Application Research of Big Data Technology in Engineering Cost

Songyang Yan*
School of North China Electric Power University, Baoding, China

*Corresponding author e-mail: 201806000725@ncepu.edu.cn

Abstract. In view of the frequent use of information technology in practical engineering cost projects in recent years, the engineering cost industry slowly accumulated a lot of information data. The development of engineering cost industry has entered a new era of data era. The disadvantages are that there is no unified standard for project cost information data, information interconnection is difficult, the frequency of updating project cost data is slow, and the value of information has not been fully reflected. Therefore, in order to solve these problems, the most advanced information technology is applied to the construction of the project cost platform by relying on large number technology, and the project cost information data is collected, sorted out and analyzed, so that these data are full of its deserved value. The process of building a platform of the need to establish a unified standard, engineering cost data shall be carried out in accordance with the standard information collection, sorting and processing, for project cost related units to query, data model, is applied to analyze the information data at the same time, to invest in the project cost, engineering material price analysis, all staff to provide decision support for engineering cost industry. This paper USES questionnaire survey to obtain data, and analyzes the data to find that 50% of the cost information obtained comes from professional journals, and only 19% of engineering cost workers can use the Internet.

Keywords: Big Data Technology, Project Cost, Questionnaire Survey

1. Introduction
At present, with the rapid development of big data technology, new mobile phone technology represented by 5G technology, Internet of Things, cloud computing and other applications widely appear in all social strata, constantly changing people's production and life style, and promoting great social progress [1]. The application of information technology is inseparable from big data. The key role of "big data" is to evaluate and predict the changes of things, that is, to use mathematical algorithms to find the relationship between things in a large amount of data and to predict the development and changes of things. With the extensive application of information technology in the field of engineering cost, a large number of data have accumulated in the engineering cost information data [2]. In order to make better use of the data, in addition to the classification and storage of the data, it is also necessary to sort and classify the data. In the face of these massive project cost data, the
simple use of non-automatic processing not only costs a lot of time, but also easy to make mistakes. Therefore, modern science and technology should be used to discover the value of such data to promote the development of different industries.

With the gradual maturity of big data technology and its entry into people's vision, people begin to think about how to apply big data technology in engineering cost [3]. Mr. Fan think big data for engineering cost is to guide and regulate role, along with our country government and the market is becoming more and more attention of cost of the project, the project cost has a little research significance and realistic significance of engineering cost as the breakthrough point, respectively, this paper expounds the general research methods and index system of project cost, enumerated instances to analyze big data under the perspective of the application of the engineering cost index [4]. Some professional scholars use fuzzy mathematics, neural network technology and gray prediction model. Fuzzy mathematics method W fuzzy mathematics knowledge is the main method to determine the degree of similarity of engineering projects and estimate the cost of engineering projects [5]. BP neural network is an operational model, which is composed of a large number of neurons and the interconnections between neurons. Neural network has memory learning ability, can approximate a certain algorithm or function with arbitrary precision, can excavate the logic relation of data and express it. Gray prediction is to dig up the past and present data to build a partial information uncertain model from the history to the future, thus predicting the development of the system [6]. Other research groups used fuzzy mathematics theory and big data technology, combined with the actual situation of the project, through the project characteristic media, and applied the membership function and the closeness degree between the completed project and the project to achieve the project cost estimation [7]. Some excellent teams in the world have studied the role of BIM and big data technology in project cost management, pointed out the advantages of big data technology in cost information acquisition and other aspects, and pointed out that one of the major difficulties in the current construction of cost big data is the non-standard cost management [8].

Although these studies are very fruitful, but there are still some shortcomings. In this paper, based on large data technology support, with big data for research in the application of the engineering cost, using the method of literature, questionnaire survey set experimental methods such as come to the conclusion, on the basis of large data, under the technical support, process engineering cost information transition from paper to electronic and network methods, improve the cost effectiveness of the engineering process.

2. Method

2.1. Big Data Technology

With the rapid development of 5G technology and the continuous reform of the business model of high-tech Internet enterprises, the global data volume has been in the stage of rapid development for a long time. Data is the foundation of business analysis. In the traditional sense, data refers to more "digital" information that can be directly analyzed, also known as structured data. Facing the increasing amount of information, text, video, audio and image data also contain a lot of knowledge that can be mined [9, 10]. Big data usually has four characteristics: first, the amount of data is huge, moving from TERabytes to petabytes. Second, there are many types of data, including many types of structured data. Third, the processing speed is fast. Data processing speed up to 1 second. Fourthly, as long as the data are fully mined and analyzed, the price will be very high [11].

On large data processing, refining data source first, then the data sorting, extract the relevant contact and entity, after adopting relationship and polymerization method, in the process of refining and finishing, and big data, only through this link, data extraction and integration of talent with high quality and feasibility, finally in the right way to store large data. At present, the mainstream database technology has a mature extraction and integration pattern.
2.2. Project Cost

The cost of engineering products is called engineering cost. According to the different pursuit of economic interests between the main body of supply and demand, it can be defined from two aspects. From the owner's point of view, it can be defined as the total cost of construction investment, the total cost of fixed assets investment estimated or actually used in the construction project. From a management perspective, refers to the project price, including the project, project, equipment, termination of contract, installation price, etc. [12].

The characteristics of engineering cost knowledge are summarized as follows:

(1) Situational Dependence

The dependence on project cost projections means that any knowledge of project costs is generated in a specific context and must be in a specific context. This means that the knowledge of project cost is related to specific knowledge and practical activities in a specific context, which is specific or partial. Beyond this scope, its accuracy may be affected.

(2) Intention

The intentionality of engineering cost knowledge refers to the unexpressible nature of some engineering cost knowledge, namely the unexpressible and unexpressible nature of coding, which makes it unable to be transmitted and learned.

(3) Discreteness

Due to the division of labor, there are many market entities in the construction field, such as the construction unit, the construction unit, the supervision unit, the auditing unit, the consulting unit and the government management department, etc., and the engineering cost knowledge is also distributed among these different units [13].

2.3. Establishment of Prediction Model

Grey prediction is an important link in the grey system theory, it USES poor information, based on the original data processing to find the change rule of system, to generate strong regular data sequence, and then establish the corresponding differential equation model, every stage of the highway engineering construction cost is dynamic and uncertain, conforms to the characteristics of gray system, it can be seen as an independent of grey system [14].

Let's say the original number column:

\[
X^{(0)} = \{X^{(0)}(1), X^{(0)}(2), \ldots, X^{(0)}(n)\}
\]  

This is a set of grey quantities with incomplete information and great randomness. It is accumulated and generated to provide more useful information, and the result of m times accumulation is:

\[
X^{(m)}(k) = \sum_{i=1}^{k} X^{(m-1)}(k); (k = 1, 2, \ldots, n)
\]  

Make an accumulative generating operator, denoted as 1-AGO;

\[
X^{(1)}(k) = \sum_{i=1}^{k} X^{(i)}
\]

A cumulative generation sequence is obtained:

\[
X^{(1)} = \{X^{(1)}(1), X^{(1)}(2), \ldots, X^{(1)}(n)\}
\]

Generally, the data can be generated by one accumulation to show certain rules. If the regular data is not enough, the data can be accumulated for many times.
3. **Experiment**

In this paper, the index of project cost case is described in the form of slot, and the main design estimate index is selected, which makes it the characteristic attribute of case retrieval. Then, based on case-based reasoning system, the case of the database for primary retrieval and advanced retrieval. Then, the correlation coefficient matrix of characteristic factors is constructed by using factor analysis tool of SASS software, and the weight of characteristic factors is calculated, and then the similarity between cases is calculated by using weighted Euclidean distance method. Finally, the set of cases with the largest similarity with the target case is obtained. On the basis of full consideration of expert opinions, combined with the arithmetic average method, the unilateral estimated cost of the matching case of the target case is revised, and the reasonable unilateral estimated cost of the target case is obtained [15].

4. **Discussion**

4.1. **Check and Analyze Abnormal Data**

This paper proposes a method to solve the problem that the traditional anomaly data detection method cannot detect the anomaly data that is related to the list description. Here, it compares the method in this paper with the traditional anomaly data detection method based on distance. In this method, K=6.

| Evaluating indicator | Accuracy | recall |
|----------------------|----------|--------|
| Distance based approach | 0.9 | 0.65 |
| Method based on list classification | 0.92 | 0.86 |

The traditional method based on distance is mainly to calculate the difference between test data and training data. If the difference between test data and training data is less than a certain value and the number of training data is greater than the threshold value, then it is a normal data value, otherwise it is an exception. It can be seen from the experimental results that the list classification method proposed in this paper is more accurate than the traditional method in detecting the abnormal data of the comprehensive unit price, and has a better recognition effect for the abnormal data. The main reason is that the traditional method only considers the distribution law of the comprehensive unit price in the normal data, but fails to consider the relationship between the comprehensive unit price and the list description, which makes the traditional method more difficult to detect the abnormal data inconsistent with the list description and the comprehensive unit price, and the detection effect is not good.

4.2. **Engineer Access to Information Channel Investigation and Analysis**

![Figure 1. Acquisition Channels of Project Cost Information](image-url)
It is not difficult to see from Figure 1 that 50% of the cost workers prefer to obtain the most comprehensive and up-to-date project cost information from engineering cost information journals and websites. Although the publication cycle of journals and magazines is relatively long, the cost workers instinctively regard them as the main reference object, because compared with other channels, except for the shortcoming of the long cycle, the content published is more guaranteed, is authoritative, and the coverage of information is more comprehensive. The good news is that the younger generation of engineering cost workers will use the Internet more frequently, which creates better conditions for the construction cost of big data.

4.3. Software as a Service (Saas) Opportunity
SaaS applications adopt integrated multi-tenant development and usage patterns to provide high-performance software replication services. It can use a low-margin but fast-selling marketing model to help software developers achieve economies of scale and expand their market share.

Table 2. Comparison of Cloud Software and Traditional Software

| The characteristics of | Traditional packaged software | Software in SaaS mode |
|------------------------|------------------------------|-----------------------|
| Data storage location  | The local computer           | Local computer, cloud |
| Use/upgrade            | Download and install from your local computer | Software providers unified upgrade, network access services |
| Using the characteristics of | PC hardware and software requirements are high, slow running speed | Can be used in any computer software, running speed Take up less storage space on your local computer |
| Payment mode           | Purchase once, upgrade extra charge | Pay by usage or time, no upgrade worries |
| To prevent piracy      | Piracy is rampant and a lot of effort goes into it | Effective prevention of piracy |

In recent years, the market competition of engineering cost software is very cruel. There are not only legal copies in the market, but also various types of piracy. Engineering cost software development companies have spent labor and money trying to prevent piracy, but it hasn't worked. So it is putting the cart before the horse, with consumers footing the bill for all the costs of cracking down on piracy. If the whole set of software costs more than 10,000 yuan, and many expensive management software are attached, the price is several thousand yuan. But pirated software costs only a few hundred yuan and functions as well as real software.

As a new emerging technology, big data technology provides new possibilities for the development of software cost and also provides an important opportunity for the development of engineering cost industry. Big data technology organizes software on the programmer's server, and users get unified services through the Internet and different levels of access according to payment standards. Everyone has his own account. After the software is upgraded, all users can directly access the latest software services. This technology also limits software piracy to a certain extent, and ensures the security and authority of data.

4.4. Type of Information Obtained by Engineering Cost Engineer
Figure 2. The Types of Information the Respondents Get From The Internet

From Figure 2, we can infer information about labor price/material/machinery changes and industry trends. Policies and regulations are the most expected information. Further tracking interviewees, the important reason for relatively low demand for completed project case analysis is that the amount of comprehensive project information released by the cost center is too small, the project type is unified, and there is no in-depth analysis of the method to solve differences and the characteristics of the project. The information is simple and the guidance to practical work is limited. In practice, the project cost workers basically agree that the completed project summary information plays an important role in the new project cost work.

5. Conclusion

The construction of project cost information management platform is a complex, extensive and highly relevant work. This is a systems engineering. It requires knowledge of computer technology, project management, mathematical modeling, and so on. How to make full use of these knowledge, reasonable planning and cost information management platform is a huge problem.

(1) Carry out more in-depth research on big data technology, especially data mining, and apply more information technology means to the project cost information management platform to improve the efficiency of data extraction from the platform, mining its greater value and providing basis for decision makers.

(2) For the calculation of engineering cost investment estimation index and engineering cost index, the mathematical mining algorithm based on big data needs to be further optimized to make the algorithm more reasonable and provide a better basis for decision-making.

(3) Further organize and improve the standard format and interface of national project cost information; We will effectively implement national information interconnection and completely eliminate "information islands."

With the discovery and innovation of science and technology every year, the technology and mode of engineering cost information management will be constantly innovated under the new science and technology in the future, so as to realize the leapfrog development of engineering cost information management.

References

[1] Vikram Puri ; Sudan Jha ; Raghvendra Kumar ; Ishaani Priyadarshini ; Le Hoang Son ; Mohamed Abdel-Basset; Mohamed Elhoseny; Hoang Viet Long, A Hybrid Artificial
Intelligence and Internet of Things Model for Generation of Renewable Resource of Energy, IEEE Access, Vol 7, No. 1 (DOI: 10.1109/ACCESS.2019.2934228)

[2] Elsayed, W., Elhoseny, M., Sabbeh, S., & Riad, A. (2018). Self-maintenance model for wireless sensor networks. Computers & Electrical Engineering, 70, 799-812.

[3] Ibrahim M. El-Hasnony ; Sherif Barakat ; Mohamed Elhoseny ; Reham R. Mostafa, Improved Feature Selection Model for Big Data Analytics, IEEE Access, Vol 8, No 1, pp. 66989-67004, 2020 (DOI: 10.1109/ACCESS.2020.2986232)

[4] Fan Zhu. Application of engineering cost index in power grid project based on large data [J]. Journal of natural science of Harbin Normal University, 2017, 033 (001): 79-82

[5] Ying Tang, Mohamed Elhoseny, Computer network security evaluation simulation model based on neural network, Journal of Intelligent and Fuzzy Systems, May 2019 (DOI: 10.3233/JIFS-179121)

[6] Lee, Won Sang, E. J. Han, and S. Y. Sohn. "Predicting the pattern of technology convergence using big-data technology on large-scale triadic patents." Technological Forecasting and Social Change100 (2015):317-329.

[7] Liu, Zhifeng, et al. "Design and manufacturing model of customized hydrostatic bearing system based on cloud and big data technology." International Journal of Advanced Manufacturing Technology 84.1-4(2016):261-273.

[8] Lv, Rong Sheng, and C. H. Wang. "A Comparative Study between Sino-US and Japan-US Trade Frictions with Big Data Technology." Applied Mechanics & Materials 687-691(2014):4950-4954.

[9] Sun J. 124.Construction of Dynamic Management System of Engineering Cost Based on Exponential Model [J]. Revista de la Facultad de Ingenieria, 2017, 32(11):830-835.

[10] Xiaohui Yuan, Daniel Li, (2018) "Deepankar Mohapatra, Mohamed Elhoseny, Automatic Removal of Complex Shadows from Indoor Videos Using Transfer Learning and Dynamic Thresholding"," Computers and Electrical Engineering, 70, pp. 813-825.

[11] Abdurrahman, Lukman, et al. "Information Technology Value Engineering Model and Cost Efficiency in IT-Based Firms." IEEE Systems Journal (2017):1-12.

[12] Reflections on Construction of Engineering Cost Major for the Open University of China [J]. 2019, 023(001):23-26.

[13] Discussion on Countermeasures of Engineering Cost Management under Background of Big Data [J]. 2019, 046(005):89-90.

[14] Gaertner W, Schwettmann L. Burden sharing in deficit countries: a questionnaire-experimental investigation [J]. Series, 2017, 8(2):1-32.

[15] Liu Jian, Zhao Dan, fan Zhuzhu. Design and implementation of an online questionnaire survey platform [J]. New business week, 2019, 000 (013): p.288-288