

Available online at www.sciencedirect.com**ScienceDirect**

Procedia Engineering 123 (2015) 606 – 614

**Procedia
Engineering**www.elsevier.com/locate/procedia

Creative Construction Conference 2015 (CCC2015)

A system for tender price evaluation of construction project based on big data

Yongcheng Zhang^{a*}, Hanbin Luo^a, Yi He^a^a*Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan and 430074, China*

Abstract

Tender price evaluation of construction project is one of the most important works for the clients to control project cost in the bidding stage. However, the previously underutilization of project cost data made the tender price evaluation of new projects lack of effective evaluation criterion, which brings challenge to cost control. With the improvement of companies' information technology application and the advent of big data era, the project cost-related data can be completely and systematically recorded in real time, as well as fully utilized to support decision-making for construction project cost management. In this paper, a system for tender price evaluation of construction project based on big data is presented, aiming to use related technique of big data to analysis project cost data to give a reasonable cost range, which contributes to obtaining the evaluation criterion to support the tender price controls. The paper introduced the data sources, data extraction, data storage and data analysis of the system respectively. A case study is conducted in a metro station project to evaluate the system. The results show that the system based on big data is significant for tender price evaluation in construction project.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the organizing committee of the Creative Construction Conference 2015

Keywords: System; Bid price evaluation; Construction project; Big data

1. Introduction

It is of great significance for clients' construction projects success to carry out construction cost management during bidding stage. The clients want to make projects be built with limited economic resources; while the bidder is in pursuit of maximum economic efficiency [1, 2]. Therefore, during the construction project bidding phase, clients evaluate bids submitted by bidders, which we called bid evaluation—mainly considered on the basis of economic

* Corresponding author. Tel.: +86-18086482233; fax: +86-027-87557124.
E-mail address: cqzhyh@126.com

standards and technical standards.. Evaluation of the technical standards often can get perfect results with the support of experienced technical experts. However, as for evaluation on economic bid, due to the technical dependency of project cost, and market-related prices and other characteristics in economical bid, there exists some limitations in current economic evaluation, including mainly focusing on evaluating the total price and unbalanced quote as well as difficulty in assessing the price of engineering projects subheadings [3]. Kunhui Ye(2013)pointed that the factors in compiling tender prices were complicated and used a mix of research methods findings the key factors[4]. The method of limiting score within the bidding price was used in bid evaluation, which setting control range for bid prices which participate in rating (typically + 3% to -8%). The bidder whose bid price is over this range will lose qualification for bid evaluation and bid wining. This evaluation method adds to the risk of revealing base price. Reasonable low price bid evaluation method is used in construction projects in China now.it is also called multi-criteria method, which focus on the evaluation of the price. The arithmetic average of bidders' price and price from cost consulting company are the two main ways to get the tender price. The former method usually leads to tender price higher artificially under the condition of incomplete competition market for bidders' own interests and price "distortion" duo to together-conspired bidding. The latter method produces a time lag on the quota cost reference compiled by cost consulting company for previous cost data underused [5]. How to evaluate the reasonableness of the tender quoted price has become an important issue for successful bidding. Solution to this problem is to make full use of previous project cost data, conducting data mining analysis and providing a reference for tender evaluation of new construction projects. Before the advent of the era of big data, construction project cost data did not get deserved attention, which was reflected in that the project cost data was frequently narrowly thought as the project cost report data. Data storage formats and storage media are different, as well as serious data partitioning, etc. Ballesteros Pérez etc.(2014) built a new bidding model, as a practical tool, was used to improve bidders' strategy and increase their chance to win the contract before bidding based on experience and cost data .It called Smart Bid Model, but this model is mainly used in capped tendering[6].

From the perspective of big data, the concept of project cost data should be generalized, including engineering information data, construction cost data and technology scheme data etc. And the types of these data include documents, forms and other types. Considering from the amount and updates of data, project cost data has typical features of big data: large, emerging quickly, multi-sourced and heterogeneous [7, 8]. Big data technologies application in the Internet, medicine and other fields can provide reference for construction cost management. The most typical cases are that Google predicted the occurrence of flu and Wal-Mart's marketing of beer and diapers [9, 10]. At present, it is rare to see introduction of relevant storage and analysis technology of big data based on traditional relational database for conducting cost control during bidding phase in construction cost management field [11-13].

The main purpose of this article is to use big data technologies to study quoted price evaluation in cost control during construction bidding phase. The project cost data collection, extraction, storage, transmission and analysis are carried out and engineering subheadings reasonable price range is obtained in this paper, which is reasonable to evaluate the bidder's tender price, and the decision support system platform was developed.

2. Big data and related technology

2.1. Big data

With the all-round development of network and hardware facilities such as sensor and server, the era of big data has already come. Big Data technologies promote many firms to integrate their needs, creating unimaginable economic benefits, and achieving great social value with high commercial potential. Different industries use big data to generate great value and benefits, which shows unprecedented social potential not just the data itself. As for academia, Nature has launched a special issue of Big Data in 2008[14]. In February 2011, Science launched the special issue "Dealing with Data", mainly discussing the problems of big data scientific and explaining the importance of big data science [15].

As for big data, it has not yet come up with an accepted definition. The definition of 3V is more representative that considering big data should meet three characteristics: volume, variety and velocity. In addition, the definition of 4V is put forward by other person to trying to add a new feature on the basis of 3V. International Data Corporation (IDC) has agreed that the data should also have a quality of value [16]. The value of big data often exhibit characteristics of rarefaction. By the use of new processing mode, big data has more decision-making power and insight to optimize processes to achieve high growth rates and process vast amounts of diverse information assets. Big Data can be defined as a social process that collect large-scale data within a reasonable time and become more effective decision-making to help users. Table 1 is comparison between typical big data application.

Table 1 comparison between typical big data application

| Application | Examples | Number of users | Response time | Data scale | Reliability | Accuracy |
|--------------------|-------------------------|-----------------|---------------|------------|-------------|-----------|
| Internet of things | Sensor network | Large | Fast | TB | High | High |
| Mobile data | Mobile phone | Very large | Fast | Tb | High | High |
| Social network | Facebook | Very large | Fast | PB | High | High |
| Finance | High –frequency trading | Large | Very fast | GB | Very high | Very high |
| Scientific | Bioinformatics | Small | Slow | TB | Moderate | Very high |

2.2. Big Data technologies

Reflecting the great value of big data requires coordination of multiple technologies. File system provides the support to bottom storage. In order to manage the data, it needs to establish a database system on the file system. By establishing indexes and providing efficient data query functions, useful knowledge is extracted from big data by using analysis techniques in the database. Big data technology is a series of collection, storage, management, processing, analysis, and visualization technologies.

Cloud Computing is a large data base platform and supporting technology, involving a lot of technologies and algorithms. In 2006, Google proposed the concept of cloud computing firstly. The development of a variety of big data internal applications in Google are supported by a series of cloud computing open-source tools, which integrate to a platform called Hadoop[17]. Hadoop has become a complete ecosystem including the file system (HDFS), database (HBase, Cassandra), data processing (Map reduce) and other functional modules. Data processing involves series of algorithms, such as Genetic algorithms, neural networks, regression analysis, clustering analysis and association rule learning. To some extent, it can be said that Hadoop has become the real standard for processing big data tools.

In addition, the big data technology includes series algorithms, including genetic algorithms, neural networks, data mining, regression analysis, clustering analysis and association rule learning, classification analysis, data integration and fusion, machine learning etc. These algorithms can be evaluated and extended, which are involving most the algorithms of data mining. The cloud computing and the algorithms are the important support for big data mining.

3. Design for construction project bid price evaluation system based on big data

3.1. System framework

System framework provides an important basis for the system software development. Bid price evaluation system based on big data includes four levels: construction cost data collection layer, data acquisition and integration layer, data analysis layer and user-oriented application layer. Fig1 gives the system frame of project bid price evaluation.

(1) Data collection layer is the basic layer of the system, which mainly focuses on data collection related to project cost data. Basic information, construction technical scheme information and project cost information about the project are included. Structured, semi-structured and unstructured data are the main type of the information, through automatic collection mechanism, the completed project cost data are collected continuously into the system.

(2) After identifying the source of data, it is necessary to extract and integrate construction cost data, including automated data extraction, relationship building, data cleaning, and data quality assessment and so on. It aims to transform the raw data to those stored in certain rules and form big data center.

(3) Data analysis layer. Further data processing and analysis of business applications are continued by using mathematical and statistical methods, data mining and machine learning algorithms etc. it can provide the correlation law discovered from cost data for managers decision making. At the same time, data visualization techniques are considered to present the analysis results in most easily understood way.

(4) The top layer is the user layer, namely layer of human-computer interaction, including data input, data output, results of statistical analysis etc.

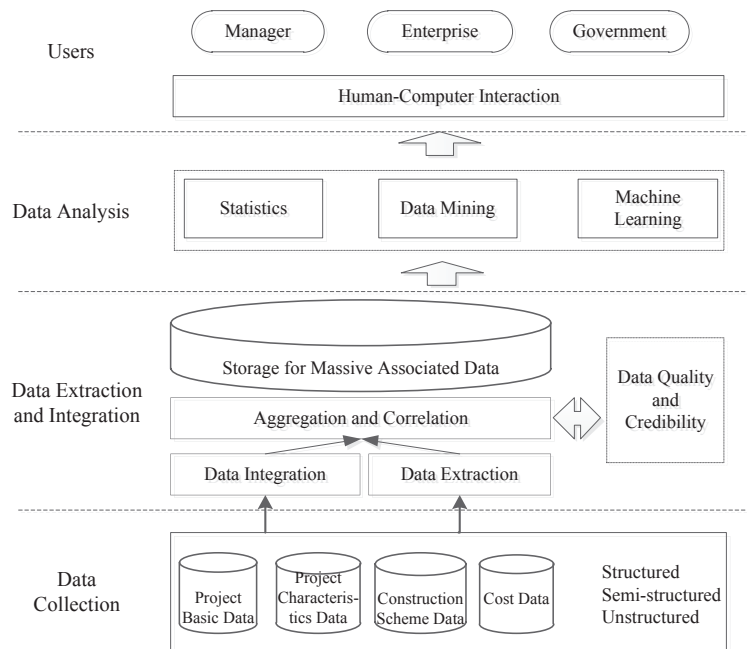


Fig. 1. System frame of construction project bid price evaluation

3.2. Research on big construction cost data collection system

(1) The source of the construction cost data

It is first step for the construction cost data analysis to collect the cost data. Currently, the main source of construction cost data derived from the following three aspects. The first aspect is the accumulated research cases of the team on practical projects. The second aspect data is from the cooperation agreement signed with the companies to dig the value from cost data. These companies included the clients, contractors and cost consulting firms. The third aspect is to sign a cooperation agreement with the government construction project management departments and analysis these construction cost data. The three aspects of data sources constitute the data collection source of construction cost big data together. In order to ensure the timeliness and accuracy of data sources, cost data from the relevant parties can be linked directly to the system and imported into the system after the signing a cooperation agreement with the parties. As a result, construction cost data can continue increasing.

(2) The content of the construction cost for big data collection

Basic project information, construction technical scheme information and project cost information are included in construction cost data acquisition. Basic project information collection includes project number, project name, project year, construction site, construction companies and the total duration of the basic project profiles. Projects characteristic information is based on different type of projects. Setting metro station project as an example: its feature information contains the project type, station type, station excavation depth, station planar shape, the surrounding environment, soil conditions, hydrology, and construction method and so on. Construction technical scheme information is mainly to complete the construction project process engineering methods, machinery, materials and other information. Project cost information mainly contains the total price, the unit construction cost, cost of division component project, the main quantity and price. The collected data will be integrated and merged, then form data cube and dimension on integrated circumstances. Fig2 shows a local project cost cube.

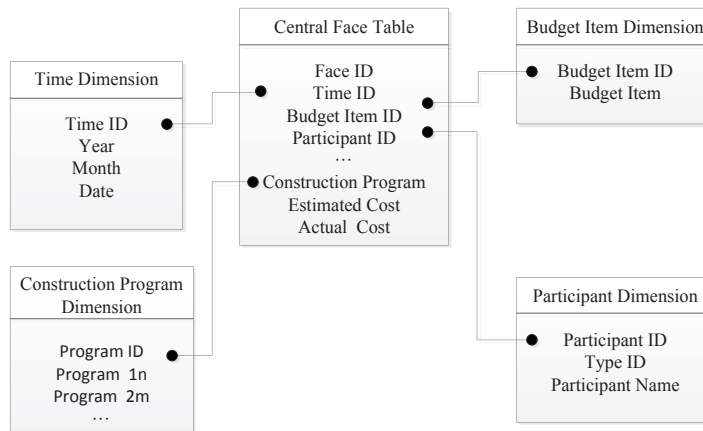


Fig. 2. Data cube and dimension design

(3) Methods of construction cost data extraction

At present, construction cost data in different structural forms are stored in different databases, including formatted report data stored in the system, text data stored on the computer, tabular data, picture data and various projects cost data stored in paper. Different types of data demand different data processing method. The common way is to transform the unstructured data into a semi-structured data or structured data. The semi-structured data has been able to meet data processing demand stored in big data centre. On the contrary, using processing method including manual sorting, automatic processing of artificial intelligence algorithms, such as automatic text scanning, text recognition, image recognition, text mining methods, transform into structured data, which may use these proper methods to extraction project cost data.

3.3. Construction Cost Data Security Analysis

Construction cost data is commercial secrets of enterprises, which is the core of enterprise data. Therefore, the safety management of project cost data is extremely important. In order to strengthen the safety management of project cost data, it mainly focuses on the data encryption processing, management level of permissions settings and accountability mechanisms. Firstly, project cost data needs to be anonymized by using an encryption algorithm. Secondly, access permissions which differ from varies levels are set for all the relevant parties with access to the data and the responsible person. Finally, accountability mechanism is established for relevant parties and responsible persons. Entities and individuals who brought negative influence to project cost or caused loss of data leakage by personal reasons should be punished. Through the design of the three levels, the project cost data security can be ensured.

3.4 Methods for Construction Cost Data

Cost management is supported by cost data analysed by different algorithms, which includes time series analysis, clustering analysis, statistical algorithms, case-based reasoning and other artificial intelligence algorithms. Here are a few algorithms.

Case-based reasoning method is one of the commonly used correlation analysis algorithms. A popular explanation is to find a new solution of practical problems. Firstly, similar problems are found in the library experience from similar problems in the past solution, and use it as a starting point for solving practical problems solution. The solution is obtained by adaptations to new problems. Fig3 is the process of case-based reasoning[18].

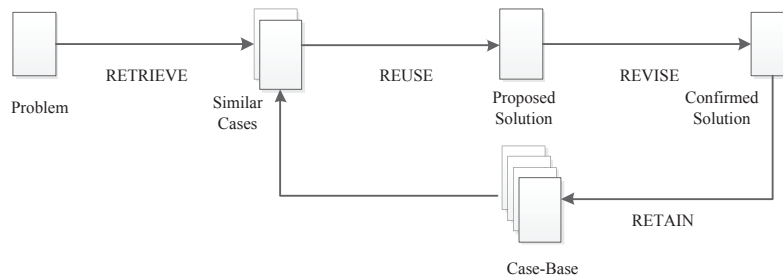


Fig. 3. The process of case-based reasoning

The steps of the CBR as follows:

Step 1: Cases Representation. The case is a structured description of some solutions for solving problems in the application. The description method determines the case index, case retrieval and case storage methods etc. All knowledge representation method can be used in the case, and different methods of knowledge representation are chosen according to the features of application domain. Case knowledge representation methods mainly include two categories, namely the logical method and the method based on logic.

Step 2: Cases Retrieve. Case retrieval aims to search and compute the most similar cases with the current problem. Recent adjacent algorithm, the decision tree method and knowledge guidance method are the common case retrieval methods.

Step 3: Cases Revise. It is hard to find the same problems with the new problem exactly in case-based reasoning, so it is often necessary to modify reuse cases solutions to adapt to the new problems. Many methods can be used in the case revision, such as Inductive learning method and satisfaction algorithm.

Step4: Cases Re-learn. The ultimate goal of case study is to store the new methods, solutions and event information together, which make the system acquire the ability to learn. Also, it will increase the knowledge and experience of the system.

Time series analysis is a statistical method for dynamic data processing. The time series is a set of chronological sequence of numbers. Time series analyses use this set and the application of mathematical statistics to predict the future development of things. The method is easy to grasp, but the accuracy is poor. Generally it's only suitable for the short-term prediction. The method can be used for decision-making of the project sub item price reasonable interval calculation and analysis. Take the single-element regression model as an example, suppose there are n data points $\{(x_i, y_i), i = 1 \dots n\}$. The function that describes x and y is:

$$y_i = a + bx_i \quad (1)$$

The goal is to find the equation of the straight line:

$$y = a + bx \quad (2)$$

This would provide a “best” fit for the data points. Here the "best" will be understood as in the least-squares approach: a line that minimizes the sum of squared residuals of the linear regression model. In other words, a (the y-intercept) and b (the slope) solve the minimization problem. Then Calculate a and b . single-element regression model is built and used to predict the variable.

4. Case study

It has been developed that bid price evaluation system prototype of construction project based on collecting a large number of subway station project cost data. The system platform which is called metro project management platform has been preliminary applied in some subway station projects, and the application effect is rather good.

When potential contractors get the quantities of a new subway station project, they will carry out project bidding. Then the potential contractors will make a bidding quotation after the overview of the construction project. The most important information includes the type of the station(X1), excavation depth(X2), flat shape(X3), hydrological conditions(X4), soil conditions(X5) and surrounding environment(X6). These variable' values are different with different project and these variables mainly for discrete variable. The values of discrete variable have been confirmed by fuzzy processing methods. For example, X3=(X31, X32), X31= standard shape of the station, X32= non-standard shape of the station, X6=(X61, X62, X63), X61=Very complex, X62=Complex, X63=General complex. X61 mains that surrounding buildings situation, pipeline situation and traffic condition are very complex.

The main bidding quotation includes bill of quantities quote and quotation analysis of comprehensive unit price. These two aspects will be investigated by tenderer, especially the quotation analysis of comprehensive unit price. The bid price evaluation system can analyse construction project characteristic information and the cost data of the project that has been finished. It uses the data association (PDA) algorithm, statistic and time series method to get a reasonable price range for the bidding quotation which is suitable for the cases that has same characteristics in the same period, and this bidding quotation range regard as a reference standard for the construction bid assessment. It likes the laboratory sheets of the hospital which including both experimental value and reference value, marking some test value weather is high or low to give patient a reference.

When the bidders log in the system and fill in system with construction project data while the information format is required strictly, system will evaluate quotation automatically and alarm the tenderer which project unit price is high and low in visual way. The experts will censored this unreasonable project unit price by this warning and asking the bidder to explain in order to avoid the unreasonable and unbalanced quotation, etc. Fig4 and Fig5 are the interfaces of the system.

Fig. 4. The input interface of system

| 一级工程单元 | 二级工程单元 | 单位 | 合理区间价 (万元/) | 投标报价 (万元/) | 价格评价 |
|--------------------|---------|----|----------------|------------|------|
| 主体围护结构工程 | 地下连续墙 | m3 | [0.320, 0.350] | 0.351 | • |
| 主体土(石)方工程 | 机械挖土方 | m3 | [0.010, 0.015] | 0.012 | |
| 主体结构垫层工程 | 混凝土垫层 | m3 | [0.110, 0.130] | 0.122 | |
| 主体结构防水工程 | 卷材防水 | m2 | [0.010, 0.014] | 0.012 | |
| 主体结构钢筋混凝土工程 | 钢筋混凝土结构 | m3 | [0.170, 0.200] | 0.19 | |
| Construction Items | 地下连续墙 | m3 | [0.310, 0.340] | 0.352 | • |
| | 机械挖土方 | m3 | [0.010, 0.015] | 0.012 | |
| 出入口结构垫层工程 | 混凝土垫层 | m3 | [0.110, 0.130] | 0.123 | |

Fig. 5. The result of system analysis

It can be seen from the above table, calculating a reasonable range for the bidding quotation, when the bidders submit bid price, the system can make a compare to tender with the reasonable range automatically. It shows the bid price range whether beyond the control or in the interval of range in a visually way, which can give tenders a useful reference to revalue the bid price. With construction cost data increasing rapidly, the price range calculated by system will be more accurate.

5. Conclusion

With the advent of the era of big data, it will produce huge impetus to construction cost management with the steady accumulation of cost data. In this paper, we have studied the problem of evaluation on tender price. We have presented a system called SDPE that uses algorithms to mine cost data based on the platform. The framework and prototype implementation described in this paper illustrates the data collection, data security and method for data processing. It gives the feasibility of system application. Through the metro construction cost data analysis, it has shown the result of the case provided by the system. The decision support which is made by the system is reasonable and dynamic in metro construction project case. But how to refine and revise the algorithms to improve the result accuracy and the degree of automation of data processing, which will support the clients or tenders to evaluate the bidding quotation usefully.

Acknowledgements

We would like to thank Wuhan metro group co., LTD for help with data collection at the construction site. We would additionally like to thank Shengyu Guo for many interesting discussions.

References

- [1] OECD (Organization for Economic Co-operation and Development) 2007. Bribery in procurement, methods, actors and counter-measures, Paris.
- [2] OECD (Organization for Economic Co-operation and Development) 2009. Guidelines for fighting bid rigging in public procurement, Paris.
- [3] Yong Gang Zhang et al., 2014, Applied Mechanics and Materials, 501-504, 2710
- [4] Ye, K., Li, B., and Shen, L. Key Factors Considered in Compiling Tender Prices for China's Public Works Projects. J. Manage. Eng. 29(3), (2013), 206–215.
- [5] Yu, W., Wang, K., and Wang, M. Pricing Strategy for Best Value Tender. J. Constr. Eng. Manage., 139(6), (2013), 675–684.
- [6] Ballesteros-Pérez, P., et al.. Estimating future bidding performance of competitor bidders in capped tenders. Journal of Civil Engineering and Management 20(5): (2014)702-713.
- [7] L. Kart, High-Tech Tuesday Webinar: Big Data Opportunities in Vertical Industries, Gartner, Webinar, ID No. G00236293, Aug. 8, 2012.
- [8] Barwick H. The " four Vs" of Big Data. Implementing Information Infrastructure Symposium. [2012-10-02]. http://www.computerworld.com.au/article/396198/iis_four_vs_big_data.

- [9] Du, D., et al.. Survey on the Applications of Big Data in Chinese Real Estate Enterprise. *Procedia Computer Science* 30(0): (2014), 24-33.
- [10] Shin, D.-H. and M. J. Choi. Ecological views of big data: Perspectives and issues. *Telematics and Informatics* 32(2): (2015), 311-320.
- [11] Jeffrey Dean. MapReduce: Simplified Data Processing on Large Clusters, *Communications of the ACM - 50th anniversary issue: 1958 - 2008 Volume 51 Issue 1, 2008*, 51(1): 107-113, ACM New York, NY, USA.
- [12] Shen DR, Yu G, Wang XT, Nie TZ, Kou Y. Survey on No SQL for management of big data. *Ruan Jian Xue Bao/Journal of Software*, 2013,24(8):1786-1803 (in Chinese).
- [13] Ni MX, Luo WM. Technology revolution in the age of data exploding. *CCF Communications*, 2011, 7(7):12□□0 (in Chinese with English abstract).
- [14] Howe D, Costanzo M, and Fey P, et al. Big data: The future of biocuration. *Nature*, 2008, 455(7209): 47-50.
- [15] Bryant R, Katz R H, Lazowska E D. Big-data computing: creating revolutionary breakthroughs in commerce, science and society. 2008.
- [16] L. Kart, "High-Tech Tuesday Webinar: Big Data Opportunities in Vertical Industries," Gartner, Webinar, ID No. G00236293, Aug. 8, 2012.
- [17] Hadoop [EB/OL].[2012-10-02].<http://hadoop.apache.org/index.html>.
- [18] Aamodt A, Plaza E. Case-based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *Artificial Intelligence Communications*, 7(1) (1994), 39–59.