Research on Classification and Hierarchical Management of Industrial Data Based on Entropy Method-TOPSIS Theory

Jian Xu1, Ren Qiang2, Ziqi Zhou1, Wei Luo1*, Qinglin Yuan2 and He Cai2

1CHN Energy Dadu River Big Data Services CO., LTD, CHN Energy Dadu River Hydropower Development CO., LTD, Chengdu, Sichuan, 610041, China
2Beijing Engineering Research Center of Energy Electric Power Information Security, North China Electric Power University, Changping District, Beijing, 102206, China

*Corresponding author’s e-mail: wei.luo.ap@chnenergy.com.cn

Abstract. In the era of accelerating digitalization, digital infrastructure is one of the most popular topics today, and digital infrastructure dominated by the industrial sector has become the vanguard of development. In response to this, industrial companies and platform companies attach importance to industrial data. Increasingly, the classification and hierarchical management of industrial data has become an indispensable procedure for the effective processing and safety management of industrial data. This article classifies the vertical data of each domain from the general process of industrial production, namely R&D, production, operation and maintenance, management and other classification domains, and according to the industrial production, self-economic benefits and social benefits brought by industrial data. Potential impact on industrial data can be classified by horizontal data classification of primary data, secondary data, and tertiary data from data features such as data importance, tamper-proof, and anti-corruption, and use entropy weight method to calculate the weight of each indicator feature and evaluate the data importance, and finally apply TOPSIS theory for data grading.

1. Introduction
With the sudden emergence of the "digital new infrastructure" development strategy, the application of big data in the industrial field has become more and more extensive, and the digitization, intelligence, and networking of industrial production and manufacturing have become an important carrier for the development of digital infrastructure. Under this new development model and new format[1], industrial data is developed around the business process of the company from product research and development design, manufacturing, operation management, application services[2], etc. During the circulation of industrial data in the industrial production process, due to its data characteristics such as the expansion of the data volume, the variety of names and purposes, the wide range of collection, the diversification of flow directions and transmission paths, and the diversity of storage locations[3]. The classification and hierarchical management of data has become the top priority to ensure data security and ensure the production safety of industrial enterprises[4].

In 2017, the report of the 19th National Congress of the Communist Party of China proposed that the building of a manufacturing power should be accelerated, and the development strategy of the deep integration of the Internet, big data, artificial intelligence and the real economy should be promoted, and in the second collective study of the Politburo of the 19th CPC Central Committee, General Secretary Xi Jinping emphasized the important concept of “to implement the national big data strategy and promote the innovation and development of the big data technology industry”. It can be seen that the
industrial development in the era of big data has become the trend of the times, and the application of data and the widespread dissemination of data have become an irreversible general trend[5]. After that, the Fourth Plenary Session of the 19th Central Committee of the Communist Party of China proposed for the first time that “data” is used as a production factor to participate in the distribution. Therefore, the classification and hierarchical management of industrial data has become an essential way to effectively mine the value of data and realize the transformation of enterprise production methods. As early as 2015, the State Council issued the "Outline of Action to Promote the Development of Big Data", which has clearly defined the need to develop and establish a data classification catalog and other standard and normative systems. The "Big Data Industry Development Plan (2016-2020)" and "Industrial Control The Guidelines for System Information Security Protection" clearly require enterprises to categorize data information. In 2018, the national standard "Data Management Capability Maturity Assessment Model" (DCMM) was released, requiring that all data be classified above the "managed level". Since then, in 2020, the Standing Committee of National People's Congress has successively issued the "Data Security Law (Draft)" and "Personal Information Protection Law (Draft)", indicating that there are clear new requirements for data security at the national level. In March of the same year, the Ministry of Industry and Information Technology issued the "Industrial Data Classification and Grading Guidelines", which further clarified the data generated and used by relevant industrial enterprises and platform enterprises in product design and development, manufacturing, management and operation, operation and maintenance services, and industrial Internet platform enterprises in equipment access, platform operation, industrial software applications, etc. The data generated and used in the process should be classified[6] and graded. In view of this, on the basis of analyzing a series of security problems faced by industrial data protection, this article further explores the classification method of industrial data, and classifies the importance of data according to the entropy method-TOPSIS theory, in order to achieve industrial data security control and protection.

2. Security Issues Facing Industrial Data Protection

Data security is related to the survival of an enterprise. Once a security issue occurs, it will be an unlimited responsibility to hold accountability. Paying attention to the development of data security work, that is, standardizing data management, improving data security protection capabilities, and preventing data security risks are indispensable[7]. For industrial data in the entire production process of an enterprise, data security control is a necessary link for the normal operation of the enterprise production and data exchange process. For data security issues, it can be summarized as data security management information asymmetry, data management personnel are lagging behind in obtaining information, and the data management and control system is not yet complete, data collection and identification are difficult, data transmission, monitoring and traceability are not convenient enough, data storage and classification and hierarchical management work are still not perfect, data security protection core technology cannot be mastered, data application process Credibility and sharing are difficult to ensure all items.

- Data security management information is asymmetry, data management personnel are lagging behind in obtaining information, and the data management and control system is not yet complete. While industrial enterprises have gradually improved their awareness of data security protection, the actual application situation has not been satisfactory. More industrial enterprises have focused on project-level data management, but their awareness of overall data is still relatively weak. Most companies are still stuck in the simple storage and management of data, and have not yet formulated a comprehensive and complete data security management plan, and have failed to make the management and control of data a daily work arrangement for the company.
- Data collection and identification are difficult. The distribution of industrial data is very extensive, and a large number of industrial systems and equipment have huge Data set. In addition, due to the non-standard and inconsistent data interfaces of various industrial
manufacturers, the resulting data island phenomenon is extremely severe, causing the problem of data difficult to identify.

- Data transmission monitoring and traceability are not convenient. Today's industrial system cannot be separated from big data, artificial intelligence, and cloud computing. With the application of many emerging technologies, industrial data is more complex and diversified in the process of circulation. In view of this, it is difficult to effectively capture abnormal or sensitive data in a large-volume virtualized environment.

- Data storage and classification and hierarchical management are not perfect. The data storage process involves data collection and screening, so it is necessary to divide the database or domain according to the type and level of the data, and set the data access key to improve the data protection level. However, due to the diverse forms and complex formats of industrial data, it is difficult to carry out the classification and hierarchical management of data.

- The core technology of data security protection has not been fully grasped. From the perspective of industrial system protection, traditional industrial system data security protection technical means are not practical enough, and the application of key technology products is limited. Key technologies such as credible defense, lightweight encryption, data desensitization, and data traceability for industrial data have not yet been tackled.

- It is difficult to ensure the credibility and sharing of the data application process. The analysis and utilization of industrial data is an important way for industrial development, and data rights and responsibilities, security and credibility, and other issues have effectively hindered the sharing of data.

3. Industrial data classification and hierarchical management method

Industri al data has attracted great attention from many parties because of its huge value. In the current stage of data interoperability and information interconnection, various industrial enterprises and enterprise platforms are facing security threats such as data tampering, data theft, and data leakage all the time. The extremely high commercial value is related to the production and operation of the enterprise. Once the above-mentioned threats are encountered, it will not only affect the production and operation safety of the enterprise, but may even threaten the national security[8]. In addition, due to the huge amount of industrial data and the diversity of forms, it is difficult to protect data security. In view of this, the classification and hierarchical management of data has not only attracted widespread attention at the enterprise level, but also has issued numerous laws and regulations at the national level to comprehensively restrict enterprise data management to ensure the overall security of industrial data.

3.1. Industrial data classification method

Industrial data has its own characteristics such as complexity and difference, which makes it difficult for enterprises to manage data. From the perspective of its form, industrial data has multiple forms such as time series, non-time series, structured, unstructured, etc., and the carrying information and application fields of different data are also different, and the requirements for real-time, continuity, and stability are quite different. From the perspective of importance, industrial data can be divided into general data suitable for cloud access and sharing, important data related to enterprise product quality, business interests, production safety, and sensitive data that affect national security. From the perspective of data flow, industrial enterprises and platform enterprises in their internal product development, production, operation and maintenance, management and other links constitute the enterprise production process, upstream and downstream enterprises, platforms, or equipment manufacturers, industrial enterprises, platform enterprises, service providers the flow and circulation[9] of data must be carried out between. There are many prerequisites for data management, for the development and utilization of industrial data, it is necessary to comprehensively sort out and understand the existing industrial data, carry out data integration work and conduct correlation analysis, and finally form a high-quality data source for data production decision-making, thereby laying the foundation for the improvement of enterprise production efficiency and industry innovation. Regarding the circulation
and sharing of industrial data, first of all, it is necessary to vigorously promote its enterprises to use the cloud, and determine the scope of data sharing to break the industrial data island, so as to further accelerate the transformation of enterprise production to intelligent, digital, and convenient.

The classification of industrial data should be combined with the production and manufacturing model and service operation model. Considering the actual situation, the data of industrial enterprises and platform enterprises should be merged and integrated, and the secondary sub-categories should be subdivided to form a list of enterprise industrial data classifications. For example, industrial enterprises can be divided into R&D domains, production domains, operation and maintenance domains, management domains, and external domains. Platform companies can be divided into platform operation data domains and enterprise management data domains. The reference table of enterprise and platform enterprise data classification is shown in table 1 and table 2.

Table 1: Industrial Enterprise Data Classification

| Data field          | Data Classification                  | Data Sources                                                                 |
|---------------------|--------------------------------------|-------------------------------------------------------------------------------|
| R&D domain          | R&D design data                      | Computer-aided system (CAD), Engineering simulation analysis system (CAE), Industrial software development system, Industrial system testing tools, etc. |
|                     | Development design data              |                                                                               |
| Production domain   | Control information                  | Manufacturing execution system (MES), Programmable logic controller (PLC), Data acquisition and monitoring system (SCADA), Distributed control system (DCS), Working condition database, etc. |
|                     | Industrial control status            |                                                                               |
|                     | Process parameters                   |                                                                               |
|                     | System log                          |                                                                               |
| Operation and       | Logistics data                       | Product logistics system, Product after-sales status tracking system, After-sales service management system, etc. |
| maintenance domain  | After-sales maintenance data         |                                                                               |
| Management domain   | System equipment asset information   | Product life cycle management system (PLM), Supply chain management system (SCM), Quality management system (QMS), Enterprise resource planning system (CRM), Warehouse management system (WMS), etc. |
|                     | Customer and product information     |                                                                               |
|                     | Product supply chain data            |                                                                               |
|                     | Business statistics                  |                                                                               |
| External domain     | Data shared with other subjects      | Access to supply chain systems and collaborative R&D systems of other companies. |

Table 2: Classification of platform enterprise data

| Data field             | Data Classification      | Data Sources                                                                 |
|------------------------|--------------------------|-------------------------------------------------------------------------------|
| Platform operation     | IoT data collection      | The production data obtained by the platform from the customer's industrial control system and the monitoring data collected through peripheral sensors |
| domain                 |                          |                                                                               |
| Customer application   | Customer application     | Customers rent data generated by various systems running independently of platform computing, storage and other resources |
| system data            | system data              |                                                                               |
| Knowledge base, model  | Knowledge base and       | Knowledge base and mechanism model library provided by PaaS layer |
| library data           | mechanism model library  |                                                                               |
| Analyze data           | Results, reports and     |                                                                               |
|                        | other data generated by  |                                                                               |
the platform through big data analysis

| Platform configuration data | Configuration data of platform users, equipment, application services, PaaS capabilities, etc. |
| Platform application data | Carrying data generated and applied by industrial APP |
| Development data | The development source code, open source tools, commercial tools, and test cases accumulated during the development of the platform |
| Technical and management data | Platform architecture design, software development, interface design and other process technical specifications, security software management, configuration and patch management, physical and environmental security management, identity authentication management, access security management, personnel management, testing and emergency response, asset management and other system documents |
| Enterprise Management Domain | Customer and solution data |
| Customer and solution data | The various data in the platform customer and product management system mainly include the customer’s basic information, behavior characteristics, usage records, customer service and maintenance records, etc., as well as the complete set of solutions customized and deliverable by the platform for customers |
| Business cooperation data | Various data in the platform enterprise business management system, mainly including data on strategic agreements signed with customers, product purchase and sales contracts, etc. |
| Personnel financial data | Various data in the personnel management system of the platform enterprise, including basic employee information, salary information, asset ledgers, financial statements, audit information, etc. |

3.2. Industrial data hierarchical management method

In the "Industrial Data Classification and Grading Guidelines", it is emphasized that the industrial data is classified into the first level according to the potential impact that different types of industrial data may have on industrial production, economic benefits, etc. after being tampered with, destroyed, leaked or illegally used, Level 2 and Level 3.

For the first-level data, its division has little impact on the normal production and operation of industrial control systems and equipment, industrial Internet platforms, etc.; the negative impact on the company is small, or the direct economic loss is small; the number of affected users and companies is small, the production and living areas are small, and the duration is short; the cost of restoring industrial data or eliminating negative impacts is small these four methods. Regarding the secondary data, its classification is likely to cause large or major production safety accidents or environmental emergencies,
which may cause large negative impacts on the enterprise, or large direct economic losses; the cascading effect caused is obvious, and the scope of impact is involved multiple industries, regions, or multiple companies in the industry, or the impact lasts for a long time, or may lead to the illegal acquisition of a large number of suppliers and customer resources or the disclosure of a large amount of personal information; the cost of restoring industrial data or eliminating negative impacts is relatively high these three methods. For the three-level data, it can be divided into two types that are prone to cause particularly serious production safety accidents or sudden environmental incidents, or cause particularly huge direct economic losses; and cause serious impacts on the national economy, industry development, public interests, social order, and even national security.

The hierarchical management of industrial data is divided into five parts: government responsibilities, enterprise responsibilities, hierarchical protection, information sharing, and emergency response. The government has overall guidance and coordination, and clarifies territorial management tasks; enterprises establish systems, establish relevant departments and institutions, and arrange professions Personnel conduct regular review; the hierarchical protection function can resist different intensities of external attacks; for information sharing, the sharing control is mainly for the first and second categories of data classification; emergency management can be regarded as the backup function of industrial data protection, once activated Emergency management procedures, immediate disposal and supplementary reporting of disposal conditions[10].

4. Industrial data importance hierarchical management evaluation method
The importance level of industrial data is determined by the following process. The project evaluation index system is established, each index is quantified, the index value is obtained, and then the index value is data processed, and the objective weight is determined by the entropy weight method, and finally obtained through the TOPSIS theory The progress of the project is posted, so as to finally clarify the heavy classification of industrial data. The evaluation process is shown in Figure 1.

![Evaluation process of the program selection](chart.png)
4.1. Construct a judgment matrix and perform data processing

(1) Construct a judgment matrix. Assuming that the m preferred schemes in the original data are judged by n evaluation indicators, a judgment matrix with m rows and n columns is formed, which is denoted as:

\[
X = \begin{bmatrix}
    x_{11} & x_{12} & \cdots & x_{1n} \\
    x_{21} & x_{22} & \cdots & x_{2n} \\
    \vdots & \vdots & \ddots & \vdots \\
    x_{m1} & x_{m2} & \cdots & x_{mn}
\end{bmatrix}
\]  

(1)

In the formula, \(x_{ij}\) represents the j-th evaluation index of the i-th plan to be optimized, where \(i = 1, 2, \ldots, m; j = 1, 2, \ldots, n\).

(2) Normalization of evaluation indicators. The evaluation indicators for the priority classification and optimization mainly include three types of indicators, namely the completeness of the data, the reliability and the density of the data.

(3) Dimensionless processing of data. Since there are many evaluation indicators and the units are not uniform, it is necessary to use the efficiency coefficient method to perform non-dimensional processing on the original data. The calculation formula is as follows.

\[
y_{ij} = \frac{x_{ij} - \min_{1 \leq l \leq m} (x_{lj})}{\max_{1 \leq l \leq m} (x_{lj}) - \min_{1 \leq l \leq m} (x_{lj})}
\]

(2)

Among them, \(y_{ij}\) is the dimensionless value of the index \(x_{ij}\).

4.2. Entropy weight method objectively determines the index weight method

According to the explanation of the basic principles of information theory, information is a measure of the degree of order of the system, and entropy is a measure of the degree of disorder of the system; according to the definition of information entropy, for an index, entropy can be used to judge the degree of dispersion of an index, The smaller the information entropy value, the greater the degree of dispersion of the index, the greater the impact of the index on the comprehensive evaluation (ie, the weight), if the values of an index are all equal, the index will not play a role in the comprehensive evaluation. Therefore, the tool of information entropy can be used to calculate the weight of each indicator to provide a basis for comprehensive evaluation of multiple indicators.

(1) Entropy calculation.

\[
e_j = -\frac{1}{\ln(m)} \sum_{i=1}^{m} (Q_j \times \ln(Q_j))
\]

(3)

\[
Q_j = \frac{x_{ij}}{\sum_{i=1}^{m} x_{ij}}
\]

(4)

Among them, \(e_j\) is the entropy value of the evaluation index \(j\); \(Q_{ij}\) is the proportion of the i-th evaluated object under the j-th evaluation index.

(2) Calculate the weight \(\omega_2\).

\[
\omega_2 = \frac{(1-e_j)}{\sum_{j=1}^{n} (1-e_j)}
\]

(5)

Among them, \(\omega_2\) is the weight value of the entropy method of the evaluation index \(j\).

4.3. TOPSIS theoretical method

The TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution) method was first
proposed by C. L. Hwang and K. Yoon in 1981. The TOPSIS method ranks according to the closeness of a limited number of evaluation objects to the idealized target. The relative pros and cons of the objects are evaluated. The basic principle is to calculate the optimal solution and the worst solution, and then sort according to the distance between each solution to be optimized and the optimal solution and the worst solution. If the solution to be optimized is the closest to the optimal solution and the farthest away from the worst solution The solution is the optimal solution. The theoretical steps are as follows:

1. Construct a weighted judgment matrix.

\[ Z = (z_{ij})_{m \times n} \]  

\( Z \) represents the weighted judgment matrix, and has

\[ z_{ij} = y_{ij} \omega \]  

(2) Determine the positive ideal solution \( Z^+ \) and the negative ideal solution \( Z^- \). The maximum value of each evaluation index after all normalization constitutes a positive ideal solution, and the minimum value of each evaluation index after all normalization constitutes a negative ideal solution, which can be denoted as,

\[
\begin{align*}
Z^+ &= (z^+_1, z^+_2, \ldots, z^+_n) \\
Z^- &= (z^-_1, z^-_2, \ldots, z^-_n) \\
z^+_j &= \max(z_{ij}), z^-_j = \min(z_{ij})
\end{align*}
\]  

Among them, \( Z^+ \) indicates the maximum value of the \( j \)-th evaluation index; \( Z^- \) indicates the minimum value of the \( j \)-th evaluation index.

(3) Determine the distances \( D^+_i \) and \( D^-_i \).

\[
\begin{align*}
D^+_i &= \sqrt{\sum_{j=1}^{n} (z_{ij} - z^+_j)^2} \\
D^-_i &= \sqrt{\sum_{j=1}^{n} (z_{ij} - z^-_j)^2}
\end{align*}
\]  

The above formula shows the distance \( D^+_i \) from the positive ideal solution \( Z^+ \) and the distance \( D^-_i \) from the negative ideal solution \( Z^- \) for each preferred solution.

(4) Determine the relative posting progress \( C'_i \) of the plan to be optimized.

\[ C'_i = \frac{D^-_i}{D^+_i + D^-_i} \]  

(5) Sort the preferred solutions according to the value of \( C'_i \). The larger the value of \( C'_i \), the better the solution.

4.4. Construction of industrial data importance grading index evaluation system

Industrial data importance grading is mainly considered from three types of criteria: data integrity indicators, data reliability indicators, and data confidentiality indicators. Through specific analysis from three different indicator dimensions, an industrial data importance grading evaluation index system is established, as shown in Figure 2.
Industrial Data Importance Grading Evaluation Index System

Data Integrity Index Q1
- Redundancy X1
- Storage rate X2
- Degree of polymerization X3

Data reliability index Q2
- Logical order X4
- Structured rate X5

Data Confidentiality Index Q3
- Share rate X6
- Tampering rate X7
- Key Level Index X8

Figure 2. Industrial Data Importance Grading Evaluation Index System

Among them, data integrity index $Q_1$ mainly includes data redundancy ($X_1$), data storage rate ($X_2$), data aggregation degree ($X_3$); data reliability index $Q_2$ mainly includes data logical ordering degree ($X_4$), structure rate ($X_5$); data confidentiality index $Q_3$ mainly includes data sharing degree ($X_6$), data tampering rate ($X_7$), key level index ($X_8$). Through the calculation of the weight of the above quantitative indicators and the TOPSIS theory, the grading level of the importance of industrial data can be determined.

5. Conclusion

Data classification and grading is an important foundation for establishing a data lifecycle security protection system, and it is also one of the core tasks of data security governance. Data classification protection and grading protection complement each other. Data classification combines data with common attributes. Data grading constructs a technical protection system based on the consequential standards of data, and finally incorporates it into the corresponding data grading system through data categories. On the basis of hierarchical classification, specific requirements for data protection, setting of different access rights, encrypted storage and transmission of important data, desensitization processing of sensitive data, audit records and analysis of important operations, etc., it can form an effective data security protection system. This article first introduces the security issues faced by industrial data protection, interprets related security issues, and clarifies the importance of industrial classification and hierarchical management at this stage, and then uses the entropy method-TOPSIS theory to explain the importance of industrial data evaluation methods to further determine industrial data the importance level of the enterprises is conducive to the safety management and protection of the industrial data of various enterprises. For future data protection work, we should further upgrade the classification and grading of industrial data and security management and control under the guidance of existing policies, so that industrial data can better serve the rapid development of modern society.
Acknowledgments
We would like to express our sincere gratitude to everyone who helped in this research work. This project is supported by Hydropower big data classification and hierarchical management system research project (DSJ-KY-2021-003).

References
[1] Chen X.H., Yang S.F., Zhang X.Y. (2021) A Brief Talk on Industrial Internet Data Security Protection. J. Automation Expo., 38: 15-17.
[2] Zhang X.Y., Yang S.F., Wang C.H., Li J., Chen X.H. (2021) Research on Industrial Internet Data Security Classification and Hierarchical Protection Framework. J. Information Technology and Network Security., 40: 2-9.
[3] Wang W.L. (2021) Industrial Data Governance Based on Value Chain: Model Construction and Practice Direction J. Science and Technology Management Research., 40: 233-239.
[4] Wang D.Y., Wang J.X., Li Y.S., Meng K. (2020) Analysis and research on the application of classification and classification of industrial data in the construction machinery industry J. Construction Machinery, 09: 8-11.
[5] Guo M. (2020) Industrial data grading method based on risk assessment J. Instrumentation Standardization and Metrology, 04: 25-28.
[6] The Ministry of Industry and Information Technology (2021) Guidelines for Classification and Classification of Industrial Data (Trial) J. Automation Expo, 38: 5.
[7] Wang Q.M. (2020) Give full play to the engine role of data resources J. China Science and Technology Awards, 06: 10-11.
[8] Wang Y, Xu J.X. (2020) Interpretation of industrial safety protection under the guidance of "Guidelines for Classification and Classification of Industrial Data (Trial)" J. Automation Expo, 37: 52-54.
[9] Ministry of Industry and Information Technology (2020) Strengthen the classification and hierarchical management of industrial data and create a vigorous future for the digital economy J. Cybersecurity and Informatization, 04: 4-6.
[10] Chen X.H., Yang S.F., Liu C.Y. (2019) Thoughts on Industrial Internet Data Security Classification and Grading J. Network Security and Informatization, 08: 112-114.