ^{b,*}wuhuiying2021@163.com,
^chdymq2014@163.com, ^dchenbiao351@163.com

Abstract. For power grid engineering projects, the project consists of a number of projects to clear a certain amount of engineering, the completion of the project involves a series of different categories of costs, the existing technology of these categories are usually defined as labor costs, mechanical equipment costs, material costs, management fees, profits and taxes. Therefore, how to efficiently analyze these huge category of data based on data analysis technology has become an urgent technical problem to be solved. Therefore, in order to solve the defects in the existing technology that it is difficult to compare and analyze the cost data of many power grid categories, this paper puts forward the statistical analysis technology of the cost of power grid project categories based on big data analysis technology.

Keywords: big data analysis technology, power grid engineering, project category cost, statistical analysis technology

1 Introduction

The project cost based on the quota valuation mode is difficult to reflect the actual consumption of the construction department, cannot accurately deal with the market price changes, and ultimately affects the results of the project cost analysis, and directly affects the correctness of the investment decision of the project, and has a significant impact on the bidding project of the power grid project and the cost management in the later period. At the same time, the amount of power grid project cost data is large, involving multiple budget, bidding, settlement and other subsystems. Manual collection of these data and cost analysis are easy to lead to data inconsistency and low analysis efficiency.

Literature [1] uses the work breakdown structure (WBS) method to decompose and classify the intrinsic attributes, physical structure, and technical characteristics of the distribution network project itself, and gives the basic principles and ideas of coding and setting for the subdivided project level and data unit. Literature [2] Using multi-grid technology and VAR model, a set of power grid engineering cost data infor-

mation analysis methods were established to achieve accurate analysis and prediction of construction costs. Literature [3] proposes a big data-based investment budgeting, accounting, analysis and evaluation method for the main distribution network of smart economic parks, and budgets the main network and distribution network separately through different strategies. Literature [4] Based on the big data environment, the data management process and management framework of power grid engineering cost are proposed, hoping to strengthen the management and control of power grid engineering cost. On the basis of combing the full cost data system of power grid engineering, literature [5] systematically proposes a normalized management framework for power grid engineering cost analysis under big data storage from three aspects: division of labor and cooperation, supervision and management, and comprehensive evaluation. In literature [6], an intelligent cost management and control model based on edge computing was constructed for the confirmation of the whole link unit of power transmission and transformation engineering, and the immune particle swarm optimization algorithm based on hybrid strategy was used to optimize the model. In literature [7], a data processing method based on Particle Swarm and Genetic Hybrid Optimization Algorithm (PSO-GA) was constructed to improve the timeliness and accuracy of cost analysis of power transmission and transformation projects. Literature [8] proposes a method for predicting the cost data of power transmission and transformation projects based on the MK-TISM method. Literature [9] combined with the problem that there are a large number of pictures in the cost data source file of power transmission and transformation projects and the difficulty of data extraction, a picture text recognition method based on image preprocessing is proposed, and the B+ tree and index technology are combined to retrieve data. In Ref. [10], the influencing factors of 110kv substation project cost were qualitatively identified, and the principal component analysis, multiple linear regression and path analysis methods were used to screen out the key influencing factors of 110kv substation project cost with the help of the data analysis function of SPSS software, and their influence degree was ranked. Literature [11] highlights the impression and effect on decision-making through big data. The investigation's rationale is to dig deep insight into the buzzword to enable stakeholders to understand the challenges and opportunities that big data has brought in the current business scenarios. It also discusses applications of big data-influenced decision-making, along with state-of-the-art big data techniques and technologies. Literature [12] presents an investigation of the linkages between Big Data Analytics (BDA) capabilities, circular economy (CE) practices, and sustainable supply chain (SSC) flexibility on the SSC performance.

To sum up, the current research on power grid project cost is more based on the perspective of project management, with the purpose of serving project budget and process cost control, while the analysis and comparative analysis of the cost data of different types of power grid projects is relatively weak.

The proposed power grid project category cost statistical analysis technology based on big data analysis technology solves the defects that it is difficult to compare and analyze the cost data of many power grid categories compared with the existing technology. The research ideas of this paper are as shown in Figure 1 below:

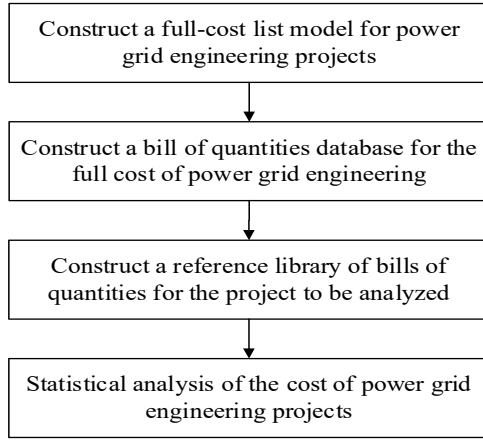


Fig. 1. Technical Roadmap of this Paper.

2 Construction of Full List Model of Power Grid Project

The purpose of constructing the full cost list model of the power grid engineering project is to determine the apportionment coefficient model θ through a certain method, and then determine the full cost list model M , and allocate the cost collection of the cost document to the bill of quantities collection, so as to realize the labor cost, material cost, machinery use fee, measure project cost, management fee, profit, regulation fee and tax are all allocated to the comprehensive unit price of the list. In this way, the comprehensive unit price of the list after cost allocation reflects all the costs of the power grid engineering project, and the relationship with the engineering cost of the project is more direct, and it is convenient to directly analyze or compare the comprehensive unit price of the project to be analyzed. The specific steps are as follows:

(1) Import data, import project cost file F_i based on bill of quantities; split project cost file F_i , split into bill of quantities set and cost set, wherein bill of quantities:

$$M_i = \{ID, M_i^a, M_i^{Norm}, M_i^{Unit}, M_i^{P_0}, M_i^\omega, M_i^p\} \quad (1)$$

ID indicates the fixed number of the list item M_i ; M_i^a represents the quantity of the inventory item M_i , M_i^{Norm} represents the characteristics of the inventory M_i , M_i^{Unit} represents the unit of the inventory item M_i , $M_i^{P_0}$ represents

the basic cost of the inventory item M_i , M_i^ω represents the apportioned cost of the inventory item M_i , and MP represents the full cost of the inventory item M_i . Here, the properties of the bill of quantities can be added or removed based on the project file to be analyzed. The formula for fee items is as follows:

$$O_i = \{ID, O_i^{Name}, O_i^p\} \tag{2}$$

ID indicates the fixed number of the O_i of the expense item, O_i^{Name} indicates the name of the O_i of the expense item, and O_i^p indicates the cost of the expense item.

(2) Share the cost of the bill of quantities, and share the cost set of the project cost document to the set of the bill of quantities. The allocation method is as follows:

$$\forall M_i \in M, M_i^p = M_i^{p0} + \omega_i \tag{3}$$

Among them, it represents the share of the cost collection allocated to the project quantity clearance item:

$$\omega_i = \sum_{o_i \in 0} (\theta_i \times o_i^p), M_i^\omega = \omega_i \tag{4}$$

Here are three ways to obtain them:

Example 1: Share the cost items according to the weight method:

There are two types of A and B for the project quantity clearance of A cost document, among which A has installation requirements, and B has no installation requirements to share the installation cost to A and B. According to the requirements, the installation cost allocation coefficient of A is 1, and the installation cost allocation coefficient of B is 0.

Example 2: Share the cost item according to the category method θ_i :

For example, there are two kinds of quantity items A and B in A cost document, and share the construction cost C to A and B, the allocation amount of A and B is $C / 2$. At this time, the allocation coefficient of A and B is $1 / C (M)$, in which C (M) indicates the number of quantity items. θ

Example 3: Share the cost items according to the mean method:

The quantity items of A cost document are A and B, including A and B; now share the installation cost C (whether there are installation requirements), the installation cost C of A and B is; then the allocation coefficient of any project quantity is, which represents the total project quantity. $\frac{C}{a+b} * a - \frac{C}{a+b} * b \theta 1 / count \ count$

Thus, record the specific allocation coefficient of the cost set O of each single item for subsequent comparison analysis:

$$M_i = \{ID, \theta_1, \theta_2, \dots, \theta_n\} \quad (5)$$

Generate the project BOQ model: and the matching set of allocation coefficients:

3 Construction of the Full-Cost Bill of Quantities Database for Power Grid Engineering

Build the full cost bill of quantities database, read the historical power grid project documents, and build the full cost bill of quantities database and the corresponding allocation coefficient database according to the historical power grid project documents. In this step, it is the bill of quantities based on the historical project documents. According to the principle of consistency of fixed number and uniqueness, the quantities of all historical project documents are combined to form the full cost bill of quantities database and the corresponding allocation coefficient library. The specific steps are as follows:

(1) According to the full cost list model M and the apportionment coefficient model θ of the historical engineering project documents, the bill of quantities set M' ($M'_i = \{ID, M'_i{}^a, M'_i{}^{Norm}, M'_i{}^{Unit}, M'_i{}^{P_0}, M'_i{}^\omega, M'_i{}^p, \dots\}$) and the apportionment coefficient set θ' ($M'_i = \{ID, \theta'_1, \theta'_2, \dots, \theta'_n\}$) are constructed. The bill of quantities model library F ($F : F_1, F_2, \dots, F_i, F_n$) is constructed in the following steps:

$$\underbrace{\quad}_{M' \theta}$$

1) According to the full cost list model M and the apportionment coefficient model θ of all engineering project documents, all attributes and apportionment factor attributes of the bill of quantities items in all engineering project documents are merged and the number of times is set for each numerical attribute;

2) Merge all bill of quantities items into bill of quantities set M' and allocation factor set θ' according to the full cost bill model M and apportionment factor model θ of all project documents;

3) M' contains all kinds of bill of quantities items that appear in all historical engineering project documents. For any bill of quantities item M , when merging to M' , if M_i does not contain the attributes of M_i , the attribute value of the attribute is left blank;

4) θ' contains the set of allocation factors for all bill of quantities items that appear in all historical project documents;

5) For any bill of quantities item, when merging to θ' , if the M_i does not contain the attribute of M' , the attribute value of the attribute is left blank.

(2) To build the full cost quantity bill library, according to the quantity clear single fixed number consistency principle and quantity clear single character attribute matching principle, the bill of quantities set and allocation coefficient set clustering, to be scattered in the multiple historical project files of similar clear single aggregation

statistics, to make the whole cost bill of quantities in the same kind of clear single item only, easy to treat the analysis of project file clear single comparative analysis, the steps are as follows:

According to the collection structure of the bill of quantities, the establishment of the full cost bill of quantities library, structure, its field is composed of all the attributes in the bill of quantities collection, the main field includes the fixed number of the quantity item, the quantity item, the specification description of the quantity item.

M' in the bill of quantities set M'_i is filtered out from the bill of quantities set M' according to the fixed number ID of the bill of quantities item in turn; all the filter information of the bill of quantities item M'_i is classified, mainly according to the fixed number ID and character attribute of the list item: all the same attribute values of the number ID and character type attribute are the bill of quantities items in the same class case; the reference information of the numerical attribute of the bill of quantities item in the same class is calculated, so that the record of the bill of quantities item with the same attribute value of the number ID and character type attribute in the full-cost bill of quantities library M'' is unique, wherein the calculation formula for calculating the full-cost M_i^p of a certain class of the bill of quantities item M'_i is as follows:

Calculated average values M_i^p :

$$\bar{M}_i^p = \frac{\sum_{M'_i} M_i^p}{c(M'_i)} \quad (6)$$

Calculated variance M_i^p :

$$S^2(M_i^p) = \frac{\sum_{M'_i} (M_i^p - \bar{M}_i^p)^2}{c(M'_i)} \quad (7)$$

Reference information counted M_i^p :

$$M_i^p = \bar{M}_i^p \pm S^2(M_i^p) \quad (8)$$

The statistics of the quantity of single reference information into the full cost of the bill of quantities library, establish the allocation coefficient of the quantity of single item library structure, the field of allocation coefficient set, all the allocation attributes, the main field including quantity clear item fixed number and a series of allocation

attributes. The reference information of the single allocation coefficient in the statistical allocation coefficient collection and stored in the apportionment coefficient database, the set of the quantity clearance selects all the information from the allocation coefficient collection according to the fixed number; the reference information of each allocation coefficient of the statistical quantity clearance item, so that the records of the fixed number in the allocation coefficient database are unique.

Calculate the mean value of each allocation coefficient:

$$\forall \theta'_i, \bar{\theta}'_i = \frac{\sum \theta'_i}{C(\theta'_i)} \quad (9)$$

Calculate apportionment variance:

$$\forall \theta'_i, S^2(\theta'_i) = \frac{\sum (\theta'_i - \bar{\theta}'_i)^2}{C(\theta'_i)} \quad (10)$$

Reference information for statistical allocation coefficient:

$$\theta''_i = \bar{\theta}'_i \pm S^2(\theta'_i) \quad (11)$$

It represents the number of times that the apportionment coefficient of the item appears in the set of apportionment coefficient.

Store the apportionment coefficient information of the statistical project quantity clearance item into the apportionment coefficient database:

Bill of quantities library:

$$M_i'' = \left\{ ID, M_i^{a''}, M_i^{Norm''}, M_i^{Unit''}, M_i^{P_0''}, M_i^{\omega''}, M_i^p'', \dots \right\} \quad (12)$$

Sharing coefficient library:

$$M_i'' = \left\{ ID, \theta_1'', \theta_2'', \dots, \theta_n'' \right\} \quad (13)$$

4 Construction of Reference Library of Engineering Bill to be Analyzed

(1) Construct the full cost list model of power grid engineering projects to be analyzed.

Read the power grid project documents to be analyzed, split the power grid project documents to be analyzed, and divide them into bill of quantities collection and cost collection.

$$M_i^t = \left\{ ID, M_i^{t,a}, M_i^{t, Norm}, M_i^{t, Unit}, M_i^{t, P_0}, M_i^{t, \omega}, M_i^{t, P} \right\} \quad (14)$$

It represents the list, the fixed number, represents the quantity of the list, represents the characteristics of the list, represents the list, the unit, represents the basic cost of the list, represents the list, the apportionment cost, represents the full cost of the list.

Similarly, this Bill of Quantities attribute is added as required.

Where, the cost estimate set O:

$$O_i^t = \left\{ ID, O_i^{Name}, O_i^{t,P} \right\} \quad (15)$$

Generate the project BOQ model:

$$O_i^t = \left\{ ID, O_i^{Name}, O_i^{t,P} \right\} \quad (16)$$

5 Statistical Analysis of Power Grid Project Category Cost

By the method of project category conversion, the full cost of the bill of quantities is converted into specific category cost, and the comparative analysis is made based on the category cost. The principle of comparison is the existing traditional idea. After the data processing in the above chapter, the comparison is automatically analyzed by computer. The specific steps are described as follows:

(1) Conversion of the project cost category. Convert the full cost of the bill of quantities into labor cost R, mechanical equipment cost J, material cost C, management fee G, profit L and tax S for category analysis.

The conversion of the project cost category described consists of the following steps:

Generate the reference model of category cost. For analysis project documents, the bill of quantities reference library and bill of quantities allocation coefficient reference library, for any quantities clear the reference cost allocation, according to the bill of quantities allocation coefficient of the corresponding quantities clear single allocation coefficient allocation to category cost reference model, generate category cost reference model, the expression is as follows: Δ^t

$$\Delta^t : M_i^t (ID, \Delta_R, \Delta_J, \Delta_C, \Delta_G, \Delta_L, \Delta_S) \quad (17)$$

$ID, M_i^t, \Delta_R, \Delta_J, \Delta_C, \Delta_G, \Delta_L, \Delta_S$ Among them, it is the number of the project quantity clearance item, which corresponds to the specific value of the labor cost R, mechanical equipment cost J, material cost C, management fee G, profit L and tax S respectively. M_i^t

$$\Delta_R = \theta_R^{t^*} * M_i^{t, \omega^*} \quad (18)$$

$$\Delta_J = \theta_J^{t''} * M_i^{t,\omega''} \quad (19)$$

$$\Delta_C = \theta_C^{t''} * M_i^{t,\omega''} \quad (20)$$

$$\Delta_G = \theta_G^{t''} * M_i^{t,\omega''} \quad (21)$$

$$\Delta_L = \theta_L^{t''} * M_i^{t,\omega''} \quad (22)$$

$$\Delta_S = \theta_S^{t''} * M_i^{t,\omega''} \quad (23)$$

The apportionment coefficients of labor cost (R), machinery and equipment cost (J), material cost (C), management fee (G), profit (L), and tax (S) correspond to their respective cost categories in the allocation framework.

Generate the reference model of the category to be analyzed. According to the reference database of the bill of quantities of the project documents to be analyzed, use the generation method of generating the category cost reference model to generate the reference model of the project to be analyzed. The expressions are as follows:

$$\Delta'' : M_i^{t''} (ID, \Delta_R, \Delta_J, \Delta_C, \Delta_G, \Delta_L, \Delta_S) \quad (24)$$

$$\Delta_t : M_i^t (ID, \Delta_R, \Delta_J, \Delta_C, \Delta_G, \Delta_L, \Delta_S) \quad (25)$$

Compare and analyze the project category and cost of the project. Analyze the total category cost of the engineering project, which is expressed as follows:

$$R^t = \sum_{M_i^t} M_i^a * \Delta_R, R^{t''} = \sum_{M_i^{t''}} M_i^a * \Delta_R \quad (26)$$

$$J^t = \sum_{M_i^t} M_i^a * \Delta_J, J^{t''} = \sum_{M_i^{t''}} M_i^a * \Delta_J \quad (27)$$

$$C^t = \sum_{M_i^t} M_i^a * \Delta_C, C^{t''} = \sum_{M_i^{t''}} M_i^a * \Delta_C \quad (28)$$

$$G^t = \sum_{M_i^t} M_i^a * \Delta_G, G^{t''} = \sum_{M_i^{t''}} M_i^a * \Delta_G \quad (29)$$

$$L^t = \sum_{M_i^t} M_i^a * \Delta_L, L^{t''} = \sum_{M_i^{t''}} M_i^a * \Delta_L \quad (30)$$

$$S^t = \sum_{M_i^t} M_i^a * \Delta_S, S^{t''} = \sum_{M_i^{t''}} M_i^a * \Delta_S \quad (31)$$

$$\sum^t = R^t + J^t + C^t + G^t + S^t \tag{32}$$

$$\sum^{t''} = R^{t''} + J^{t''} + C^{t''} + G^{t''} + S^{t''} \tag{33}$$

R^t, J^t, C^t, G^t, S^t are the costs of all items of the project to be analyzed $R^{t''}, J^{t''}, G^{t''}, S^{t''}, \sum^t, \sum^{t''}$ are the reference costs for the completion of all items of the bill of quantities; the total reference costs of the project and the total reference costs of the project.

Analysis of the project total category cost and reference total category cost difference and analyze the causes of the difference, determine whether the project total category cost within the range of the total category cost fluctuation R^t, J^t, C^t, G^t, S^t and $R^{t''}, C^{t''}, J^{t''}, G^{t''}, S^{t''}$, the difference between the corresponding category of cost, judgment to analyze all kinds of cost in the range of cost of reference.

Project category cost category analysis of a single project quantity clearance item, analyze the difference between the cost category of a project quantity clearance item and the reference cost category:

$$\begin{cases} M_i^{t''}(ID, \Delta_R, \Delta_J, \Delta_C, \Delta_G, \Delta_L, \Delta_S) \\ M_i^t(ID, \Delta_R, \Delta_J, \Delta_C, \Delta_G, \Delta_L, \Delta_S) \end{cases} \tag{34}$$

Compare the cost of the same project, analyze the differences between the corresponding cost categories, and judge whether the cost of the project to be analyzed is within the fluctuation range of the cost of the reference project quantity.

6 Conclusion

Reasonable control of power grid project cost, improvement of project quality and investment benefit have an important meaning for strengthening the project cost management of power grid enterprises. Therefore, this paper puts forward the statistical analysis technology of power grid project category cost based on big data analysis technology, so as to realize the efficient analysis of the huge category data of power grid engineering and improve the efficiency of data analysis.

References

1. Feng Kaijian, Wang Bingqian, Yuan Zuwei, etc. Structure decomposition and classification of cost data of distribution network transformation project based on WBS [J]. Electric Power Survey and Design, 2021, No.158(09): 74-80.
DOI: 10.13500/j.dlkcsj.issn1671-9913.2021.09.014.

2. Ding Yan, Zhang Haiwen, Sun Yongyan. Research on the information analysis method of power grid engineering cost data based on multi-grid technology [J]. *The Electronic Design Engineering*, 2021, 29(19): 35-39. DOI:10.14022/j.issn1674-6236.2021.19.007.
3. Cui Jindong, Zheng Que, Zhou Yancheng, etc. Research on the cost management method of power grid Engineering in Smart Economy Park based on Big data [J]. *Science and Technology Management Research*, 2018,38 (06): 166-174.
4. Song Yuqian. The analysis and control of power grid engineering cost under the big data environment [J]. *Information Construction*, 2016, No.213(02):87.
5. Liu Hongzhi, Tu Qingbo, Han Yanfeng. Research on the cost analysis and control system of power grid engineering under the big data environment [J]. *East China Electric Power*, 2014,42 (12): 2722-2727.
6. Luan Ling, Pan Lianwu, Yan Lei, et al. Research on Precise Cost Intelligent Control Technology for Whole-link Unit Confirmation of Power Transmission and Transformation Engineering Based on Edge Computing [J]. *Computer Science*,2021,48(S2):688-692.)
7. Ye Qing, Liu Xiaomeng. Data processing of the whole process cost management of power transmission and transformation engineering based on particle swarm and genetic hybrid optimization algorithm [J]. *Electric Power and Energy*, 2021, 42(05):521-526+542.
8. Ding Zhengzhong, Peng Luwei. Cost data prediction method of power transmission and transformation project based on MK-TESM method [J]. *Journal of Shenyang University of Technology*, 2021, 43(02):126-131.)
9. Chen Kailing, Shen Hong, Gu Wen, et al. Research on intelligent extraction and integration technology based on power transmission and transformation project cost data [J]. *Project Management Technology*, 2020, 18(12): 136-141.)
10. Hu Jinlan, Gao Xiaobin. *Journal of North China Electric Power University (Social Science)*, 2016(05): 55-59+64. DOI: 10.14092/j.cnki.cn11-3956/c.2016.05.009.
11. Naqvi R, Soomro T R, Alzoubi H M, et al. The nexus between big data and decision-making: A study of big data techniques and technologies[C]//The international conference on artificial intelligence and computer vision. Cham: Springer International Publishing, 2021: 838-853.
12. Edwin Cheng T C, Kamble S S, Belhadi A, et al. Linkages between big data analytics, circular economy, sustainable supply chain flexibility, and sustainable performance in manufacturing firms [J]. *International Journal of Production Research*, 2022, 60(22): 6908-6922.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

