

Feature Vector Extraction Algorithm Based on Big Data in Engineering Quality

Fan Zhang^{1,*}, Yuhua Yang²

¹ Beijing Jiaotong University, Infrastructure and Planning Department, Beijing, China

² Beijing University of Chemical Technology, Logistics Support Department, Beijing, China

Abstract. With the advent of the information age, the network has played a role in promoting the development of various industries. As a construction enterprise, it is necessary to integrate new technologies to achieve scientific management and construction. Engineering quality control management is the lifeblood of determining the merits of a project, which is the life of construction engineering and the key to winning users, developing enterprises and occupying the market. Based on the current problems encountered in the construction quality control of China's construction industry, a comprehensive evaluation system based on network big data in the paper is proposed, and the data of method in the engineering quality risk eigenvector model are extracted, processed and analyzed. In the paper, the engineering quality risk feature vector model is designed. The genetic algorithm is used to solve the function as a nonlinear optimization problem. The vector feature extraction algorithm is optimized. The data projection vector of the feature vector data processing is used to define the quality influencing factor evaluation value. The quality of the project is analyzed. After testing and analyzing the model, it proves that the data based on big data extraction is more objective and reasonable from engineering quality risk analysis, risk generation mechanism and optimization risk indicators, which provides reference for China's construction engineering enterprises.

1 Introduction

Engineering quality control management is the lifeblood of determining the merits of a project, which is the life of construction engineering and the key to winning users, developing enterprises and occupying the market. Since quality control management runs through the entire construction process, it plays a key role in the design, preparation, specific operation, or in the final stage. Therefore, no matter from the national, social level, or from the enterprise and individual level, new requirements and new goals for advancing with the times have been put forward for the quality control management of construction projects. This requires enterprises to change the extensive quality management work mode under the new situation, and to explore the new methods and new ideas of construction engineering quality control management while pursuing the shortening of construction time and saving construction costs to maximize profits. Fine quality management and high-quality construction products have won a place in the fierce market competition. To this end, the integration of big data technology in the paper will be explored as a new information technology into quality control management, which achieve the transformation of enterprises into quality intensive growth, and effectively improve the safety and reliability of construction engineering products to meet social needs [1-3].

The research content of the paper is based on the current problems encountered in the construction quality control of China's construction industry. It is aimed at the event management mode in traditional engineering, that is, in the construction process, the quality problems in the progress of construction is checked; or the engineering quality problems that have already occurred are fed back to reach the control of project quality. Such a management method is relatively lacking in quality supervision and regulation, and the management means is single, and the management effect cannot be clarified. Moreover, such management requires the engineering enterprise to revise the project quality management plan by itself, mainly based on the fact that the centralized analysis method based on its own experience is also prone to subjective deviation problem. Therefore, based on the network big data, a comprehensive evaluation system of engineering quality in the paper is established, and the data in the engineering quality risk eigenvector model is extracted, processed and analyzed. The vector feature extraction algorithm is optimized, the data projection vector of the feature vector data processing defines the quality influencing factor evaluation value to expresses the data analysis method of the engineering quality [4].

*fanz@bjtu.edu.cn

2 Engineering quality feature vector extraction based on big data

In big data technology, the informatization supervision platform is mainly composed of servers and browsers, which provides accurate data support for managers. The main functions of the informatization supervision platform include that, (1) Managing the project library. The project library mainly includes the basic information of the relevant project, the construction situation of the project, the time of project construction and the operating cost of the project. In the process of managing the above data and information, in order to ensure management efficiency, the project supervision platform established after integration into the big data platform is shown in Figure 1. In this platform, it can be classified according to the type of project, or can be classified according to the construction time of the project. In addition, relevant personnel can inquire about the project by logging in to the supervision platform, which include engineering application, formalities, project reporting and engineering rectification. (2) Quality supervision of the project. The main contents of quality supervision include the quality application of the project, the quality inspection during the construction process and the quality management after the construction and the regular quality return visit. Relevant personnel input the corresponding code of each management project into the supervision platform by means of logging in to the information platform, and query and process the required information. All project processes in the management project will be included in the supervision platform, which include project quality notice, supervision notice, special inspection notice, quality spot check notice, project rework notice and stoppage notice. In addition, after the completion of the project, the platform will conduct regular visits and quality feedback. Relevant feedback results will be released to the platform's information processing system as soon as possible. After a series of verification and tests, the system will post the final test results to the regulatory platform to provide relevant query services [5].

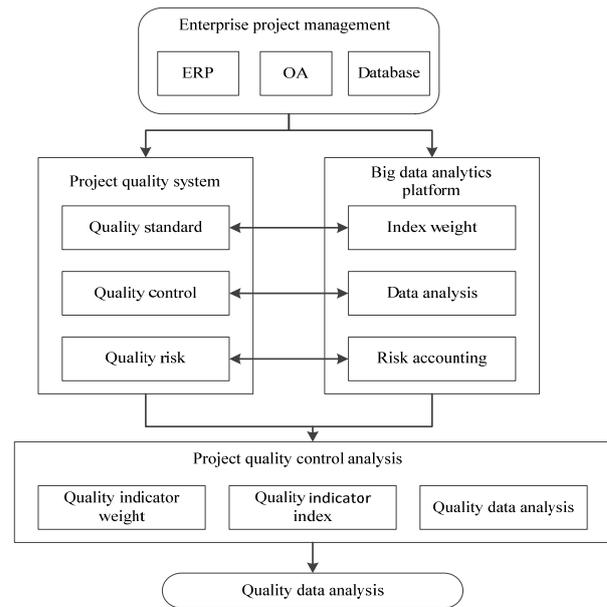


Figure 1. Engineering enterprise big data project management platform.

The project quality control platform based on big data is to combine the frontier management theory, technical measures and supervision methods to optimize the design of traditional engineering quality objectives. The system platform first collects data on all aspects of construction site materials, machinery, personnel, methods and environment, and then stores and analyzes the data in each server in the data processing layer, and realizes quality monitoring and early warning in the application layer of the data. Finally, at the sharing layer, the project leader is provided with decision support analysis through the quality monitoring application platform [6-8].

In the evaluation model of engineering quality influencing factors, the effects of different environments are not consistent according to the quality influencing factors in different projects, so the proportion weight analysis is adopted to weigh all kinds of influencing factors. The main content of influencing factors of engineering quality is shown in Table 1.

Table 1 Factors of Engineering Quality

Number	Progress Risk Factor	Source of Risk
01	Shortage of funds	Internal factors
02	Quality defect rework	Internal factors
03	Design drawing delivery delay	Internal factors
04	Material shortage	Internal factors
05	Land acquisition and demolition delay	Internal factors
06	Bad weather	Force majeure
07	Owner delay payment	Internal factors
08	Force majeure such as natural disasters	Force majeure
09	Design changes during construction	Internal factors
10	Site handover delay	Internal factors
11	Insufficient experience in design units	Internal factors

12	The contractor's bid price is lower	Internal factors
13	Material price change	External factors
14	Limited labor level	Internal factors

The construction of the quality control platform of engineering projects under the background of big data needs to pay attention to the following three aspects. (1) Effective monitoring of materials to achieve tracking of the entire process. Specifically, when building materials are assigned their unique codes from procurement, entry, inspection, acceptance, and use, real-time collection of building material quality data can be achieved. (2) Recording management of the construction project main body and first-line construction personnel. Specifically, it is to carry out electronic code numbering for construction companies and construction personnel, and then record each numbered activity to implement comprehensive real-time supervision. (3) Recording management of the

supervision company and subcontracting company. Enterprises within the scope of registration should establish corresponding identification codes according to the principle of territoriality and supervise and record their behaviors.

In the project quality management, the integration of big data platform can comprehensively utilize the network big data information, establish a scientific and complete, analyze the pre-project existence and potential quality risks, and take relevant risk reduction measures in advance to achieve the purpose of avoiding risks. With the rapid development of big data in modern society, the use of big data manage engineering quality risks and combine with the relative quality risk management model, which can effectively reduce the cost of risk management, improve the efficiency of quality management, and reduce Engineering quality risk. For example, the feature vectors of piling operations in construction projects are extracted, as shown in Table 2.

Table 2 Quality Index of Construction Projects Based on Large Data Feature Vector Extraction

Number	Work Item	Item	Project Process	Explanation
GC3801	Main process of pile foundation process GC38011	Shuttle pile GC380111	The pile material is sent from the pile driver to the pile driver or the pile driver to ensure the complete completion of the pile material.	Piles are overweight and super-large materials. They must be operated by more than 2 people. They should be directed by professionals and assisted by auxiliary workers.
		Piling GC380112	The operator is responsible for operating the pile driver to drive the pile into the pile position.	This process is a special machine operation, the operator must be qualified for construction, and other personnel must not operate.
		Pile GC380113	The interceptor and measurement work of the pile is carried out by the operator or the reserve.	This process is a special machine operation, the operator must be qualified for construction, and other personnel must not operate.
	Pile-based process auxiliary process GC38012	Welding GC380121	It is the responsibility of the construction personnel to complete the required welding work and ensure the welding quality.	It must be supported by qualified employees of welding qualification, and the welding quality meets the requirements.
		Earthwork GC380122	The earthwork is completed by the auxiliary personnel as required	The operator needs to have certain earthwork experience and complete it as required.
		Mapping GC380123	Calibration of pile position and piling accuracy using precision instruments	Surveying and mapping personnel must have certain basic mapping and data statistics capabilities to ensure the accuracy of surveying and mapping data.
		Handling GC380124	Carrying out the task of carrying out the work of the pile and the ingredients	The porter must first ensure the physical fitness, have the experience of heavy material preservation, and perform the work in accordance with the instructions.
	Pile foundation safety process AQ-ZJ-01	Safety supervision	The supervisors shall carry out safety management work throughout the project, assist the responsible person in supervision and management, and prevent safety accidents from happening.	According to the detailed management of the pile foundation process, the safety supervision of the whole process is carried out to avoid any safety problems.

In Table 2, according to the process safety index extracted from the big data of construction safety and quality, the main process, auxiliary process and safety process of pile foundation process are explained. The explanation of these processes can clarify the construction process and the matters needing attention of personnel operation in the construction process, and play a role in the standardization of construction quality management. In the past traditional management, construction managers rely on experience to direct on-site operations, which is easy to produce construction errors, and is not conducive

to the development of safe construction, quality can't be guaranteed. Procedure project description extracted from large data is more scientific and standardized than empirical management, which guarantees the quality of construction and safe operation.

3 Establishment of Engineering Quality Risk Characteristic Vector Model

The enterprise engineering management platform based on big data can get more data support and scientific

management experience in quality management of the project. The engineering quality risk feature extraction studied in the paper is based on big data. The data projection vector processed according to the feature vector data is defined by the equation for evaluating the quality influencing factors, which is as shown in the (1).

$$K = L \times P \quad (1)$$

Where L represents the probability of occurrence of the influence factor of the quality condition; P represents the degree of influence of the influence factor of the quality condition.

All quality state evaluation values correspond to A1 indicating that the quality state is excellent, A2 indicating that the quality state is qualified, and A3 indicating that the quality state is unqualified. For example, if the influence degree of a certain quality influencing factor is 2, and the probability of occurrence of the influencing factor is 3, the corresponding interval of the influencing factor is A2, and the quality state is more qualified. In the evaluation of the factors affecting the quality of the project, L is the ordinate and P is the abscissa, and the quality impact factor evaluation matrix is established, as shown in Figure 2.

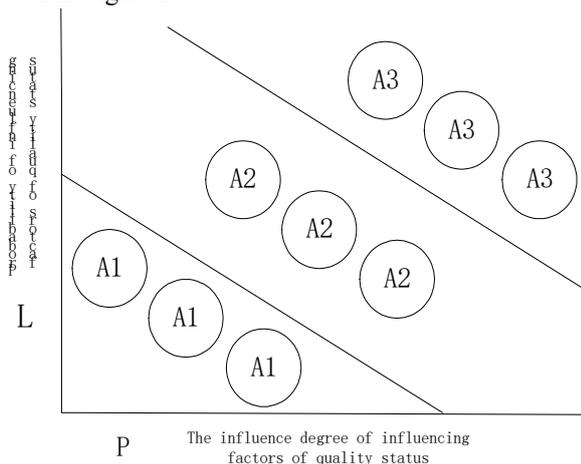


Figure 2. Quality impact factor evaluation matrix.

3.1 Data Preprocessing

In the data pretreatment stage, systematic construction and revision are needed for the quality standards and judgments of Engineering projects, which requires more improvement in the preliminary work, in order to realize the eigenvector based on engineering quality. Specific implementation methods are adopted, e.g., the establishment of engineering quality index system. The project quality index system studied in this paper is extracted from large data. According to the evaluation matrix of quality influencing factors, the project quality index model corresponding to the project can be established. According to the eigenvectors of the model, the project quality index model can be obtained from large data.

Take the content of this paper as an example, the establishment of project indicators will be based on project quality standards, each independent quality standard category will be planned as catalogue category,

and more detailed quality indicators will be a1, a2. For statistical indicator sample set data, the larger is the better, or the smaller is the better, and the data needs to be processed uniformly. The indicator sample set is

$$a(x, y) = \frac{\sum_{x=1}^m a^x - a^y}{\sum_{y=1}^n a^y - a^x} \quad (x = 1, 2, \dots, m; \quad y = 1, 2, \dots, n) \quad (2)$$

Where $\alpha(x, y)$ represents the sample of safety indicators collected; a represents the category of quality indicators, x is the number of samples, and y is the number of indicators.

$$\alpha(x, y) = \frac{\alpha(x, y) - \alpha_{\min}(y)}{\alpha_{\max}(y) - \alpha_{\min}(y)} \quad (3)$$

For the smaller sample is the better, the indicator is

$$\alpha(x, y) = \frac{\alpha_{\max}(y) - \alpha(x, y)}{\alpha_{\max}(y) - \alpha_{\min}(y)} \quad (4)$$

Where $\alpha_{\max}(y)$ is the maximum value of the y indicator; $\alpha_{\min}(y)$ is the minimum value of the y indicator.

3.2 Function Construction

Supposing that $\lambda(y)$ be the projection vector, and the projection of the sample x in this direction is

$$z(x) = \sum_{y=1}^n \lambda(y) \quad \lambda(x, y) \quad (5)$$

Where, the sub-features of $\|\lambda\| = 1$ and $z(x)$ should be satisfied: the projection points are as dense as possible, and spread as much as possible as a whole. The degree of dispersion between classes is represented by the standard deviation of $z(x)$.

$$S_z = \sqrt{\frac{\sum_{x=1}^m [z(x) - E(z)]^2}{y-1}} \quad (6)$$

Intra-class density is represented by the function D_z as

$$D_z = \sum_{x=1}^m \sum_{y=1}^n [R - r(x, y)] u [R - r(x, y)] \quad (7)$$

Where $E(z)$ is the mean value of the projection value $z(x)$; R is the local density window value, generally $0.1S_z$; $r(x, y)$ is the distance $r(x, y) = |z(x) - z(y)|$ between the sample projection values $z(x)$ and $z(y)$; $U(t)$ is the unit-order function, when $t \geq 0$, its value Take 1, then $t < 0$, its

value takes 0. From the above, the configurable indicator function is

$$Q(\lambda) = S_z D_z \quad (8)$$

The index function $Q(\lambda)$ will change with the change of the projection vector λ . The projection pursuit method is to find the corresponding projection vector λ , so that the maximum value of $Q(\lambda)$ is obtained, which the objective function is

$$\max Q(\lambda) = S_z D_z \quad (9)$$

The constraint is

$$\sum_{y=1}^m \lambda^2(y) = 1 \quad (10)$$

3.3 Function Solution

The constructed function is a nonlinear optimization problem. The traditional method is difficult to solve. This paper uses genetic algorithm to solve the problem. However, it is difficult to directly solve the problem. The problem of the maximal function with constraints is transformed, which the objective function is

$$\min Q(\lambda) = \frac{1}{S_z D_z} + m \left[1 - \sum_{y=1}^m \lambda^2(y) \right]^2 \quad (11)$$

In the (11), we should ensure that the square of the projection vector is 1, m can be set to a maximum value, and $m=10000$ in the paper. The genetic algorithm mainly includes the following steps (1) generating an initial population. Randomly generate data sets to form an initial group. Each data set contains 23 variables, the projection variables. (2) Exchanging. Each of the two parents selected by the exchange probability generates a new individual by exchanging the different partial genes. (3) Variations. A certain number of individuals are randomly selected in the group, and the value of a certain data gene is changed with a certain probability for the selected individual. Like the biological world, the probability of

occurrence of the mutation is very low. (4) Fitness. Substituting the generated n data sets into the (11), the smaller the target value function value, the better the adaptability, and the larger the fitness value, the larger the next exchange probability. (5) Aborted. The genetic algebra $mgen$ is given, the algorithm stops when iterating to this algebra.

3.4 Classification and Optimal Arrangement

By genetic algorithm, the optimal projection variable ax can be obtained and substituted into (4) to obtain the projection vector Z_x which can best reflect the characteristics of high-dimensional data, and the Z_x is arranged in order from small to large.

4 Model simulation test and data analysis

The engineering quality algorithm studied in the paper uses model naturalization simulation to test in the feature vector extraction process. Firstly, the characteristics of the project are extracted, and some of the decisive factors of the project quality are taken as test references. Then, the naturalization of the feature vector model is based on the network big data, and the obtained model data is brought into the test. The weighting parameters of the index are calculated in the engineering quality algorithm, and the data weighted by these weighting parameters are used as the analysis data. Finally, the effects of different weights of the conditional factors on the quality of the project are studied at each stage of the project.

The engineering quality inspection index is collected and statistically analyzed, and normalized for processing. Each value in the projection vector is within the range. The software MATLAB7.1 is used, the projection pursuit model is solved by the optimized algorithm. The calculation results are shown in the Table 3, and the projection value is visualized, which as shown in Figure 3. According to the calculation results, the best projection vectors are in the range, and the sum of squares is 1, and the calculation result satisfies the constraint condition.

Table 3. Projection Vector of Engineering Project Quality Index

Project Progress	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Proje-ct1	0.7913294	0.2998938	0.5934193	0.6928328	0.8921392	0.2932414	0.6918314	0.9128218	0.5121921	0.6917254
Proje-ct2	0.6197138	0.7173929	0.3621829	0.6938138	0.613613	0.5922972	0.6941921	0.693811	0.6519341	0.5153883
Proje-ct3	0.5192961	0.5573128	0.6932713	0.7184713	0.6391393	0.5183228	0.5681403	0.1381224	0.7512582	0.2861383

Proje-ct4	0.7132384	0.658822	0.6927192	0.3592112	0.6261926	0.6181258	0.8235729	0.6283791	0.4148273	0.5183828
Proje-ct5	0.284432	0.2682908	0.7283273	0.4971127	0.8382029	0.3138344	0.7623924	0.4181624	0.4927218	0.6101826

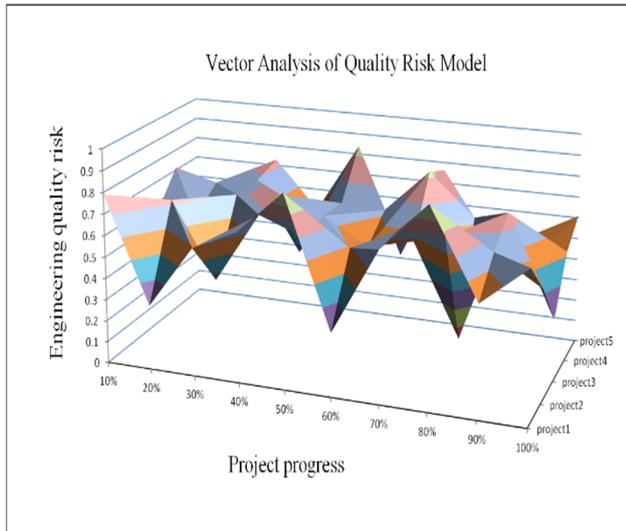


Figure 3. Visualization analysis of quality risk model data.

After the above test verification process, the weight analysis of the engineering quality risk model studied in the paper is determined by the progress and effect of the indicator in the model projection. In the data analysis, the index parameters of the project quality risk will not decrease with the completion of the project schedule, which includes the risk difference of the overall project quality during construction and after completion, and the standardized quality certification. The decisive indicator vector such as system and certification system is regarded as the basis of engineering quality, and its data will not cause large fluctuations due to the impact of construction quality; but the quality of the project is based on non-master factors such as design, materials, new process and engineering outsourcing to form the impact of the main quality risk indicators.

The engineering quality feature extraction method studied in the paper is built on the network big data, which avoids the long process of the quality data database relying on the accumulated data analysis after the long-term data collection and early warning of the traditional data analysis model. Moreover, the data based on big data extraction is more objective and reasonable from the engineering quality risk analysis, risk generation mechanism and optimization risk indicators, and can also reduce the process of index weight adjustment caused by subjectivity.

5 Conclusions

In the process of engineering quality control, construction projects will continuously generate a large number of progress, quality and investment cost data. Through the big data platform, the data can be correlated and analyzed, and the basic objectives of the project can be monitored

and adjusted in real time. In the aspect of quality control, the control platform based on big data can replace the post-inspection method of traditional methods, warn the abnormal and uncontrolled processes in the construction process, store and analyze the engineering quality data of each server in the data processing layer, realize the automatic monitoring of engineering quality in the data application layer, and use the quality control platform. Data analysis technologies, such as model training, algorithm prediction, effect statistics and online analytical processing, visualize all kinds of cost data and provide decision support analysis for quality management.

From the perspective of big data, it is an inevitable trend of the Internet era and the development of information society to realize the management and control of engineering quality risks and the participation of big data in the whole process. The paper integrates engineering projects into big data under the big data platform, optimizes the index vector extraction algorithm, and verifies the effect of the effect on the application of engineering enterprises in project management, and provides reference for China's construction engineering enterprises.

References

1. Q. M. Tenailleau, Frédéric Mauny, D. Joly, S. Franois and N. Bernard, (2015) Air pollution in moderately polluted urban areas: How does the definition of “neighborhood” impact exposure assessment, *Environmental Pollution*, vol. 206, 437-448.
2. L. Feng and W. J. Liao, (2016) Legislation, plans, and policies for prevention and control of air pollution in China: achievements, challenges, and improvements, *Journal of Cleaner Production*, vol. 112, Part 2, 1579-1558.
3. L. Y. Zhang, J. J. Du and S. S. Zhang, (2014) Solution to the Time-Cost-Quality Trade-off Problem in Construction Projects Based on Immune Genetic Particle Swarm Optimization, *Journal of Management in Engineering*, vol. 30, no. 2, 163-172.
4. R. Osei-Kyei and A. P. C. Chan, (2015) Review of studies on the Critical Success Factors for Public-Private Partnership (PPP) projects from 1990 to 2013, *International Journal of Project Management*, vol. 33, no. 6, 1335-1346.
5. F. A. Mir and A. H. Pinnington, (2014) Exploring the value of project management: Linking Project Management Performance and Project Success, *International Journal of Project Management*, vol. 32, no. 2, 202-217.
6. M. D. Assunção, R. N. Calheiros, et al, (2015) Big Data computing and clouds: Trends and future

directions, *Journal of Parallel and Distributed Computing*, vol. 79-80, 3-15.

7. O. Taylan, M. R. Kabli, C. Porcel, et al, (2017)Contractor Selection for Construction Projects Using Consensus Tools and Big Data, *International Journal of Fuzzy Systems*, vol. 20, no. 4, 1267-1281.
8. M. K. Pekerici, B. Akinci, I. Karaesmen, (2003)Modeling Information Dependencies in Construction Project Network Organizations, in Proc. 4th Joint International Symposium on Information Technology in Civil Engineering, Nashville, 513-525.