A Method for Assessing the Security of Energy Trading Data Based on Big Data and the Pagerank Algorithm

Weibin Ding[1*], Deqi Zhang[1], Junjia Yang[1], Yanzuo Chen[2]

[*] Corresponding author: 277322541@qq.com

State Grid Zhejiang Electric Power CO., LTD., Hangzhou 310007, China[1]
State Grid Zhejiang Economic Research Institute, Hangzhou 310012, China[2]

ABSTRACT: The current assessment of the security of energy transaction data has vague evaluation dimensions, resulting in large errors in the assessment results. To this end, we propose a method for assessing the security of energy transaction data based on big data and the Pagerank algorithm. Determine assessment requirements based on big data and obtain assessment requirements based on six dimensions. Construct a security assessment model and calculate the indicator weights for this model. The assessment results are obtained dynamically based on the Pagerank algorithm. Experiments show that the evaluation results of the method have a small error, with an average error of only 0.22%, which has a high application value.

Keywords: energy transaction data; security assessment; big data; Pagerank algorithm; assessment methods

1. INTRODUCTION

Energy trading is a very important part of trade transactions. The main areas involved in energy trading are natural gas, crude oil, petrochemicals, etc. With the development of internet technology, energy trading is no longer limited to the offline level and has given rise to many online energy trading platforms [1]. In this process, a large amount of data and information on energy transactions is generated. As one of the largest transactions in the trade, the data generated during the energy trading process is very important. If this data is lost or leaked, it will cause incalculable losses for both parties. Therefore, the security of energy transaction data is important. Assessing the security of energy transaction data can effectively improve the security of energy transaction data. Big data technology enables access to large amounts of data information. Through big data methods, the security indicators of past energy transaction data can be obtained and used as indicators for security assessment [2]. The Pagerank algorithm is a process of making hyperlinks between web pages to complete the calculation. The method is based on the hyperlinks in the web pages as a step in the algorithm's analysis to determine the importance of the data information. If the importance of that data is high, the more references are made to that data. Based on the results of the Pagerank algorithm, the number of times the energy trading data has been cited is known. Based on this data, a model for assessing the security of energy trading data is constructed. The indicators and weights of the model are calculated to complete the security assessment process [3]. The method of assessing the security of energy trading data based on big data and the Pagerank algorithm can

ICBBEM 2023, May 19-21, Hangzhou, People's Republic of China

DOI 10.4108/eai.19-5-2023.2334237

significantly improve the accuracy of security assessment and provide strong support for realistic energy trading data management work.

2. DEFINE SECURITY ASSESSMENT REQUIREMENTS BASED ON BIG DATA

IIn assessing the security of energy trading data, it is first necessary to identify the individual aspects of data security that need to be assessed. In this paper's approach to assessing the security of energy trading data, the assessment requirements are set out in six different dimensions, as shown in the table 1:

Table 1 Energy trading data security assessment requirements

Assessment requirements	Content
Data confidentiality	The right to open energy trading data to some people
Authentication of both parties to a transaction	Equipment and entity authentication methods
Data integrity	Right to modify, delete, recreate and copy energy trading data
Non-repudiation	The non-dependability of the behaviour of both parties to an energy transaction
Access control	Authorisation of security scales for energy trading data
Data availability	Network Data Reconstruction Service

As shown in Table 5, based on these six dimensions, big data technology is used to obtain relevant information. Fast access to relevant information in the storage terminal for energy trading data[4]. The terminal vector space v for energy transaction data is calculated as equation (1):

$$v(d_i) = (w_1 \times d_1, w_2 \times d_2, \cdots, w_i \times d_i) \tag{1}$$

In equation (1), i denotes the ith data information of the energy transaction data in the terminal. wi is the feature value of the ith data information. di is the weight of the feature term of the ith energy transaction data information. In the above-mentioned energy transaction data vector representation space, big data techniques are used to obtain information on the security assessment dimensions in this space. Feature items are set for six security assessment dimensions [5]. The energy transaction data information with a high degree of association is extracted in this terminal vector space through the dimension and the feature term association degree of the data information. Let the feature term dimension of the security assessment dimension be l. Then the big data acquisition associated energy transaction data set F is calculated as equation (2):

$$F = \log \frac{N}{d \times l_i} \tag{2}$$

In equation (2), N is all the energy transaction data in the terminal vector space. li is the energy transaction data containing the l-dimensional feature term in the terminal.

3. BUILDING A MODEL FOR ASSESSING THE SECURITY OF ENERGY TRADING DATA

Retaining the amount of information in the security assessment requirements dimension, on the basis of which it corresponds to a valid assessment factor [6-8]. First, in the information flow of any security assessment requirement in the energy transaction data space, let a be an individual numerator of a node in that information flow, then the transformation relationship c between an energy transaction data information node and its corresponding assessment numerator is calculated as equation (3):

$$c(a) = e \times \cos a ´ V_a ´ O + e_a ´ E(a) \tag{3}$$

In equation (3), e is the weighting factor of the security assessment node. Va is the components of the energy trading data security assessment model. o is the feature identification factor for the security assessment requirements in the assessment model. ea is the transformation relationship between the assessment factor and the feature identification relationship. E(a) is the node structure relationship for security assessment. Based on the above calculations, an energy transaction data security assessment model is constructed. The assessment process of the model is shown in the Figure 1:

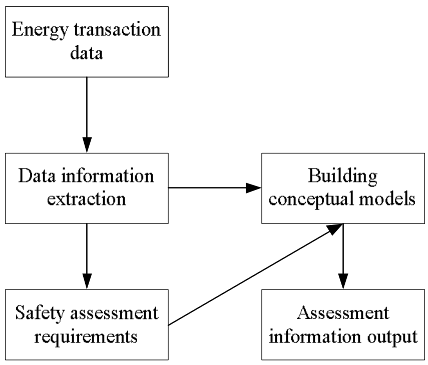

Figure 1 Security Assessment Process

4. CALCULATING THE SECURITY ASSESSMENT MODEL INDICATOR WEIGHTS

The AHP hierarchical analysis is divided into three calculation parts, namely the target, intermediate and factor layers [9-11]. The target layer corresponds to the security outcome of the energy trading data. The factor layer is the factors that influence the security of the energy trading data. The intermediate layer is the process of sorting the data information [12]. First, a judgment matrix was constructed based on AHP hierarchical analysis, with each energy transaction data listed as a different scale, as shown in the table 2:

Table 2 Energy trading data scales

Scale	Implications
1	Of equal importance
3	One data is slightly more secure than the other
5	One data is significantly safer than the other
7	One data is strongly more secure than the other
9	One data is more extremely secure than the other
2,4,6,8	Median of the above two adjacent judgements
Countdown	Comparative data judgement

As shown in Table 2, determine the security relationship between the two energy transaction data. Let the two energy transaction data be denoted as i and j, respectively, then the security relationship between them xij is calculated as equation (4):

$$x_{ij} = \frac{1}{a_{ij}} \quad (4)$$

In equation (4), aij is the individual numerator of the two energy transaction data. Based on the results of equation (4), the consistency indicator CI of this security assessment model is defined as equation (5)

$$CI = \frac{\lambda - n}{n - 1} \quad (5)$$

In equation (5), λ is the feature vector of the model. n is the security weight of each energy transaction data after normalisation. According to the calculation result of equation (5), the indicator weights of the evaluation model meet the criteria when the CI calculation result tends to be close to 0. The final weight matrix Z of the security assessment model is obtained as equation (6)

$$Z = \begin{bmatrix} CI_{11} & CI_{12} & CI_{1n} \\ CI_{21} & CI_{22} & CI_{2n} \\ & & \\ CI_{k1} & CI_{k2} & CI_{kn} \end{bmatrix} \quad (6)$$

In equation (6), k is the number of weights for energy transaction data security after normalisation. After completing the calculation of the index weights of the energy transaction data security assessment model, the model was trained in order to further improve the accuracy of the assessment results.

5. OBTAINING SECURITY ASSESSMENT RESULTS BASED ON THE PAGERANK ALGORITHM

Based on this core function, the Pagerank algorithm is used to obtain the output of the energy trading data security assessment model. the process of obtaining the assessment results for the Pagerank algorithm is shown in the Figure 2 :

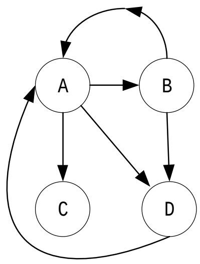

Figure 2 Schematic diagram of the process for obtaining evaluation results

As shown in Figure 2, the results of the query are aggregated to form an assessment of the security of energy trading data under the six assessment dimensions. In this paper, the assessment results are expressed in terms of ratios, and the proportion of each assessment dimension is shown in the table 3:

Table 3 Scale of assessment dimensions

Assessment dimensions	Proportion
Data confidentiality	20%
Authentication of both parties to a transaction	10%
Data integrity	30%
Non-repudiation	10%
Access control	20%
Data availability	10%

As shown in Table 3, the security assessment results output from the assessment model are transformed according to that ratio. Let the security assessment result for energy trading data be R, which is calculated as equation (7):

$$R = R_1 * 20\% + R_2 * 10\% + R_3 * 30\% + R_4 * 10\% + R_5 * 20\% + R_6 * 10\% \tag{7}$$

In equation (7), R1, R2, R3, R4, R5 and R6 represent the evaluation results queried by the Pagerank algorithm from the energy transaction data security evaluation model respectively. Through the calculation of the above equation, the final result of the security assessment of energy trading data in the form of percentage is obtained, and the security assessment of energy trading data is completed.

6. EXPERIMENT

6.1 Experiment preparation

In order to verify the feasibility of the energy trading data security assessment method based on big data and Pagerank algorithm proposed in this paper, comparative experiments are designed. The data in an energy trading platform was used as the experimental object for the comparison experiment. The energy trading platform contains a total of 16,365,577 energy transaction data. In this paper, 16,065,577 of these data messages were used to train the security evaluation model. The remaining 300,000 energy transactions were used for comparison experiments. The final result is the combined security evaluation of these 300,000 energy transactions.

The experimental environment chosen for this experiment is shown in the table 4:

Table 4 Experimental environment parameters

Projects	Parameters
Computers	Lenovo superx5b
CPU	Inter(R) Core(TM)i7-10510U(1.80GHz)
Memory	32GB
Operating systems	Windows10
Main Frequency	1.6Hz
Hard Disk	32TB
Running memory	4GB
Development languages	Python
Compiler environment	jupyter notebook
Algorithm implementation	VC++6.0

6.2 Analysis of results

The results of the different energy trading data assessment methods derived from the above methodology are shown in the table 5:

Table 5 Results of the safety assessment of different methods

Number of experiments	Real results	Method 1 results	Method 2 results	Method 3 results
First experiment	90.56%	89.64%	83.60%	76.32%
Second experiment	90.23%	90.58%	98.77%	78.54%
Third experiment	91.17%	91.06%	77.58%	79.58%
Average results	90.65%	90.43%	86.65%	78.15%

As shown in Table 5, The average results of the other two methods differed from the real results by 4.00% and 12.50% respectively. This shows that the method of assessing the security of energy transaction data based on big data and Pagerank algorithm proposed in this paper has high practical application value.

7. CONCLUSION

This paper addresses the problem of large errors in aligning the assessment results with the real security in the security assessment of energy transaction data, and proposes a method for assessing the security of energy transaction data based on big data and the Pagerank algorithm. By refining the dimensions of security assessment, the method effectively avoids the interference brought by the energy transaction data itself. The proposed method significantly reduces the error of the assessment results and has high application value in the actual energy transaction data security assessment work.

REFERENCES

[1] DEVARAJ, JAYANTHI, ELAVARASAN, RAJVIKRAM MADURAI, SHAFIULLAH, G. M., "A holistic review on energy forecasting using big data and deep learning models," International journal of energy research, 45(9), 13489-13530 (2021).

[2] CHAN, D., LEE, K., LEE, PAN., "Leveraging Big Data To Save Energy," ASHRAE Journal, 63(10), 66-69 (2021).

[3] HARDING, MATTHEW C., LAMARCHE, C., "Small Steps with Big Data: Using Machine Learning in Energy and Environmental Economics," 13469-488 (2021).

[4] ZENG, X., CHEN, M., "A Novel Big Data Collection System for Ship Energy Efficiency Monitoring and Analysis Based on BeiDou System," Journal of advanced transportation, 2021(Pt.7), 9914720.1-9914720.10 (2021).

[5] TIAN, P., ZHANG, L., "Big data mining based coordinated control discrete algorithm of independent micro grid with PV and energy," Microprocessors and microsystems, 82(Apr.), 103808.1-103808.8 (2021).

[6] ZAOUI, S., OKBA, K., BRAHIM, L., "An intelligent system for energy management in smart cities based on big data and ontology," Smart and sustainable built environment, 10(2), 169-192 (2021).

[7] KUMARI, SANJU, KUMAR, NEERAJ, RANA, PRASHANT SINGH. "BIG DATA ANALYTICS FOR ENERGY CONSUMPTION PREDICTION IN SMART GRID USING

GENETIC ALGORITHM AND LONG SHORT TERM MEMORY," Computing and informatics, 40(1), 29-56 (2021).
[8] SUN, P., GU, L., "Energy big data acquisition and application based on service portfolio quality," 45(Jun. Pt.2), 101134.1-101134.8. (2021).
[9] SARAH ABDULLAH AL MUHAYSH, THANAA SAAD AL SALEM, FATIMAH ABDULLAH AL JAAFARI, et al. "A Big Data Application for Energy Consumption Management of Commercial Buildings," 2(1), 1-20 (2021).
[10] WANG, Q., VELASCO, LEONARDO, BREITUNG, BEN, et al. "High-Entropy Energy Materials in the Age of Big Data: A Critical Guide to Next-Generation Synthesis and Applications," Advanced energy materials, 11(47), 2102355.1-2102355.18. (2021).
[11] LI, P., ZHANG, Y., ZHANG, Y., "Prediction of electric bus energy consumption with stochastic speed profile generation modelling and data driven method based on real-world big data," 298(Sep.15), 117204.1-117204.14 (2021).
[12] XU, S., ALTURKI, RYAN, REHMAN, ATEEQ UR, "BP Neural Network Combination Prediction for Big Data Enterprise Energy Management System," Mobile networks & applications, 26(1), 184-190 (2021).