Journal of Architectural Research and Development, 2024, Volume 8, Issue 6
https://ojs.bbwpublisher.com/index.php/JARD
Online ISSN: 2208-3537
Print ISSN: 2208-3529

Research on Big Data Coding System Based on the Classification of Artificial Materials and Mechanical Equipment in Construction Engineering

Zuguo Bai[1]*, Sheng Han[1], Zhan Wei[2], Jinping Chu[2]

[1]Hubei Urban Construction Vocational And Technological College, Wuhan 430205, China
[2]Hubei Construction Engineering Standard Quota Management Station, Wuhan 430205, China

***Corresponding author:** Zuguo Bai, hellobai@126.com

Abstract: By analyzing and comparing the current application status and advantages and disadvantages of domestic and foreign artificial material mechanical equipment classification coding systems, and conducting a comparative study of the existing coding system standards in different regions of the country, a coding data model suitable for big data research needs is proposed based on the current national standard for artificial material mechanical equipment classification coding. This model achieves a horizontal connection of characteristics and a vertical penetration of attribute values for construction materials and machinery through forward automatic coding calculation and reverse automatic decoding. This coding scheme and calculation model can also establish a database file for the coding and unit price of construction materials and machinery, forming a complete big data model for construction material coding unit prices. This provides foundational support for calculating and analyzing big data related to construction material unit prices, real-time information prices, market prices, and various comprehensive prices, thus contributing to the formation of cost-related big data.

Keywords: Data of labor; Materials; Equipment; Classification; Big data coding system

Online publication: November 27, 2024

1. Introduction

In the current era of "big cloud intelligent connection," the construction industry is also increasingly dependent on big data and information data, of which the "construction engineering artificial material equipment machinery data standard" is one of the most important basic data support. The standard aims to improve the efficiency and quality of modern computer applications in construction engineering. Its advantages include providing a set of unified static coding machines, using the method of line and surface combination, to realize the unified management of the information data of the machine throughout the project life cycle [1–3].

2. Analysis of the coding system

Foreign OmniClass, a classification system for the construction industry, has been widely adopted and applied in North America. IFC/IFD technology is the definition of building and construction engineering data, aiming to promote different specialties in the construction industry and different software in the same profession to share the same data source, and to achieve data sharing and interaction [2–4].

The development and application of foreign coding systems provide a good reference for the information development of China's construction industry. The national standard GB50851-2013 "Code for Data Classification and Coding of Artificial Materials and Machinery Equipment" is formulated based on the OmniClass standard, and the specific classification and coding numbers are adjusted. The basic principles and methods of information classification are proposed, which lays a theoretical foundation for the development of China's building coding system. The Ministry of Housing and Urban-Rural Development has also promulgated relevant standards and specifications, such as "Unified Standard for Application of Building Information Model" GB/T51212-2016, "Standard for Classification and coding of Building Information Model" GB/T51269-2017 and "Standard for storage of Building engineering Information Model" GB/T51447-2021, and many more [5]. The continuous improvement and upgrading of these coding technical standards, jointly promote the reform and upgrading of the coding system, realize the sharing and interaction of data, adapt to the growing market demand and technological development, improve the digitization level and information level of the construction industry, and promote the transformation and upgrading and sustainable development of the construction industry [4].

However, the existing coding system primarily involves issues such as serial number coding, the mixing of names and attributes, inability to generate codes automatically, non-convertible units, and the inability to automatically assign codes to new materials. Therefore, it is necessary to develop more effective coding methods to improve the efficiency and practicality of the labor and machinery coding system, align with the national cost reform policy, and adapt to a market-oriented economy.

3. Coding ideas and calculation model

3.1. Material attribute labeling

If a certain class of material has *n* feature items and *Y* represents the commonly used attribute values under a feature item, the material can be expressed as a vector with *i* rows and 2 columns. It is denoted as:

$$M = \begin{pmatrix} 1 & Y1 \\ 2 & Y2 \\ 3 & Y3 \\ . & . \\ . & . \\ n & Yn \end{pmatrix}$$

Where the first column represents the feature item, and the second column represents the number of attribute values under the feature item. For example, Y_1 means that the common attribute value of the first feature is Y_1, and so on. As shown in the example in **Table 1**, 0101 steel bars can be expressed as:

$$M = \begin{pmatrix} 1 & 6 \\ 2 & 6 \\ 3 & 29 \\ 4 & 20 \\ 5 & 11 \\ 6 & Yn3 \end{pmatrix}$$

Table 1. Examples of forward-coded feature items and attribute values

Category encoding and name	Eigenvalue serial number	Characteristic item	Number of attribute values	Attribute value 1	Attribute value 2	Attribute value 3	Attribute value 4	Attribute value 5	Attribute value 6	Attribute value 7	Total
0101 Rebar	1	Variety	6	General	Hot-rolled disc strips	Cold-rolled ribbed rebar	Threaded Rebar	Cold-rolled and twisted rebar	Reserved		$6 \times 6 \times 29 \times 20 \times 11 = 229{,}680$
	2	Levels	6	General	Level I	Level II	Level III	Reserved			
	3	Diameter	29	Comprehensive	Φ 4	Φ 4.5	Φ 5	Φ 5.5	Φ 6		
	4	Material	20	Comprehensive	Q195	Q215	Q235A	Q235B			
	5	Intensity rating	11	General	HPB235	HRB335	HRB400	reserve			
	6										

3.2. Calculation and significance of the quick count

By defining the quick calculation number, the ranking value of specific materials can be calculated quickly and conveniently. The quick number is the number of materials under a certain characteristic of that material, and also the spacing of materials with adjacent attribute values [5,6].

Define the quick number S_i: The quick number is calculated by multiplying the attribute values of the feature items in descending order of their sequence numbers. This is expressed as:

$$\text{Si} = \prod_{i=n}^{i} Y\,(\text{i}+1)$$

In particular, the quick number of the last feature is defined as 1, i.e., $S_n = 1$. For example, if a material vector of a particular property is represented as:

$$\text{M} = \begin{pmatrix} 1 & 6 \\ 2 & 6 \\ 3 & 29 \\ 4 & 20 \\ 5 & 11 \\ 6 & 3 \end{pmatrix}$$

Then the quick number vector of that material can be obtained as:

$$\text{S} = \begin{pmatrix} 114840 \\ 19140 \\ 660 \\ 33 \\ 3 \\ 1 \end{pmatrix}$$

In the example above, the quick calculation number is 114840, meaning that the variety of 0101 steel bars is hot-rolled discs, with 114840 kinds of materials available. The varieties of 0101 steel bars are hot-rolled disc bars of grade 1, totaling 19,140 types of materials. Additionally, there are 660 types of 0101 steel bars that are hot-rolled disc bars, grade 1, and have a diameter of 6mm.

For the last item in the material vector, regardless of the number of properties associated with the reserved characteristics, the quick number is 1. The quick number for the penultimate feature, "strength grade," is calculated as the number of attribute items under that feature (3) multiplied by the quick number of the subsequent feature (1):

$$3 \times 1 = 3$$

The quick calculation number for the third penultimate feature, "material," is the number of attribute items under that feature (11) multiplied by the quick number of the subsequent feature (3):

$$11 \times 3 = 33$$

Thus, the quick calculation numbers are as follows: 660 for diameter, 19,140 for grade, and 114,840 for variety.

3.3. Forward coding calculation model of coding value

If the unique attribute value of a particular material is represented as:

$$Z = \begin{pmatrix} Z1 \\ Z2 \\ Z3 \\ . \\ . \\ Zn \end{pmatrix} \text{ where } i \leq Y_i,$$

then the third column of the extended vector *M* can represent the unique attribute value positioning of a particular material. The specific material can be fully expressed as:

$$M = \begin{pmatrix} 1 & Y1 & Z1 \\ 2 & Y2 & Z2 \\ 3 & Y3 & Z3 \\ . & . & . \\ . & . & . \\ n & Yn & Zn \end{pmatrix}$$

It is clear that the total number of species for this class of materials is given by:

$$\prod_{i=1}^{n} Y_i \sum_{i=1}^{n} Si(\square)$$

For example, if the material attribute vector *C* for a particular attribute is represented as:

$$C = \begin{pmatrix} 1 & 6 & 3 \\ 2 & 6 & 4 \\ 3 & 29 & 5 \\ 4 & 20 & 6 \\ 5 & 11 & 7 \\ 6 & 3 & 2 \end{pmatrix}$$

Then the quick number vector of the material can be obtained as:

$$S = \begin{pmatrix} 114840 \\ 19140 \\ 660 \\ 33 \\ 3 \\ 1 \end{pmatrix}$$

The sequence number of the material can then be calculated as:

(Attribute Vector – 1) × Quick Number Vector

The calculation is as follows:

(3-1) ×114840+ (4-1) ×19140+ (5-1) ×660+ (6-1) ×33+ (7-1) ×3+ (2-1) ×1=289924..

The calculation process is shown in **Table 2**.

3.4. Coded backward analysis model for coded values

Similar to Section **3.3**, if a specific material is known to be coded as A and belongs to the material attribute matrix, the first attribute value number of the material can be obtained by dividing the number of codes by the first quick number and rounding up according to the reverse operation. The result of dividing the number of codes by the quick number is used to calculate the attribute value of the second layer, and this process continues

Table 2. Example of forward coding calculation procedure for coded values

Category coding	Category name	Characteristic item	Number of attribute values (PCS)	Quick count	A material eigenvalue explanation		Query feature serial number	Calculate the value	Row position	Coding	Remarks
		Variety	6	114840		Hot-rolled disc strips	2	114840			
		Levels	6	19140		Level I	2	19140			
0101	Rebar	diameter	29	660	Grade I, hot-rolled disc-bar Q235A rebar	Unlabeled diameter	1	0	0134013	0101134013X	4-bit category code +0000553+ 1-bit unit code
		Material	20	33		Q195	2	33			
		Strength rating	11	3		Unlabeled strength	1	0			
		Reserved	3	1			1	0			

until the code of the last layer is calculated.

$$M = \begin{pmatrix} 1 & Y1 \\ 2 & Y2 \\ 3 & Y3 \\ . & . \\ . & . \\ n & Yn \end{pmatrix}$$

The calculation example is as follows: Assuming that the attribute table and characteristic values of a certain type of material are shown in **Table 1**, the attribute value of a specific material coded as 113 can be obtained, and the calculation process is shown in **Table 3**.

3.5. Simplification and optimization of coded attribute values

Taking into account the various types and brands of workloads, the product models determined by different manufacturers are circulating in the market. This situation leads to a significant number of classifications based on the attribute values of specific feature items of the machinery. As a result, the "workloads metadata" must be marked in the feature items or attribute values to ensure usability. Therefore, it is essential to consider the comprehensiveness, accuracy, and ease of use of both the feature values and attribute values of the machinery [6–8].

Firstly, this coding system adds a specific attribute value of "synthesis" to each attribute value. When the user is unsure how to fill in the feature item, is unclear about the specific attribute value under a feature item, forgets to fill in the attribute value, or finds that the selected representative value is not fully included in the attribute value list, the "synthesis" option can be utilized [9].

Secondly, the representative value is selected as the specific attribute value, which can reduce the invalid input amount. For the non-representative attribute value, linear or non-linear parameters can be introduced, and the relevant attribute information can be obtained by calculating the method of inserting the value.

Thirdly, considering that there are inaccuracies in the selection of representative values, the positioning position of features and attributes, the selection and ranking order of representative attribute values need to be determined later according to the use frequency recorded by the system, to ensure that users can find the attribute faster and improve customers' sense of use [10].

4. Application of coding attributes

4.1. Screening and statistics of hierarchical attributes

Now, the attribute of a particular material is defined as, A_{ix}, where i is the sequence number of feature items and x is the sequence number of attribute values. The material can be expressed as:

$$M = \begin{pmatrix} 1 & Y1 & Z1 \\ 2 & Y2 & Z2 \\ 3 & Y3 & Z3 \\ . & . & . \\ . & . & . \\ n & Yn & Zn \end{pmatrix}$$

To calculate the number of a specific material (where the attribute is A_{ix}), the total is obtained by dividing the total number of the material by Y_i, denoted as:

Table 3. Coding inverse analysis model of coding values

Features	Code name	Number of attribute values	Quick count		Attribute value of the material		Redundant		Look up the property sheet of the material	Look up the property sheet of the material
			Formula	Results	Calculation formula	Results	Calculation formula	Results	Results Explained	Coding interpretation
Breed	X1	6	X2~X6 characteristic value concatenation	114840	The number of codes divided by the quick count is rounded up	1	The number of codes divided by the number of quick computations is complimentary	113	Value of the attribute numbered 1 in the variety: that is, synthesis	Steel with HRB500 strength grade and Q235A material without filling in the type, grade, and diameter of the steel bar
Levels	X2	6	X3 to X6 concatenation of eigenvalues	19140	The remainder of X1 is divided by the speed count and rounded up	1	Take the remainder of X1 and subtract it from the quick number,	113	Value of the attribute numbered 1 in the level: that is, synthesis	
Diameter	X3	29	X4 to X6 concatenation of eigenvalues	660	X2 Take the remainder divided by the speed count and round up	1	Take the remainder of X2 and divide it by the number of quick computations,	113	Value of the attribute numbered 1 in diameter: i.e., synthesis	
Material	X4	20	X5~X6 concatenation of eigenvalues	33	X3 Take the remainder divided by the speed count and round it up	4	X3 takes the remainder of the result and divides it by the quick number,	14	The value of the property numbered 4 in the material	
Strength class	X5	11	X6 to X6 eigenvalue concatenation	3	X4 Take the remainder divided by the speed count and round up	5	Take the remainder of X4 and subtract from the quick number,	2	The value of the attribute numbered 5 in the strength level	
Reserved	X6	3	/	1	/	3	/	0	Reserve the value of the property numbered 3 in the property	

$$\prod_{i=1}^{n} Y\,i/Yi$$

We can divide these codes into segments, each of which contains S_i numbered items:

$$\frac{\left(\frac{\prod_{i=1}^{n} Y\,i}{Yi}\right)}{Si}$$

The first of these materials is numbered (X-1) × Si + 1, and the last material is numbered X × Si. According to the database list, the material attribute values can undergo multi-level attribute screening to meet the needs of cost data and machine data statistics. Furthermore, we can utilize language expressions from computer programming statements to perform statistics, screening, expansion, and other data-related actions.

4.2. Set the weight parameters of the attribute value

Considering the complexity of the user, certain weight parameters can be set for the attribute value, to simplify and sort optimization processing in the later stage. The identity of the user is determined as the general user, supplier, consulting institution, etc., and the validity of the encoding and the meaning weight of the new encoding is judged by giving the user the weight of the identity. For example, you can give a specific attribute value of a material machine and "comprehensive" this attribute value of different weights, through the calculation of weight parameters, to get a material machine unit price [11].

5. Software development and model application

This model needs to rely on labeled material properties, so a material property list needs to be determined in advance. It can be implemented using database programs such as Matlab, MySQL, or natural language program files such as NLP. The development goal is to convert the previous manual coding into automatic coding through automated programs, thus greatly reducing the workload of manual coding. At the same time, the program is used for automatic error correction to form an accurate coding model [12]. Building on the need for an automated coding solution, the next steps involve establishing a comprehensive coding meta-system for the components of the labor and materials mechanism.

(1) Establish the coding meta-system for the components of the labor and materials mechanism, compile the "labor and materials machine metadata" coding table, expand the relevant XML and JavaScript (JS) files, and save the essential metadata in the database.

(2) Based on database technology, the development includes the material coding name normalization module, the coding calculation module, the eigenvalue calculation module, and the eigenvalue statistics module.

(3) Load the "machine metadata" coding table into the database through each module. Based on the positioning of the metadata in the matrix and the relevant attributes of the name, automatically encode the model components and perform reverse decoding calculations. According to the source of the data, assign different weights to various types of units for the given data input, and use data analysis to calculate the machine code [13–15].

(4) Establish the database file for the coding and unit price of the machine to create a comprehensive big data model of the machine code unit price. This will provide essential support for various big

data applications, including statistical analysis, unit price calculations, and real-time market price evaluations.

Of course, in developing a new coding system, it is essential to fully consider the actual conditions of project construction and to widely solicit opinions and suggestions from all stakeholders. This approach will ensure that the new coding system can be effectively utilized and promoted. Simultaneously, the management and maintenance of the new coding system should be strengthened, with timely updates and improvements made to ensure its effectiveness and practicality.

6. Closing remarks

This paper constructs a set of standard mathematical models for the data coding of worksheets and develops a comprehensive coding system. The coding scheme enables the automatic calculation of worksheet codes through forward coding, facilitating horizontal feature exploration and vertical penetration of attribute values. This approach meets the needs for slicing analysis and statistical calculations of worksheet data using big data in the future.

Additionally, the coding scheme allows for reverse analysis; by examining the coded values, one can identify the names of materials, feature items, and attribute values, ultimately determining the specifications and models of materials. This capability significantly aids the engineering industry in conducting big data statistics and analyses of material machinery. However, due to limitations in personal knowledge and experience, the model presented in this paper may contain some inaccuracies or gaps.

Through the analysis, application, and research of artificial materials and mechanical equipment, this paper establishes a standard mathematical model and coding calculation system, resulting in a unified data interaction solution.

Funding

Research project of the Construction Department of Hubei Province (Project No. 2023-64).

Disclosure statement

The authors declare no conflict of interest.

References

[1] Ministry of Housing and Urban-Rural Development, 2013, Data Standard for Artificial Material Equipment and Machinery for Construction Engineering GB/ T50851-2013. Chinese Standard.

[2] Ni J, Li J, Shang L, et al., 2016, Application and Promotion of Data Standard of Artificial Material Equipment and Machinery for Construction Engineering GB/ T50851-2013. Construction Science and Technology, 310(07): 47–48.

[3] Fu Y, 2012, Data Standard and Application Of Artificial Material Equipment and Machinery in Construction Engineering. China Construction Information, 515(20): 44–45.

[4] Sha P, 2018, Research on the Type and Coding of IFD Standard Expansion in BIM Design of Railway Bridge. Railway Standard Design, 62(11): 77–81. https://doi.org/10.13238/j.Issn.1004-2954.201801100004

[5] Ministry of Housing and Urban-Rural Development of the People’s Republic of China. Classification and Coding Standard of Building Information Model: GB/T 51269-2017. China Building Industry Press, Beijing, 2017: 4–6.

[6] Wang C, 2022, Research and Software Development of BIM Basic Data Production System for Engineering Construction. Electronic Technology and Software Engineering, 2022(16): 63–66.

[7] Yang X, Gao NP, Huang J, 2023, Research on Building Construction Cost Documentation Based on Building Information Classification System. Construction Cost Management, 2023(05): 19–24.

[8] Cai L, Wang C, 2022, Research on Identification Coding Structure of Highway Engineering Information Model. Journal of Shenzhen University (Science and Technology), 39(04): 424–431.

[9] He L, 2024, Research on Optimization strategy of Engineering Cost Management based on ERP database. Engineering Cost Management, 35(01): 30–36.

[10] Fan QX, Lu YM, Zhou SW, et al., 2019, Research and Practice on Intelligent Construction Technology System of Jinsha River Hydropower Project. Journal of Hydraulic Engineering, 50(03): 294–304.

[11] Huang Y, Du J, Huang S, 2023, Construction of Data Management System for Iron and Steel Business Owners. Shandong Metallurgy, 45(02): 53–55.

[12] Liu Z, Zhou Y, 2021, Research on Science and Technology Big Data Resources And Classification. Science and Technology Innovation, 2021(18): 123–126.

[13] Li X, Wang Y, 2004, Price Informatization of Construction Materials, Machinery and Labor. Jilin Building Materials, 2004(06): 43–44.

[14] Ma M, 2011, Discussion on the Dynamic Price Management of Construction Engineering Materials. Science and Technology Vision, 2011(03): 150–151.

[15] Ou Q, Li X, 2020, Research on Development of Price Information Compilation System of Construction Engineering Materials. Journal of Literature and Literature, 2020(41): 195.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.