

Risk assessment of data asset management based on fuzzy hierarchy method and association rules

Shuguo Chen^{1*}, Yaomao Wu¹, Yue Fu¹, Yihong Zhang²

¹State Grid Zhejiang Electric Power CO., LTD., Hangzhou, Zhejiang, 310007, China

²State Grid Zhejiang Economic Research Institute, Hangzhou, Zhejiang, 310012, China

Abstract: In the current data asset management risk assessment, the processing of assessment indicators is relatively simple, which leads to large errors in the assessment results. To this end, a data asset management risk assessment based on fuzzy hierarchy method and association rules is proposed. Identify risk factors for data asset management. Construct a risk assessment model. Quantify the model indicators based on the fuzzy hierarchy method, generate a judgment matrix and derive a comprehensive assessment vector. Assess risk levels based on association rules. Experiments show that the risk level assessment of the method has a correct rate of 100% and the assessment error of the risk score is 1.86 points, which has high application value.

1. Quotes

In the course of business operations, not only are capital assets involved, but all the data generated in the process is also an important asset. Data and the information it generates are important assets for companies. By using data assets, companies can make innovative use of data information and gain an advantage over their competitors. By managing data assets in a high quality manner, companies can reduce the cost and risk of their operations [1]. With the development of the economy and society, more and more enterprises are paying attention to the management of data assets. However, in the process of data asset management, there are certain risks. The process of data asset management includes the control and protection of data, and high-quality data asset management can substantially increase the actual value of data assets. The fuzzy hierarchy approach enables the risks in data asset management to be assessed effectively. The different levels of management risk are divided into different hierarchical structures according to certain rules. Assessment indicators for risk assessment are determined through a combination of qualitative and quantitative

approaches [2]. The indicator is set to give a quantitative reference value. The correlation method allows for better application of the assessment model through intrinsic linkages. By correlating the parameters in the data asset assessment model with the risk factors in the data asset management process, an assessment level for the risk factor can be obtained [3]. Based on the fuzzy hierarchy method and the law of association, the assessment process of data asset management is completed. This assessment method not only refines the content of the risk assessment, but also improves the degree of correspondence of the risk assessment and completes a high-quality risk assessment of data asset management.

2. Identify data asset management risk factors

In the risk assessment of data asset management, it is first necessary to identify what risks are present in the process. These risk factors are then identified according to the inherent characteristics of the different risk factors [4]. The management of data assets is divided into four main components as shown in the Figure 1:

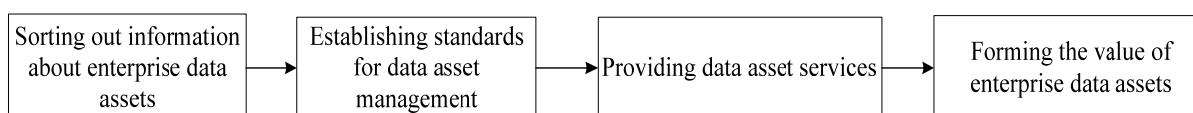


Figure 1. Data asset management process

As shown in the diagram above, there are a number of management risks in each of the four processes of data asset management. The management risk factors present in each management process are shown in the Table 1:

Table 1. Data asset management risk factors	
Steps	Risk factors
Sorting out information about enterprise data assets	Accuracy of data asset grooming methods Data Asset Storage System

*E-mail: zhang_yihong@sgcc.com.cn

Security	
Establishing standards for data asset management	Stability of data asset valuation guidelines
Providing data asset services	Customer Creditworthiness Completeness of client information
Forming the value of enterprise data assets	National policy stability Competitive market environment

A qualitative identification approach is then used to carry out the identification of risk factors for data asset management in a broad framework. Qualitative identification focuses on critical values in data asset management [5]. After a certain threshold has been exceeded, it is considered a more serious management risk. It is calculated by the formula:

$$\sum_{i=1}^N \frac{q_i}{Q_i} \geq 1 \quad (1)$$

In equation (1), N is the number of types of risk factors present in data asset management. i is the different data asset information. q_i is the actual amount of data information stored in the different data asset management sessions. Q_i is the risk threshold amount in each data asset management. Using the above formula, risk factors are identified through qualitative identification. In the actual risk assessment of data asset management, a combination of quantitative and qualitative identification is required. The identification of risk factors is done both from a detailed classification and from a broad framework of risks, thus completing the identification of risk factors for data asset management.

3. Building a data asset management risk assessment model

Based on the risk factors identified above, a model for the risk assessment of data asset management is constructed. This risk assessment model is used to describe the interrelationship between the variables in data asset management. Let the risk of data asset management be P, then this paper designs the assessment model with the formula:

$$P = \left(\sum_{i=1}^n R \times W \right) / n \quad (2)$$

In equation (2), R is the benchmark management risk of the ith data asset. W is the correction factor of the model. n is the number of risk levels set by the model. A hierarchy of risk assessment indicators and a hierarchy of risk indicators is established, as shown in the Table 2:

Table 2. Hierarchy of data asset management assessment indicators

Target level	Middle layer	Indicator layer
Data Asset Management Assessment	Basic indicators	Data type
	Quality indicators	Generation method
		Integrity
	Market Indicators	Accuracy
		Data activity
		User Comments

Based on the hierarchy of different assessment indicators shown in the table above, a hierarchy of risk prompts in a recursive structural relationship is established. The set of rubrics for the prompts, V, is:

$$V = \{5, 4, 3, 2, 1\} \quad (3)$$

In equation (3), the size of the tagline set is ranked according to the risk severity level, i.e. 5 is significant risk and 1 is no risk. After completing the setting of indicators for the data asset management risk assessment model, the weights of the model are calculated. Based on the hierarchy of assessment indicators, the priority relationship weight matrix a is constructed with the formula:

$$a = \sum_{i=1}^n \frac{a_i}{2n} + 0.5 \quad (4)$$

In equation (4), a_i is the indicator weight of the ith data asset. Then the weight parameter w of the risk assessment model designed in this paper is calculated as:

$$w = \frac{1}{n} - \frac{1}{2a} + \frac{1}{na} \sum_{i=1}^n a_i \quad (5)$$

Based on the calculation results of equation (5), the weights of this data asset management risk assessment model are completed to complete the initial construction of the data asset management risk assessment model.

4. Quantification of model indicators based on the fuzzy hierarchy method

After completing the initial construction of the above data asset management risk assessment model, the indicators of the model need to be quantified to complete a more refined risk assessment. In this paper, the fuzzy hierarchy method is used to quantify the indicators of the assessment model. The quantification process of the indicators of the fuzzy hierarchy method is shown in the Figure 2:

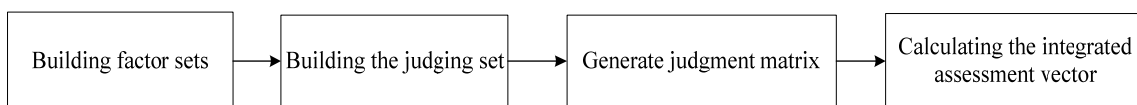


Figure 2. Quantification process of fuzzy hierarchy method indicators

As shown above, the set of factors U for the fuzzy hierarchy method is first established as:

$$U = \{u_1, u_2, \dots, u_n\} \quad (6)$$

In equation (6), $u_1, u_2, \dots, u_n$ are the various factors that influence the risk assessment results respectively. Weights are set for the set of factors. This is because the individual factors in the factor set do not all have the same degree of influence on the assessment results. In order to reflect the degree of influence of the different influencing factors, a corresponding weight L_n is set for each influencing factor, the set of component weights L is:

$$L = \{L_1, L_2, \dots, L_n\} \quad (7)$$

Once the set of factor weights has been set, each parameter in the weight set is tested to see if it meets the criteria of non-negativity and homogeneity. Generate a judgement matrix Z , calculated as:

$$Z = \sum_{i=1} L \cdot U \cdot \omega \cdot X_i \quad (8)$$

In equation (8), ω is the vector of judgment indicators of this judgment matrix. X_i is the affiliation of the risk assessment result in the rating set. Where ω is a positive indicator, in such a case that the larger the value of the calculated result of Z , the more correct that assessment result is. In the design of this paper, a risk assessment is determined to be correct when the value of Z is greater than 25. Finally, the integrated evaluation vector is calculated. According to the judgment result of matrix Z , the affiliation function is given. Let the integrated evaluation vector be B , and its calculation formula is:

$$B = R * Z \quad (9)$$

Through the calculation of the above equation, the integrated assessment vector of the fuzzy hierarchy method is derived, completing the process of quantifying the model indicators of the data asset risk assessment model.

5. Assessing data asset management risk levels based on association rules

Once the data asset management risk assessment model has been constructed, association rules are used to correlate the assessment indicators in the model with the actual risk levels to form textual information that can be read directly by managers. The correlation structure is shown in the Figure 3:

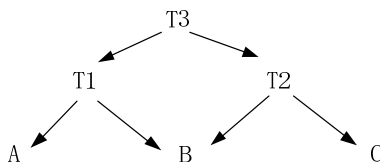


Figure 3. Risk assessment correlation structure diagram

As shown in the figure above, the assessment attributes at different levels are mapped to risk levels. The set of risk assessment items for which a high level of support was obtained. Correlations are made using the objective evaluation rule laplace, which is calculated as:

$$laplace(1 \Rightarrow 5) = \frac{\sup(1 \Rightarrow 5) + 1}{\sup(1) + b} \quad (10)$$

In equation (10), b is the class of association rules. b takes on a positive integer value. In the data asset management risk assessment method designed in this

paper, the value of b is set to 2. The two association rule categories are the leaf-level association rule and the abstract-level association rule.

6. Experiment

6.1. Experiment preparation

In order to verify the effectiveness of the risk assessment method for data asset management proposed in this paper, comparative experiments are designed. The management of data assets of an enterprise is taken as an example. The data assets of this enterprise are classified into seven risk factor categories with different management steps according to the method of this paper. Ten expert personnel in data asset management were invited to assess the risk levels of different risk factors, and the risk levels corresponding to different risk factors in the management of data assets of this enterprise were obtained as shown in the following Table 3:

Table 3. Risk rating scale

Risk factors	Risk level
Accuracy of data asset grooming	Minor Risk
Data Asset Storage System	Severe Risk
Stability of data asset valuation	Severe Risk
Customer Creditworthiness	Moderate Risk
Completeness of client	Moderate Risk
National policy stability	No Risk
Competitive market environment	Significant Risk

In order to enhance the visualisation of the results of this experiment, the 10 experts mentioned above were invited to create a table of assessment scores for the risk levels and present the results of the experiment in the form of evaluation scores. The assessment scores for the different risk levels are shown in the Table 4:

Table 4. Risk assessment score table

Risk level	Score
No Risk	0-5 points
Minor Risk	6-15 points
Moderate Risk	16-40 points
Severe Risk	41-75 points
Significant Risk	76-100 points

The PDCA model focuses on a cyclical approach to risk assessment, in which the risk assessment process for data asset management is completed in a cycle. To ensure the scientific validity of this experiment, the same level of indicator coefficients are set for all three methods.

6.2. Analysis of results

The data asset management risk assessment method based on the fuzzy hierarchy method and association rules proposed in this paper is named Method 1. the data asset management assessment method based on the improved grey fuzzy comprehensive evaluation model is named Method 2. the data asset management risk assessment method based on the PDCA model is named Method 3. according to the above scoring rules, the experimental results were obtained as shown in the Table 5:

Table 5. Assessment results of different methods

Risk factors	Real results	Method 1	Method 2	Method 3
Accuracy of data asset grooming methods	12 points	11 points	3 points	6 points
Data Asset Storage System Security	64 points	62 points	79 points	45 points
Stability of data asset valuation guidelines	71 points	70 points	46 points	37 points
Customer Creditworthiness	36 points	35 points	34 points	19 points
Completeness of client information	27 points	29 points	16 points	13 points
National policy stability	2 points	4 points	8 points	7 points
Competitive market environment	93 points	89 points	76 points	68 points

It can be seen that the data asset management risk assessment method based on the fuzzy hierarchy method and association rules proposed in this paper has a high accuracy rate for different risk factors, and the score error of the assessment results is small, which has a high application value.

7. Conclusion

This paper addresses the problem of large errors in the assessment results in the risk assessment of data asset management, and proposes a risk assessment method for data asset management based on the fuzzy hierarchy method and association rules. The method quantifies the indicators of the assessment model through the fuzzy hierarchy method, which effectively reduces the error value of the risk assessment results of data asset management. It has high application value in the realistic data asset management risk assessment work.

Cancer of the Bladder Risk Assessment (COBRA) score accurately predicts cancer-specific survival after radical cystectomy: external validation and lymphovascular invasion assessment value to improve its performance. Clinical genitourinary cancer, 2021, 20(3).

- Novoa Jurado Abel Jaime, Abellán José María. Cognitive, ethical, social and political aspects in risk assessment: regarding vaccines against SARS-CoV-2. Gaceta sanitaria, 2021, 36(1).

References

- Everett L. CEC04-03 Integration and analysis of high-throughput assays in Next Generation Risk Assessment (NGRA). Toxicology Letters, 2022, 368(S).
- GutiérrezValencia Marta, Leache Leire, Saiz Luis Carlos. Review of the validity of fall risk assessment scales in hospitalised patients. Revista española de geriatría y gerontología, 2022, 57(3).
- Riba Porquet Francesc, GuarnerArgente Carles, Solanes Cabus Mònica, Francia Santamaria Esther, GarciaIglesias Pilar, Machlab Machlab Salvador, Mascort Roca Juanjo, Mendive Arbeola Juan Manuel, Ortiz Seuma Jordi, Calvet Calvo Xavier. Indications and risk assessment of endoscopic examinations in elderly or frail people. Position paper of the Societat Catalana de Digestologia, the Societat Catalana de Geriatria i Gerontologia and the Societat Catalana de Medicina de Família i Comunitària. Gastroenterologia y hepatologia, 2022, 45(10).
- Colomer Gallardo Anna, Candela Luigi, Buisan Rueda Oscar, Freixa Sala Roger, Elias Cañavera Javier, Moschini Marco, Macek Petr, Bennamoun Mostefa, Mombet Annick, Cathelineau Xavier, Areal Calama Juan José, SánchezSalas Rafael. The