Research Article

Research on Optimization of Big Data Construction Engineering Quality Management Based on RNN-LSTM

Daopeng Wang [ID],[1] Jifei Fan [ID],[1,2] Hanliang Fu [ID],[2] and Bing Zhang[2]

[1]*School of Civil Engineering, Lanzhou University of Technology, Lanzhou 730050, China*
[2]*School of Management, Xi'an University of Architecture and Technology, Xi'an 710055, China*

Correspondence should be addressed to Daopeng Wang; wangdp@lut.cn

Construction industry is the largest data industry, but with the lowest degree of datamation. With the development and maturity of BIM information integration technology, this backward situation will be completely changed. Different business data from a construction phase and operation and a maintenance phase will be collected to add value to the data. As the BIM information integration technology matures, different business data from the design phase to the construction phase are integrated. Because BIM integrates massive, repeated, and unordered feature text data, we first use integrated BIM data as a basis to perform data cleansing and text segmentation on text big data, making the integrated data a "clean and orderly" valuable data. Then, with the aid of word cloud visualization and cluster analysis, the associations between data structures are tapped, and the integrated unstructured data is converted into structured data. Finally, the RNN-LSTM network was used to predict the quality problems of steel bars, formworks, concrete, cast-in-place structures, and masonry in the construction project and to pinpoint the occurrence of quality problems in the implementation of the project. Through the example verification, the algorithm proposed in this paper can effectively reduce the incidence of construction project quality problems, and it has a promotion. And it is of great practical significance to improving quality management of construction projects and provides new ideas and methods for future research on the construction project quality problem.

1. Introduction

As large-scale, group-oriented, and complicated construction projects, especially large-scale cluster projects, are constructed, traditional project management theories, methods, and models could not fully meet the needs of actual management anymore. Engineering quality management is one of the key contents of domestic and foreign researches, for it not only relates to the project itself but also relates to the life and property security of people. At the same time, although the management level of current construction project quality management in China is continuously improving, quality problems are also on the increase.

The 21st century is an era of data explosion. All walks of life are flooded with massive amounts of data, which contain enormous commercial value. Therefore, big data has become the focus of attention. Since the reform and opening up, China's construction industry has rapidly grown in size, promoted economic and social development, and continued to expand the market capacity of the industry [1]. However, in comparison with some architectural powerhouses such as Germany and the United States, China's construction industry still suffers from low technical level, labor-intensive and low-efficiency construction, and industrial chain fragmentation and other pain points. This is a consequence of neglecting construction process big data. In foreign countries, the information integration technology BIM based on building big data has become more mature [2]. With the aid of big data mining technology and BIM, a systematic, refined, and information-based management model has been formed [3, 4]. At this stage, most domestic companies are focusing on construction units to reduce construction costs and maximize profits. They ignore the overall control of the entire construction project life cycle, causing frequent quality problems during the construction phase [5]. However, compared with some building powers such as Japan, Germany,

and the United States, the construction industry in China still suffers from pain spots of low technical level, labor intensiveness, low efficiency, and split industrial chain, which are all the causes for frequent project quality problems. In foreign countries, the architectural big data integration management model based on building information model (BIM) technology has well developed. With the help of big data searching technology and BIM technology, a systematic, refined, intelligent, and informatized management model has been formed. At present, most domestic companies, to reduce the construction cost and maximize the interests, are focusing more on the management at the construction phase while neglecting to control the entire construction project over their entire life cycle, causing the periodicity and locality of the project management, and fragmentation. The internal link of the project has been cut down, resulting in the disruption of the information flow of the project and of information-isolated island phenomenon. And hence the entire construction project may be lacking of a unified planning and control system. Therefore, in order to further improve project quality management and meet the needs of rapid economic development in the new era, we must strive to explore new management ideas and methods.

Research results on quality issues of engineering projects at home and abroad would be of high reference value for resolving project quality problems. However, there are also many shortcomings that are insufficient to support the requirements of a quality management of modern engineering projects. Further research is needed. At present, quality issues are solved by means of postremediation, when quality problems have already brought losses to the construction and use of the project and even led to quality and safety accidents, threatening the lives and property of the people. This is inconsistent with the quality management concept of "prevention first." Therefore, in solving quality issues, a brand new perspective and thinking should be adopted to find a practical method and means. And in modern engineering project management, we should ensure the quality of construction products through prior management and implement the management philosophy of "prevention first." This paper takes the quality of construction projects as the research object, uses the big data generated by the construction engineering process as the data source of enterprise-level BIM, extracts the historical data of many project cases for data searching, deeply explores the root causes of quality problems, and discusses the loopholes in the current project quality management in China so as to optimize the engineering quality management system. The research results would have certain practical significance for improving the quality management level of construction projects and provide new ideas and methods for future researches on construction project quality issues.

This paper takes the big data generated by the construction engineering process as the data source of enterprise-level BIM, extracts the historical data of many project cases and conducts data mining, discusses the loopholes in the construction quality management of Chinese construction companies at this stage, optimizes the project quality management system, and provides a certain reference for the further development of China's construction industry.

2. Construction Quality Management Big Data Based on BIM

2.1. Management Status of Construction Project Data. Construction industry is the industry with the largest scale and the largest amount of data. The construction project has a relatively long life cycle and is generally divided into the design phase, construction preparation phase, construction phase, completion phase, and operation and maintenance phase. Each phase will generate a large amount of data, such as a large number of engineering drawings in the design and construction phase, raw materials, basic components, cost, quality, security, materials, and other information, so the entire project will produce a large amount of data from the beginning of the construction to the finalization; it can be divided into two types of structured and nonstructured and stored in the form of digital statements and text files [6]. At present, these massive engineering data materials are scattered over places. It is of great difficulty to classify, save, query, and update. And there is no system for data crating, calculating, managing, applying, and sharing. The massive data materials still need to be manually handled by project managers every day. The project managers simply rely on paper materials or personalized network communication to transmit project information, making it difficult to query the data required for project management in a timely manner. The real-time tracking of projects is also impossible, which will undoubtedly cause a great waste of data resources. The low degree of informationization has, to a large extent, constrained the transformation and upgrading of China's construction industry to become "excellent, smart, and green."

Big data has long been in physics, biology, environmental ecology, military, finance, and communication industries. However, the construction industry with huge amounts of data still does not have its own enterprise-level and project-level databases. The whole project has been isolated from the Internet and big data, so it is much weak in management, innovation, transformation, and updating. The traditional architectural engineering quality management theories, methods, and thinking pattern could no longer meet the requirements of the big data era under the new situation. To solve quality problems, we must think out of the traditional mindset and search for new theories and methods which are suitable for the new environment, new technologies, and new situations, from a new perspective and direction to guide construction project quality management activities and dissolve construction project quality problems. Big data has provided new ideas for optimizing construction project quality management. In the era of big data, we can analyze more data, extract project quality data from the big data multidimensionally, and explore root causes for quality issues with data search methods. Moreover, it can enable us to predict the key points of quality management in the implementation of new projects and solve the drawbacks of traditional quality management by relying on postevent
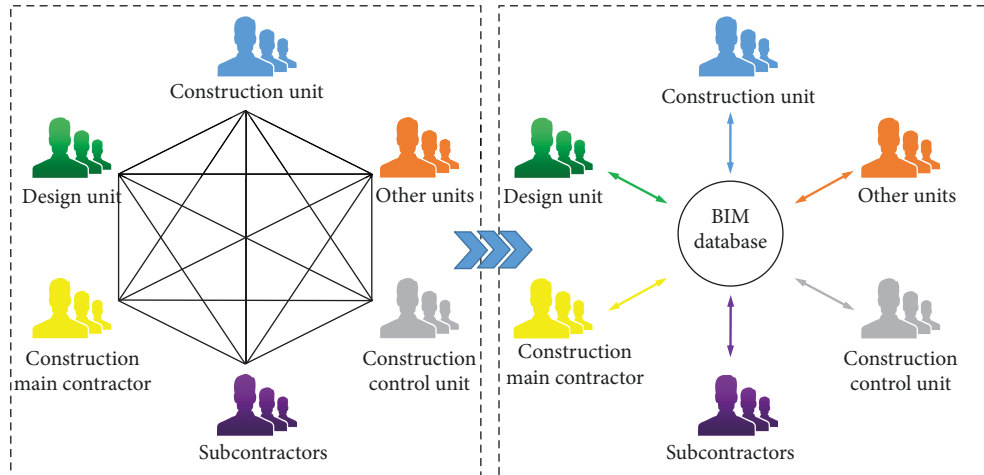
Figure 1: Changes in the mode of information exchange.

inspections to control product quality, so as to prevent the occurrence of quality problems, and realize and implement the quality management concept of "prevention first."

Considering that these building data are scattered in the localized management in China, the classification, preservation, query, and update of data are very difficult, and a systematic data management and utilization system is not formed. A large amount of data still needs a data manager to handle manually every day, while the user of the data—the project manager or technician—only depends on paper materials or personal network communication to transmit the project information [7, 8]. It is difficult to query and manage the construction materials in real time or to track the construction of the project in real time. This undoubtedly caused a waste of data resources, which, to a large extent, has constrained the development of China's construction industry.

2.2. BIM Data Integration Platform. Building information modeling (BIM) is based on various related data generated in the implementation of a construction project. It constructs a database with the data collected over the entire life cycle of a building project and breaks down the single-line links between the participants of the project. The model changes the passive situation in which traditional projects rely on paper materials or personalized network communication to deliver project information, enabling participants to understand the progress of the project in real time and using Internet technology to search for the latest, most accurate, and most complete project data and data. It reduces quality problems caused by low collaboration efficiency and is an important way to realize the refinement and information management of the construction industry.

The birth and development of BIM break the single-line contact model between participants of the project. The era of relying on paper materials or personalized network communication to deliver project information is gone forever, enabling participants to understand the progress and profile of the project in a timely and comprehensive manner which reduces many unnecessary quality issues. The emergence of

the BIM data integration platform has brought great advantages to data continuity and consistency. Project management in all stages of the life cycle is based on the 3D solid model [9]. Each participant continuously inputs and updates the BIM model and extracts basic information such as geometric parameters, physical feature parameters, and functional attribute parameters of the extracted components, as well as information on management factors such as integration project quality, safety, cost, schedule, and civilized construction, are used as the extended attribute information of the components to realize the project product quality management business process [10, 11]. Information integration, full life cycle information integration, and management organization information integration have resulted in a full information model. The entire life cycle BIM database is updated in real time. Each participant can share data information from different perspectives within their jurisdiction and work collaboratively. The change in the way information is exchanged by BIM is shown in Figure 1.

2.3. BIM: "Source Code for Construction Big Data." Big data is a collection of data that cannot be captured, managed, and processed using conventional software tools within a certain time frame. It is a vast amount of data that requires a new processing model to have greater decision-making power, insight, and process optimization capabilities. Big data is a diversified information asset with a high growth rate. The core of big data technology lies in the specialized processing of data, which searches for data information and adds value to data by increasing the "processing ability" of data. BIM has a powerful back-end storage system, including data layer, model layer, and information application layer, which creates an efficient platform for information integration. Based on the information data of the construction project, it defines basic data such as collection attributes, physical structure attributes, and functional attributes of the components and builds a 3D building information model based on these data. BIM can realize dynamic, integrated, and visual information management. Model objects are related to attribute information and report data [12, 13]. Entry, modification, deletion,
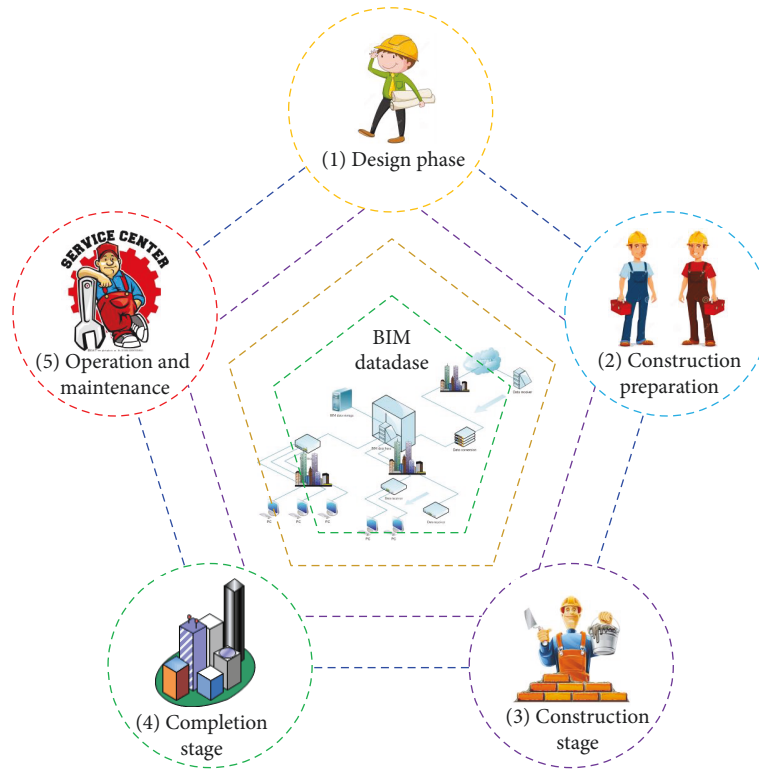
Figure 2: The data source of the BIM database.

and update of attribute information of model objects will lead to a real-time update of report data associated with them, which can ensure the dynamic transmission of information. With the progress of the project and the in-depth application of 3D building models, information on the design phase, construction preparation phase, construction phase, completion phase, and operation and maintenance phase of the construction project are continuously integrated on the basis of the 3D model to ensure the continuity and consistency of the phase information, ultimately form the BIM model of project product and business process information, life cycle information, and management organization information integration [14, 15]. The data source of the BIM database is shown in Figure 2.

Data is the support of management. The foundation of engineering project management is the management of engineering data. BIM technology accelerates the informationization of the construction industry. BIM technology can record all the data of the entire project lifecycle and create a project database. The accumulation of multiproject data based on BIM will form an enterprise database that internally stores massive amounts of data. Therefore, BIM can be regarded as the carrier and foundation of the construction industry database. It can be called the "source code of construction industry big data" [16, 17]. The key to the application of big data in the construction industry is to realize "value adding" of engineering data through data calculating, sharing, and applying and provide support for management decisions so as to promote the transformation and upgrading of Chinese construction industry to become "excellent, smart, and green" [18].

BIM technology accelerates the informatization of the construction industry. BIM, which is known as the "source code of construction industry big data," can record all historical data of the entire project lifecycle. The cycle-based accumulation of multiitem data based on BIM will form a huge database and store a large amount of data internally. Therefore, BIM can be regarded as a stable and reliable database. Based on the enterprise-level BIM database, this paper abstracts the multidimensional data from multiple cases and extracts the quality information data. Through data mining methods, it explores the root causes of many quality problems during project development and focuses on the key points in the project quality management process when the new project is launched and strengthens the ability to control the project management, which is important for the fundamental improvement of project quality management.

3. Quality Data Extraction and Preprocessing

On the basis of using BIM to generate, extract, and mine quality management data and to convert from BIM model data to BIM big data, we should pay attention to the value of big data and be the owner and beneficiary of big data. Among them, the construction engineering quality management text data as a branch of data mining is based on the knowledge discovery of text documents, mining hidden, valuable, potentially unknown information from a large-scale text collection needs to pretreat the unstructured data in the text collection, such as text cleaning, text segmentation, text clustering, semantic network analysis, and so on [19, 20].
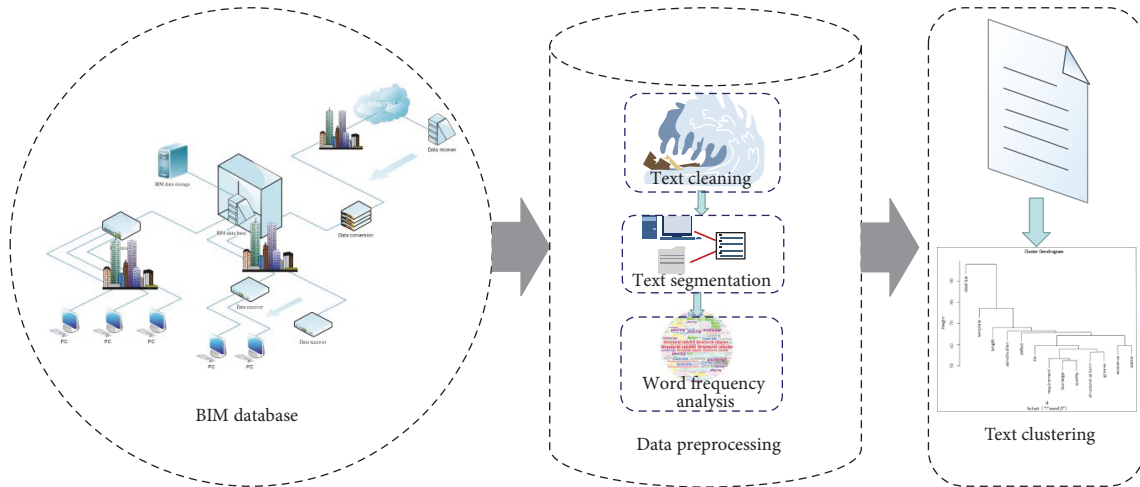
FIGURE 3: The process of text mining.

The data mining process of the text of architectural engineering quality management is shown in Figure 3.

Step 1. Analyze a large amount of unstructured text data, formulate data cleaning rules, and perform text segmentation to obtain a standardized data source.

Step 2. Extract the links between keywords and keywords by text clustering, semantic network analysis, and so on to achieve dimensionality reduction and text representation of the text, and visualize the data in the form of tools;

Step 3. According to the text content, further analyze the visual data to obtain useful knowledge information in the text.

3.1. Text Data Cleaning. In data warehouses, unstructured data are described in natural language and in various forms. Due to the recording habits, errors in records, and incomplete records, problems such as incomplete data, incorrect data, and inconsistent data are inevitable. Through the cleaning process, duplicate data, missing values, entry errors, and meaningless values are processed and abnormal data are detected and adjusted as soon as possible to provide standardized high-quality data for subsequent excavation work. Mainly, the following measures should be adopted to achieve data normalization.

3.1.1. Data Deduplication. In natural language, different expressions of the same object will form different texts, and the segmentation process will form different nodes. To avoid duplicate nodes, text mining effects are affected. First, the original text is quasisegmented and the different representations of the same object are screened; secondly, the merger rules are formulated and the duplicate expressions are merged. By deduplicating the data (see Table 1), the representation of the same object in the text is converged as much as possible, and the discreteness of the data node is reduced.

3.1.2. Correcting the Input Error. At present, the ability of automatic data acquisition is low, and most of them are manually entered, and there may be some problems with input errors. By comparing the existing input data with the contents of the standard query table and adding the artificial semantics to identify, the error information is corrected; after that, the standard query table established by the initial demand can be referred to, and the data in the table can be directly selected to reduce the input error (see Table 2).

3.1.3. Handling Missing Values. In partially missing data, infer and supplement records based on the contextual semantic environment (see Table 3), and filter data for data that has no practical significance.

3.2. Text Segmentation. Because of the two commonly used text segmentation rules (word segmentation algorithms based on dictionary matching and word segmentation based on statistics, that is, probability segmentation), they are all based on the accumulation of a large number of corpus and existing dictionaries. The Chinese expression is organized according to certain grammatical rules based on real words and virtual words. In the Chinese expression, there is no obvious segmentation between the words; in addition, different words form different words and correspond to different meanings. Based on this, for a large amount of text data, text segmentation is needed, which will help quickly parse text semantics from a large amount of text. The ICTCLAS (Institute of Computing Technology, Chinese Lexical Analysis System) 3.0 developed by the Institute of Computing Technology, Chinese Academy of Sciences, is the best Chinese lexical analyzer in the world. The R language supports word segmentation-based word processing algorithms. It is much more accurate and efficient than other word segmentation algorithms.

First of all, we segmented the texts of the main structural quality issues of the construction projects that have been collected; at the same time, phrases annotated with text and have no significant effect on text content such as "existence," "partiality," "one layer," "parts," "individual," "serious," and

TABLE 1: Mapping of synonyms for architectural engineering.

The group of words with the same meaning	Normalization
Mutual anchoring length; anchorage length; bending anchor length; elbow length; hook length	Anchorage length
Concrete; concrete; commercial concrete	Concrete
Standard curing chamber; concrete curing room	Standard curing chamber
Aseismic reinforcement; rebar with E	Aseismic reinforcement
—	—
Design alteration; altered design; change	Design alteration
Treatment of general quality diseases; requirements for prevention and treatment of general quality diseases; prevention and control measures for general quality diseases of residential engineering	Prevention and control measures of general quality diseases
Levelness; planeness	Planeness
Out of plumb; unstraight; not square	Out of plumb
Acceptance specification; acceptance specification of quality; acceptance specification of construction quality	Acceptance specification of construction quality

TABLE 2: Record error correction record.

Original record	Error correction
The cement should be fed on demand; pay attention to the waterproof; the good processing has just covered to prevent the corrosion.	The cement should be fed on demand; pay attention to the waterproof; the processed steel is covered to prevent the corrosion.
The length of the connecting beam of the steel bar anchored to the bottom of the two floors of the number 2 floor is insufficient.	The length of the steel bar anchored to the connecting beam of the two stories suspended the beam of the number 2 building.
Dismantling the floor of the floor is not in place.	The protection of the dismantling floor slab is not in place.
Stirrup	Horse stool
Silk mouth	Thread
Number 18 windowsill beam is not normally set.	Number 18 window beam Weitong long set
25-layer ALC cutting quality did not meet the requirements.	25-layer ALC block quality did not meet the requirements.

TABLE 3: Missing value processing records.

Original record	Processing method	Processing results
Sand filling, ash adding	Complete records	Sand aerated block, ash adding block
GZ1	Complete records	Steel column 1
Short sleeve	Filter	Semantic ambiguity and inability to divide subdivisional projects
ApproachΦ8, Φ10 bale bar	Filter	The semantics is not clear, and the quality problem cannot be judged

"phenomenon" are filtered out; the phrases with the frequency of 50 times or more (see Table 4) are the abbreviations for the locations and problems of common quality problems in the construction process.

3.3. *Word Cloud Visual Analysis.* Loading a custom dictionary into the R language enables text segmentation based on dictionary matching. By introducing the word segmentation results into the ROSTCM 6.0 software for word frequency statistics, the text of the subject structure quality problem is finally visualized in the form of a word cloud. For all phrases in the entire text set, word cloud expressions can display rich text information [21]. The word cloud reflects the word frequency in text size and assists in the display of colors to distinguish phrases with different frequency ranges. For word frequency statistics, in the R language, by using the wordcloud2 function for word cloud drawing (see Figure 4), the wordcloud2 function can achieve the interface of the output results, and the mouse is placed on the phrase, and the frequency of occurrence of the phrase will automatically pop up.

The word cloud can demonstrate the key summary of quality problems in the process of construction project quality management. It mainly reflects issues in two aspects. First, the distribution of high-frequency phrases is used to identify the hot spots of quality issues. By counting the changes of high-frequency words over time and identifying the trend of changes in quality, we can, to a large extent, reflect the remediation of existing quality defects and identify new emerging quality problems. This has a great role in guiding the development of quality management of construction projects. Second, the monitoring of low-frequency

TABLE 4: The text keyword table of the quality problem of the main body structure.

| Words	Freq	Words	Freq	Words	Freq	Words	Freq
Rebar	711	Stirrup	139	Frame column	84	Honeycomb	73
Concrete	361	Anchorage	120	Drag hook	83	Tie bar	71
Construction	244	Shear wall	112	Design requirement	83	Wall column	700
Wall	226	Spacing	107	Set up	82	Exterior wall	69
Pouring	208	Roof	98	Design requirement	80	Drawing	68
Lashing	197	Welding	97	Electroslag pressure welding	77	Block	68
Tectonic column	190	Main reinforcement	96	Bonded rebars	76	Dismantle	68
Structural column	188	Install	95	Joint	75	Mortar	64
Deficiency	166	Stairs	91	Postpouring belt	73	Lack of length	64
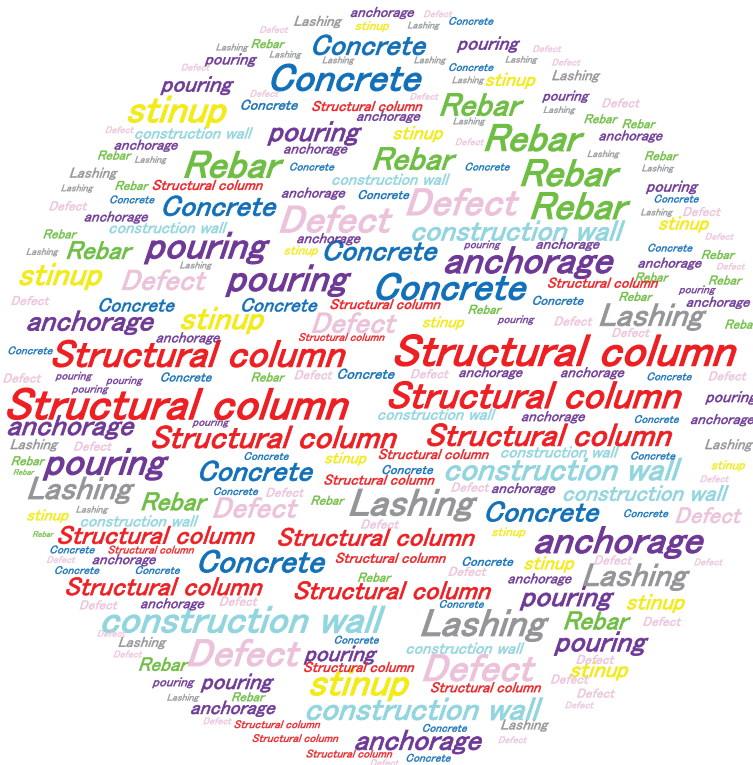
FIGURE 4: Textual nephogram of the quality problem of the main structure of the building engineering.

phrases can reveal the emerging quality of construction technology and the quality problems that are gradually exposed during its application. For the early detection of new quality problems, the prevention of the quality of new technologies is of great benefit.

It can be seen from Figure 4 that there are many problems in the quality of components and parts such as walls, structural columns, and structural columns. Among them, the quality problems of the rebar project are the most prominent. The problems are mainly stirrups, main bars, pull hooks, and pull-reinforced bars. The problems are concentrated on the rebar processing (lack of length, etc.), connections (banding, welding, especially electroslag pressure welding, and thread connection), and installation (anchors, anchorages, reinforcing bars, spacers, spacing, and joint locations). Secondly, the problems of pouring and appearance (honeycomb, pockmark, ribs, and holes) in the concrete subproject are prominent; the main problem in the subproject of the formwork is the installation of the formwork including the support system (rack, pole, and spell sewing) and demolition process.

3.4. Mining Analysis Based on the Clustering Algorithm. Convert text content into a matrix, and assign corresponding weights to each phrase. TF-IDF is used to evaluate the importance of a phrase for a text set or a text file in a corpus. The greater the TF-IDF value of a phrase is, the higher the frequency TF of the phrase appears in a text, but rarely occurs in other texts, and it can be considered that the phrase has a good category distinguishing ability. It is used as a keyword for text clustering. Furthermore, since the description texts of the quality questions studied in this paper are short texts,

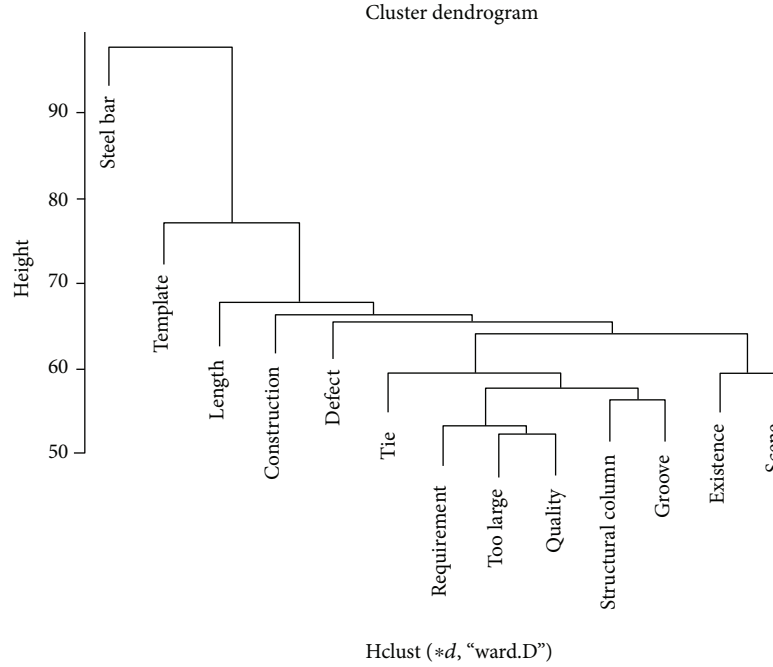Cluster dendrogram

Hclust (∗*d*, "ward.D")

FIGURE 5: Text clustering results.

the probability of repeated occurrences of the same keyword in any text is extremely low, and the calculated TF-IDF is more accurate. Therefore, in this text mining, the weight of each phrase is calculated using the TF-IDF value, and the TDM matrix (TF-IDF value of the text∗ feature phrase) obtained by the transformation is more representative as the calculation basis of the text clustering. This can not only structure the unstructured text information but also play a role in dimension reduction.

3.4.1. Mathematical Characterization of TF-IDF. The TF (term frequency) in TF-IDF represents the frequency of occurrence of a certain phrase *d* in a document *D*. In IDF (inverse document frequency), the less the text that contains a phrase *d*, the smaller the *m* and the larger the IDF, indicating that the phrase *d* has a good category distinguishing ability. The mathematical representation of TF-IDF is

$$\mathrm{TF} = \frac{t}{d},$$

$$\mathrm{IDF} = \log\left(\frac{n}{m}\right), \tag{1}$$

$$\mathrm{TF} * \mathrm{IDF} = \left(\frac{t}{d}\right) * \log\left(\frac{n}{m}\right).$$

In the formula, *t* represents the frequency of occurrence of the phrase *T* in a certain text *D*, *d* is the total number of subgroups in the text *D*, *n* represents the total amount of text, and *m* represents the number of occurrences of the phrase *T* in *n* texts.

3.4.2. Determine the Number of Clusters. Check the relevant literature, and adjust the number of sparse entries to determine the number of categories that need to be classified.

In order to achieve dimensionality reduction of high-dimensional texts and to analyze the classification results, the final choice is to remove 3% of sparse entries and divide the text into 13 categories.

3.4.3. Clustering Results. The ward.D method is a method of clustering by the sum of squared deviations and an index of the corpora after text segmentation. The index is set by using the hierarchy method, the upper class will include the content index of the lower class, and the lowest class has the clearest classification. The vertical coordinate height represents the sum of squared deviations between classes. The greater the square and increase in dispersion between classes, the more differentiated the two categories are. The results of text clustering are shown in Figure 5.

The class or cluster in the clustered tree is the center of the class or the boundary point of the class or the logical representation of the sample attributes. Therefore, the results of the clustering need to be conceptualized. The ontologies studied in this paper are the ontologies of the field of construction engineering. Ontologies in the field of quality of construction engineering follow the general definition of domain ontology, that is, research on individual (entity) or individual collections (concepts) through specific rules in the domain. The relationship between entity and concept is described by adopting some related characteristics and parameters of building engineering. Ontology in the field of architectural engineering quality links textual data of architectural engineering through certain logical relationships, allowing this disorderly information to follow certain rules and forming an information network with rich relationships, clear relationships, and large volumes through the description of ontology.

From Figure 5 and ontology reasoning in the field of construction engineering, "rebar" and "template" are individually

divided into different categories, and the "rebar" category belongs to the upper level of the "template" category. Judging from the knowledge experience method, "rebar" represents the rebar project, that is, the subitem project category; "template" represents the material name. "Length" represents the property information of the material. "Construction" represents different stages of the construction project. "Missing" is the quality description of the results of different stages of construction engineering, whether or not there is a description. The phrases associated with "site" mainly include structural columns, concrete, steel bars, construction, inspection, supervision, inspection, presence, banding, and in situ situation description. There are two types of words that have a cooccurrence relationship with "existence": first is the description of components, parts, and work, such as shear walls, structural columns, formwork, floor, parts, masonry, walls, podiums, completion, site, inspection, rectification, individual, concrete, steel, main reinforcement, process, banding, and so on. Second is the description of the existing phenomena and problems at the site, such as phenomenon, serious, missing, misalignment, poor, pockmark, honeycomb, quality, quality problems, and not in place. There is a description of the presence of a site, process, and problem. "Site" and "existence" are at the same level, with little difference and a comprehensive description of the scene. "Lashing" is the process, work, and process. "Construction column" and "wall" are at the same level as the component name category; "requirement" is the standard specification that should be performed during the work and operation; "excessive" and "quality" are at the same level and are at the "requirements." "The next level is the quality description of the work results of different stages of the construction project, whether it meets the standard description.

4. Real-Time Forecast Model of Construction Project Quality Based on Improved Recurrent Neural Network

Large-scale construction projects, obvious stages, multiparticipation, and professional-related features make it difficult to implement refined and comprehensive quality management. The cluster analysis is used to correlate the quality descriptions of the different engineering fields of the construction project ontology in different engineering phases, and the word frequency analysis is performed. However, because cluster analysis can only rely on the degree of subjective discriminant quality problems, it is impossible to quantitatively and accurately determine the probability of quality problems. This uncertainty greatly increases the construction cost. In this paper, an improved recurrent neural network is introduced to predict the probability of occurrence of quality problems in different engineering areas under field construction conditions in real time, and corresponding management measures are taken to reduce the probability of occurrence of quality in order to improve the construction efficiency.

4.1. Traditional Recurrent Neural Network. Recurrent neural network (RNN) has achieved good results in the task of sequence modeling, and many tasks in real life are time
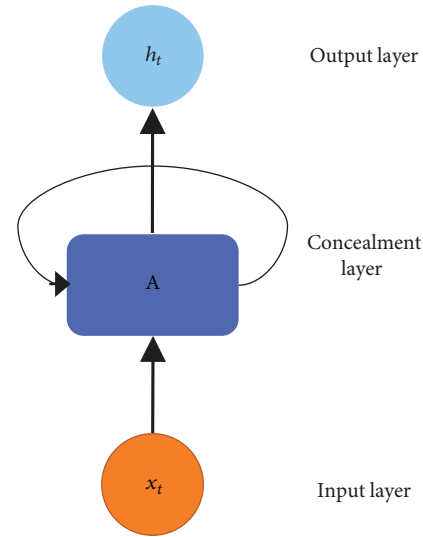

FIGURE 6: Traditional recurrent neural network.

series, such as natural language processing (NLP). The dialogue system and the machine translation system also have time series features in the collection of architectural text information during the construction period. Therefore, the recurrent neural network adapts to the real-time forecasting tasks of construction project quality management. A large number of textual data in construction projects are largely idle due to their processing difficulties. The establishment of a predictive model for the quality of recurrent neural networks makes an efficient use of architectural text data and optimizes the industrial structure. It is of great significance to accelerate the intelligentization of the construction industry.

Figure 6 is the structure of a conventional recurrent neural network. The recurrent neural network consists of an input layer, a hidden layer, and an output layer. The functions and structure of the input layer and the output layer are not different from the multilayer feed forward neural network. It can be seen from the figure that the difference between the circulatory neural network and the multilayer feedforward neural network lies in that, in addition to the connection between the layers, the circulatory neural network. At the same time, the connection is allowed to be added within the layer. The link within the layer allows the cyclic neural network to accumulate in the time domain. However, the traditional multilayer feedforward neural network has no concept of time, so only the current moment is considered in the training process. The characteristics of speech signals and the accumulation of circulatory neural networks in the time domain make the circulatory neural network more suitable for processing serial data-related machine learning tasks.

Figure 7 is a schematic diagram of the deployment of a recurrent neural network in the time domain. From the graph, each time point of the recurrent neural network includes the information of the first several moments and is calculated in chronological order when the recurrent neural network propagates forward. Postpropagation is cumulatively progressing from the gradient of the last moment.
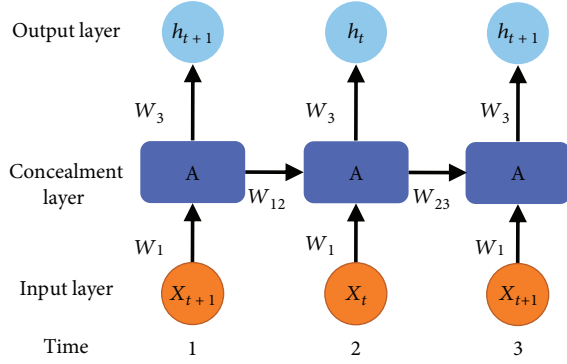
FIGURE 7: Recurrent neural network extends on time domain.

Equations (2) and (3) show the formal representation of the forward and backward propagations of the cyclic neural network, respectively.

(1) Forward propagation:

$$a_k^t = \sum_{h=1}^{H} w_h k b_h^t,$$

$$a_h^t = \sum_{i=1}^{L} w_{ih} x_i^t + \sum_{h'=1}^{H} w_{h'h} b_{h'}^{t-1}, \qquad (2)$$

$$b_h^t = \theta_h\left(a_h^l\right).$$

(2) Backward propagation:

$$\delta_h^t = \theta'\left(a_h^t\right)\left(\sum_{k=1}^{K} \delta_k^t w_{hk} + \sum_{h'=1}^{H} \delta_{h'}^{t+1} w_{hh'}\right),$$

$$\delta_j^t \overset{\text{def}}{=} \frac{\partial L}{\partial a_j^t}, \qquad (3)$$

$$\frac{\partial L}{\partial w_{ij}} = \sum_{t=1}^{T} \frac{\partial L}{\partial a_j^t} \frac{\partial a_j^t}{\partial w_{ij}} = \sum_{t=1}^{T} \delta_j^t b_i^t,$$

where a represents the sum calculated value, b represents the activation value of the value of the activation function, θ represents the activation function, t represents the time, L represents the loss function, δ represents the gradient, the angle mark, h and k represent the hidden layer and the output layer, respectively, and i and j represent the two connected layers and the number of nodes.

The circulatory neural network allows the output of the previous moment to be multiplied by the corresponding weight and then the output obtained by the activation function as the input at the current moment. Therefore, for the loop chanting network, the characteristics of the current moment often include the characteristics of the first n moments. Recurrent neural networks have better modeling capabilities for sequence data than ordinary neural networks. However, the traditional circulatory neural network does not have a good control over the accumulation of information in

time; that is, the depth in the time dimension is too deep, which often leads to the problem of gradient disappearance or gradient explosion. Therefore, the researchers are in the traditional recurrent nervous system. Based on the network, a door structure was added to control the length of memory. A well-known improvement of the circulatory neural network is described in detail below, allowing the model to adapt to the analysis and prediction of a large number of sequence data.

4.2. Improved Recurrent Neural Network. The traditional circulatory neural network incorporates the concept of time on the basis of a multilayer feed-forward neural network, provides a neural network with a memory function, and enables the neural network to show good modeling ability on time series data, but on the time dimension. The most immediate problem caused by the deep layers is that the gradient disappears or explodes. The long short-term memory neural network (LSTM) proposed by Hochreiter and Schmidhuber has well controlled this problem [22, 23].

The LSTM unit is shown in Figure 8. The LSTM unit specifically designed a memory cell to store historical information. The update and use of historical information are controlled by three doors: the input gate, the forget gate, and the output gate.

Let h be the LSTM cell output, c the value of the LSTM memory cell, and x the input data. The update of the LSTM unit can be divided into the following steps:

(1) First, we calculate the candidate memory cell value at the current moment $\overline{c_t}$, W_{xc}, and W_{hc} according to the traditional RNN formula, which are the corresponding input data and the weight of the LSTM cell output at the previous time.

$$\overline{c_t} = \tanh\left(W_{xc} x_t + W_{hc} h_{t-1} + b_c\right). \qquad (4)$$

(2) We calculate the value of the input gate i_t. The input gate is used to control the influence of the current data input on the state value of the memory cell. The calculation of all gates is affected not only by the current input data x_t and the output value of the last-time LSTM cell h_{t-1} but also by the value of the memory cell c_{c-1} at the previous moment. This scheme is called peephole connections.

$$i_t = \sigma\left(W_{xi} x_t + W_{hi} h_{t-1} + W_{ci} c_{t-1} + b_i\right). \qquad (5)$$

(3) We calculate the value of the forget gate f_t. The oblivion gate is used to control the influence of historical information on the current memory cell status value.

$$f_t = \sigma\left(W_{xf} x_t + W_{hf} h_{t-1} + W_{cf} c_{t-1} + b_f\right). \qquad (6)$$

(4) We calculate current memory cell status value c_t.

$$c_t = f_t \odot c_{t-1} + i_t \odot \overline{c_t} \qquad (7)$$

where $\odot$ represents the product of point by point. It can be seen from the formula that the state of
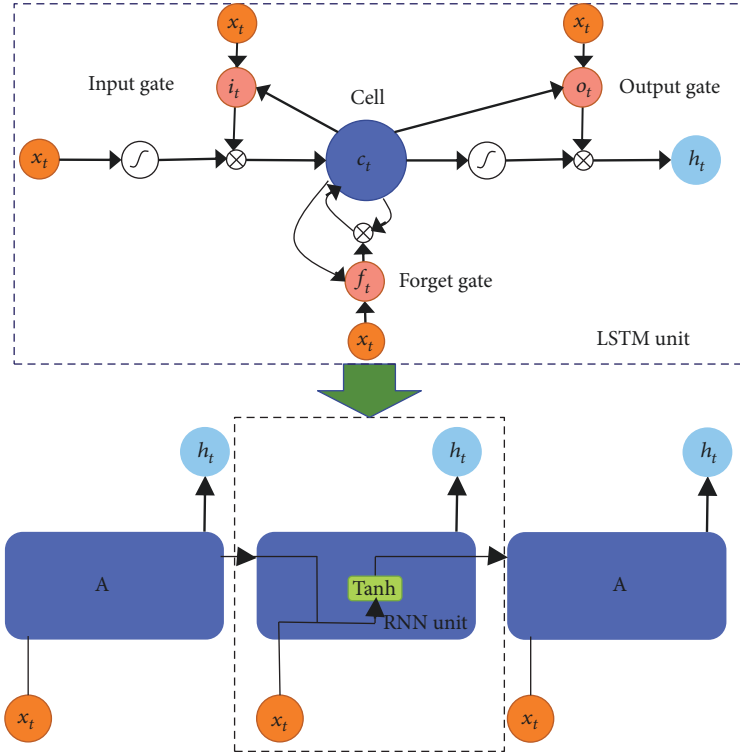
FIGURE 8: LSTM optimization of a recurrent neural network unit.

TABLE 5: The parameter of RNN-LSTM model.

(a)

Input vector $\mathbf{X}_i$	x_1	x_2	x_3	x_4	x_5	...	x_T
			Materials				
Engineering parameters	Subdivision project	Material name	Material properties	Project implementation stage	Scene description	...	Build name
Index text	Reinforcement engineering	Reinforced template	Length	Construction stage	According to the design	...	Construction column, wall

(b)

Output vector $\mathbf{Y}_i$	y_1	y_2	y_3	y_4	...	y_T
Quality problem node	Reinforcement processing		Reinforcing bar connection		...	Install
Index text	Reinforced length	Lashing	Welding	Threaded connection	...	Anchorage

the memory cell is updated depending on its own state and the current candidate memory cell values which are c_{t-1} and $\overline{c_t}$, respectively, and these two parts are adjusted by the input gate and the oblivion gate, respectively.

(5) We calculate the output gate O_t, the output for controlling the memory unit status value.

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o). \quad (8)$$

(6) The output of the last LSTM unit is given by

$$h_t = o_t \odot \tanh(c_t). \quad (9)$$

The general logistic sigmoid function in the above formula is the range of values $(0, 1)$. The design of the three doors and the separate memory cell allows the LSTM unit to save, read, reset, and update long-distance history information.

4.3. Real-Time Forecasting Model of Quality Based on RNN-LSTM Construction Project. Through the cluster analysis, the steel reinforcement project, the formwork project, the concrete project, and the masonry project are taken as separate construction project problems, and the relevant text data is analyzed and forecasted separately. Based on the BIM data platform, while assisting and optimizing the full-process

Figure 9: Network training of reinforcement engineering.
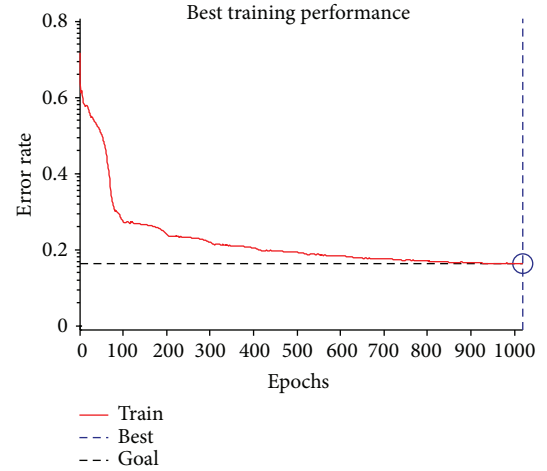

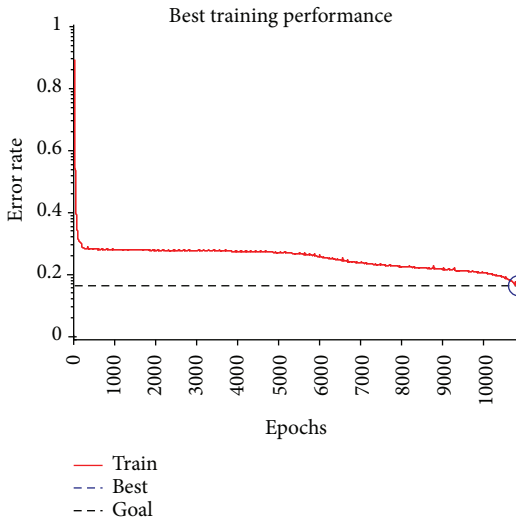Figure 11: Network training of concrete engineering.

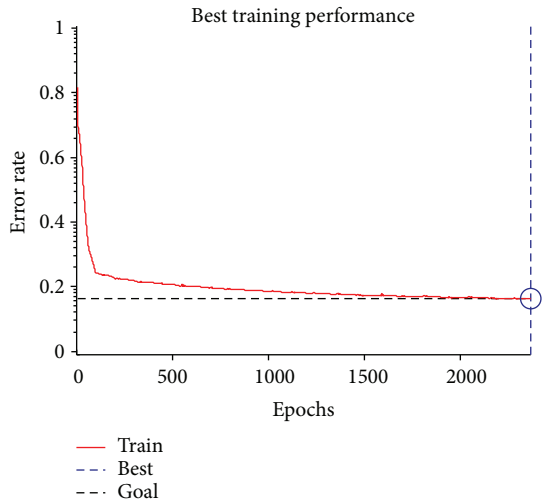Figure 10: Network training of template engineering.

Figure 12: Network training of masonry engineering.

quality management workflow in which all employees participate, the integration of heterogeneous and heterogeneous data from all stages, participants, and majors is consistent with the project quality sequence. Temporality, quality issues have some contextual relevance throughout the project construction process. That is, the input of historical engineering as a sequence requires a model that continuously learns the quality characteristics of the project before and after. The recurrent neural network (RNN-LSTM) has a memory-memory function by connecting feedback nerves. The project quality sequence forecasted in this paper is closely related to the quality of historic projects. For example, under certain construction preparation and construction conditions, there will be relatively close historical quality problems. RNN-LSTM can train sequence generation, process real data sequences at each step time, and predict what will happen next. The model is adjusted by continuously iteratively calculating the conditional probability between the input samples and the predicted results. Taking the engineering quality of steel reinforcement as

an example, the RNN-LSTM rebar engineering quality forecast was established (see Table 5).

The vector sequence $\mathbf{X}_i = (x_1, x_2, \ldots, x_T)$ is transmitted to a recurrent neural network that is linearly connected by an N layer-hidden layer, and the layer-hidden output vector sequence $\mathbf{h}^n = (h_1^n, h_2^n, \ldots, h_T^n)$ is obtained by iterative calculation. Then, pass the output vector to the output layer and calculate through the activation function to get the output result vector sequence $\mathbf{Y}_i = (y_1, y_2, \ldots, y_T)$. Each of the output result vectors is its corresponding quality problem, with quality problems represented as 1 and no quality problems represented as 0. The activation function from the moment $t = 1$ to the moment $t = T$. The hidden layer is continuously iteratively calculated by

$$h_t^1 = H\left(w_{ih^1} \cdot x_t + w_{h^1 h^1} \cdot h_{t-1}^1 + b_h^1\right). \quad (10)$$

The framework according to the model of this article only involves a layer of LSTM hidden layers; of course, this can be continuously adjusted as a parameter, where w represents the weight between the input layer and the hidden layer,

(a)

(b)

(c)

(d)

Room legend

▢ Bath	▢ Laundry
▢ Hall	▢ Living
▢ Kitchen & dining	▢ Mech.

(e)

Room legend

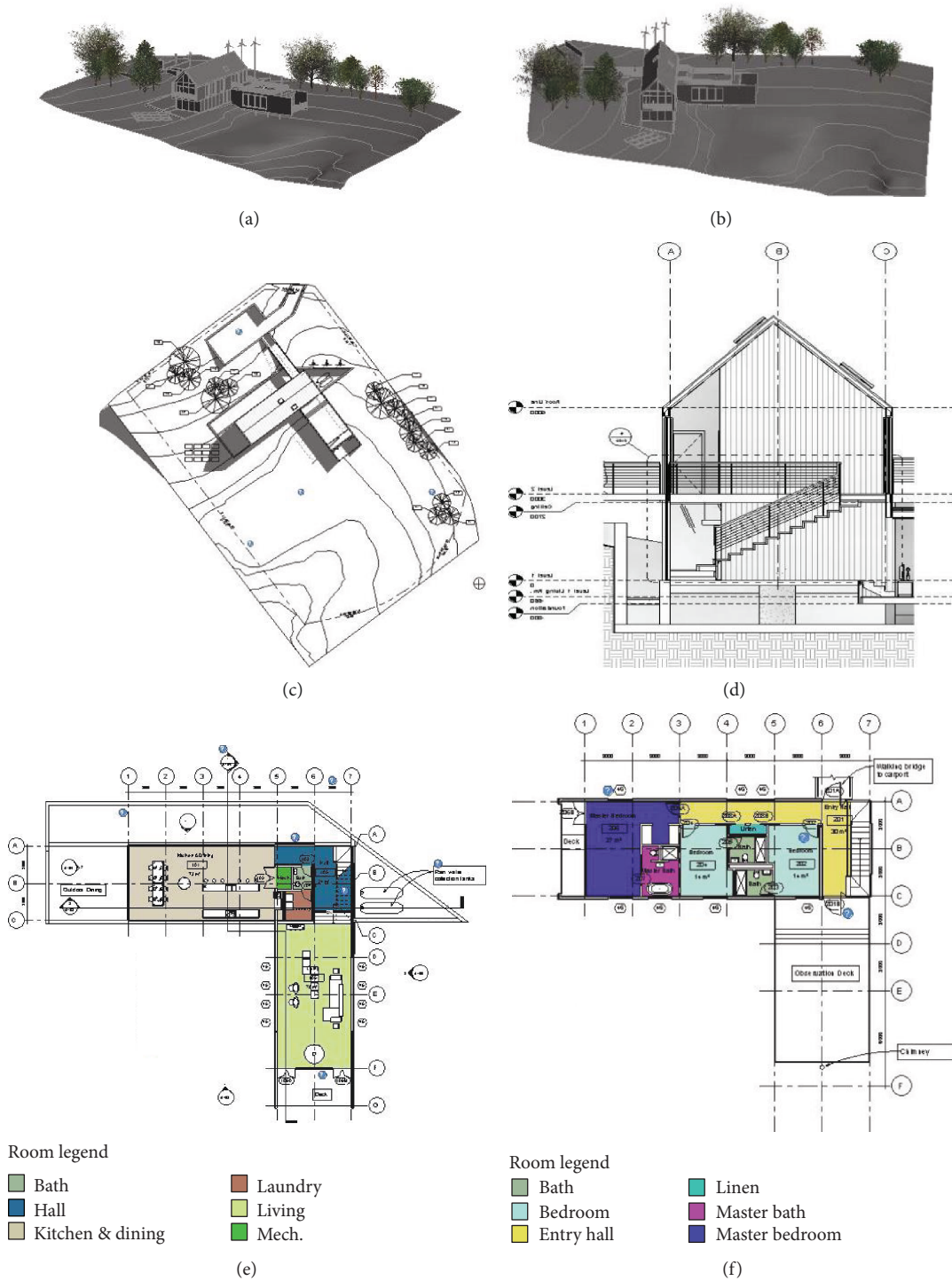|	
▢ Bath	▢ Linen
▢ Bedroom	▢ Master bath
▢ Entry hall	▢ Master bedroom

(f)

Figure 13: Examples of X engineering design.

b represents the bias term, and H is the activation function of the hidden layer.

After the clustering, the structured BIM platform big data is trained through the RNN-LSTM network. When the overall error rate is below 20%, the training is stopped.

As shown in Figures 9–12, save the trained network to guide the quality inspection of the design phase and construction preparation phase template engineering network, the concrete engineering network, and the masonry engineering network. When the output vector is $\mathbf{Y}_i = (0, 0, L, 0)$ or the frequency of occurrence of 1 is lower than the set value, it indicates that the quality of the design and construction preparation is in good effect and the construction can be implemented. When the output vector is not $\mathbf{Y}_i = (0, 0, L, 0)$ or the frequency of the occurrence of 1 exceeds the set value, the reverse modification of the design scheme and the

TABLE 6: Frequent-point statistics of quality problems.

| Subproject	Frequent points of quality problems
Reinforcement engineering	Stirrups, main ribs, pull hooks, pull tabs, concrete works
Concrete engineering	Pouring and appearance (honeycomb, matte, exposed bar, hole)
Template engineering	Installation of support systems (racks, poles, joints) and demolition
Masonry engineering	Grey joints, verticality control, retention of horse teeth and holes during masonry
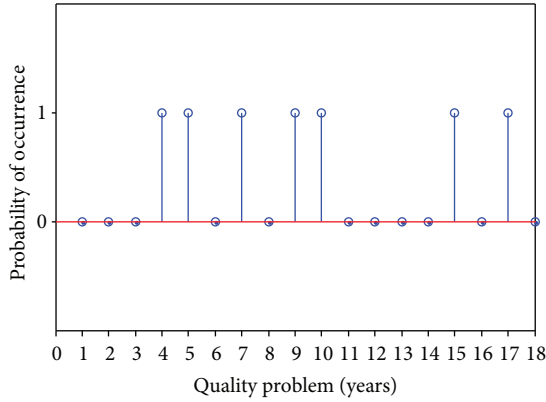
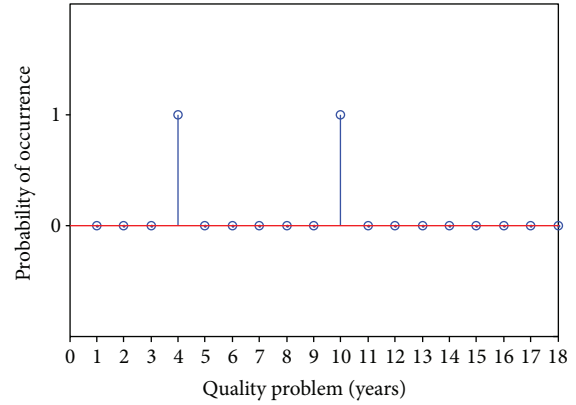FIGURE 14: Prediction of engineering quality problems.

FIGURE 15: Prediction of quality problems after revision.

material preparation scheme effectively reduces the probability of construction project quality problems.

5. Case Analysis

This article is based on a new project (X) for selected construction companies, the theme of which is to design and build a two-story villa, whose total construction area is 821 square meters, the investment of which is about 10 million yuan and the duration of which is 85 calendar days. The project completion plan and the three-dimensional model diagram are presented in Figure 13.

Figures 13(a) and 13(b) show the three-dimensional views of the completion of the X project. Figure 13(c) is a roof's plan of X project. Figure 13(d) is a cross-section. Figures 13(e) and 13(f) represent the first floor's plan and the second floor's plan of the X project, respectively. The X project is similar to the type of historical case of the construction company, and the different characteristics and past cases are as follows:

(1) The schedule is tight and the task is heavy. The total construction area is 821 square meters, and the project duration is 85 calendar days. In order to ensure the project will be completed and delivered for using within the contract period, participants must work together efficiently, maximizing the quality of production, guaranteeing completion on schedule.

(2) In large project volume, if using traditional quality management methods to carry out management activities, we are prone to quality problems due to lack of management, in turn affecting project quality and duration.

In this paper, we excavate the historical data of similar cases in the BIM database of the enterprises, discovering the entire building lifecycle quality problem of the company's past cases (see Table 6 for details).

After the training of the neural network model passed, quality-related data generated in real time during construction as input was used, to achieve real-time forecasting of project quality. Based on data mining and engineering quality prediction model, construction projects can be completed on time with quality assurance, and during the entire project term, its frequency of having quality problems will be greatly reduced (see Figures 14 and 15).

Based on the historical data mining results of previous cases and the established construction project quality forecast model, providing a starting point for optimizing the quality management system, in the light of the five stages for the operation of construction projects, the following requirements can be made:

(i) In the design stage, when drawing the design, we should pay attention to the accurate checking of the construction property, dosage, and use of subitems such as steel bar, concrete, formwork, and masonry.

(ii) In the construction preparation stage, using BIM's powerful 3D modeling tools, we need to perform a collision check on drawings and design plans, focus on model comparison of subengineering parts, and reduce the quality problems caused by drawing design deviations.

(iii) In the construction stage, through text mining, finding that the quality problems in the construction process are mostly located at the node, therefore,

enterprises should perform node management and node visualization for complex nodes. In addition, we set up a project quality rescue team, based on engineering quality prediction model, solving the problem of quality timely and effectively.

(iv) In the completed stage, we invite an authoritative quality inspection agency to carry out a quality inspection of the project.

(v) In the operation and maintenance stage, we regularly check the quality of the steel bar, concrete, formwork, and masonry structure.

6. Conclusion

This research takes the quality of construction projects as the subject, and BIM integrated construction engineering big data source as the foundation. Firstly, this article is based on BIM integrating big construction data sources, extracting the text data related to the project quality, and carrying out data cleaning and text segmentation. Then, with the help of word cloud visualization and cluster analysis, we mine the links between data structures, lock the frequent points of quality problems throughout the building's lifecycle, and turn the integrated unstructured data into structured data. Finally, using RNN-LSTM to predict the quality problems of the subdivision projects such as steel bars, formworks, concrete, cast-in-place structures, and masonry in construction projects, it is more precise to locate the node of the quality problem in the implementation of the project. Through strength verification, this method can reduce the incidence of quality problems in construction projects effectively. The empirical study proves that the method is feasible, scientific, and reasonable and can effectively reduce construction project quality problems.

Data Availability

The data used to support the findings of this study are currently under embargo while the research findings in this paper are confidential. Requests for data, after the publication of this article, will be accepted by the corresponding author.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] A. GhaffarianHoseini, T. Zhang, O. Nwadigo et al., "Application of nD BIM integrated knowledge-based building management system (BIM-IKBMS) for inspecting post-construction energy efficiency," *Renewable and Sustainable Energy Reviews*, vol. 72, pp. 935–949, 2017.

[2] A. H. GhaffarianHoseini, N. D. Dahlan, U. Berardi, A. GhaffarianHoseini, N. Makaremi, and M. GhaffarianHoseini, "Sustainable energy performances of green buildings: a review of current theories, implementations and challenges," *Renewable and Sustainable Energy Reviews*, vol. 25, pp. 1–17, 2013.

[3] H. Hamledari, B. Mccabe, S. Davari, and A. Shahi, "Automated schedule and progress updating of IFC-based 4D BIMs," *Journal of Computing in Civil Engineering*, vol. 31, no. 4, article 04017012, 2017.

[4] Z. Liu, D. Wu, H. Yu, W. Ma, and G. Jin, "Field measurement and numerical simulation of combined solar heating operation modes for domestic buildings based on the Qinghai–Tibetan plateau case," *Energy and Buildings*, vol. 167, pp. 312–321, 2018.

[5] V. W. Y. Tam, S. Senaratne, K. N. Le, L. Y. Shen, J. Perica, and I. M. C. S. Illankoon, "Life-cycle cost analysis of green-building implementation using timber applications," *Journal of Cleaner Production*, vol. 147, pp. 458–469, 2017.

[6] L. Klein, J. Y. Kwak, G. Kavulya et al., "Coordinating occupant behavior for building energy and comfort management using multi-agent systems," *Automation in Construction*, vol. 22, pp. 525–536, 2012.

[7] Z. Liu, K. Cheng, H. Li, G. Cao, D. Wu, and Y. Shi, "Exploring the potential relationship between indoor air quality and the concentration of airborne culturable fungi: a combined experimental and neural network modeling study," *Environmental Science and Pollution Research*, vol. 25, no. 4, pp. 3510–3517, 2018.

[8] H. Fu and X. Liu, "A study on the impact of environmental education on individuals' behaviors concerning recycled water reuse," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 13, no. 10, pp. 6715–6724, 2017.

[9] R. Volk, J. Stengel, and F. Schultmann, "Building information modeling (BIM) for existing buildings-literature review and future needs," *Automation in Construction*, vol. 38, pp. 109–127, 2014.

[10] R. Miettinen and S. Paavola, "Beyond the BIM utopia: approaches to the development and implementation of building information modeling," *Automation in Construction*, vol. 43, pp. 84–91, 2014.

[11] H. Fu and X. Liu, "Research on the phenomenon of Chinese residents' spiritual contagion for the reuse of recycled water based on SC-IAT," *Water*, vol. 9, no. 11, p. 846, 2017.

[12] F. Bosché, M. Ahmed, Y. Turkan, C. T. Haas, and R. Haas, "The value of integrating scan-to-BIM and scan-vs-BIM techniques for construction monitoring using laser scanning and BIM: the case of cylindrical MEP components," *Automation in Construction*, vol. 49, pp. 201–213, 2015.

[13] A. Yang, Y. Han, Y. Pan, H. Xing, and J. Li, "Optimum surface roughness prediction for titanium alloy by adopting response surface methodology," *Results in Physics*, vol. 7, pp. 1046–1050, 2017.

[14] E. Alreshidi, M. Mourshed, and Y. Rezgui, "Requirements for cloud-based BIM governance solutions to facilitate team collaboration in construction projects," *Requirements Engineering*, vol. 23, no. 1, pp. 1–31, 2018.

[15] Z. Liu, W. Xu, X. Zhai, C. Qian, and X. Chen, "Feasibility and performance study of the hybrid ground-source heat pump system for one office building in Chinese heating dominated areas," *Renewable Energy*, vol. 101, pp. 1131–1140, 2017.

[16] S. Malaikrisanachalee and H. Vathananukij, "Integration of Java-based BIM with spatial database," *International Journal of Civil Engineering*, vol. 9, pp. 17–22, 2010.

[17] K. Liu, H. Fu, and H. Chen, "Research on the influencing mechanism of traditional cultural values on citizens' behavior regarding the reuse of recycled water," *Sustainability*, vol. 10, no. 2, p. 165, 2018.

[18] C. W. Liao and Y. H. Perng, "Data mining for occupational injuries in the Taiwan construction industry," *Safety Science*, vol. 46, no. 7, pp. 1091–1102, 2008.

[19] R. K. Roul, S. R. Asthana, and G. Kumar, "Study on suitability and importance of multilayer extreme learning machine for classification of text data," *Soft Computing*, vol. 21, no. 15, pp. 4239–4256, 2017.

[20] A. M. Yang, Y. Han, S. S. Li, H. W. Xing, Y. H. Pan, and W. X. Liu, "Synthesis and comparison of photocatalytic properties for Bi_2WO_6 nanofibers and hierarchical microspheres," *Journal of Alloys and Compounds*, vol. 695, pp. 915–921, 2017.

[21] N. Shahid, M. U. Ilyas, J. S. Alowibdi, and N. R. Aljohani, "Word cloud segmentation for simplified exploration of trending topics on Twitter," *IET Software*, vol. 11, no. 5, pp. 214–220, 2017.

[22] G. Jia, Y. Lu, W. Lu, Y. Shi, and J. Yang, "Verification method for Chinese aviation radiotelephony readbacks based on LSTM-RNN," *Electronics Letters*, vol. 53, no. 6, pp. 401–403, 2017.

[23] Z. Liu, H. Li, K. Liu, H. Yu, and K. Cheng, "Design of high-performance water-in-glass evacuated tube solar water heaters by a high-throughput screening based on machine learning: a combined modeling and experimental study," *Solar Energy*, vol. 142, pp. 61–67, 2017.