Membrane Technology
ISSN (online): 1873-4049

Identification of Abnormal Bidding Behavior Types of Power Suppliers in Electricity Market Based on Multiple Isolated Forests

Yu Jiang [1,2*]

[1]*State Grid Jiangsu Electric Power Company Ltd., Nanjing 210000, China*
[2]*Jiangsu Power Exchange Center Co., Ltd., Nanjing 210008, China.*
**Corresponding Author.*

Abstract:
With the steady development of the electricity market, China's electricity market is increasingly showing a large-scale, multi-agent and other characteristics. There are many types of market violation risks and it is difficult to prevent them. Aiming at all kinds of violations in the electricity market, in order to ensure fair competition in the market and prevent and resolve market risks caused by abnormal bidding behaviors in the market, this paper proposes a method for identifying the types of abnormal bidding behaviors of power producers in the electricity market based on multiple isolated forests. Based on the identification index of multi-dimensional abnormal bidding behavior type proposed in this paper, this method first encodes the characteristics of different types of abnormal behavior in the market bidding process, and establishes the numerical characteristic table of abnormal behavior. Secondly, the multiple isolated forest algorithm proposed and constructed in this paper is used to identify the types of potential abnormal behaviors in the bidding process of e-commerce. The example part takes the actual bidding data of a power producer in the US PJM spot market as the research object for anomaly identification and analysis verification. The example results show that the method can effectively identify the abnormal bidding behavior types of the power producer in the market, verify the effectiveness of the method, provide decision support for market supervision, and help the power market run healthily and smoothly.

Keywords: electricity market, electricity trading center, power producer, abnormal bidding behavior, isolated forest, outlier detection.

INTRODUCTION

With the development of China's electricity market entering a new stage, especially the trial operation and construction requirements of the spot market, the research on the identification of potential abnormal behaviors and violations in the bidding process of market members is becoming more and more important, which has become a very concerned issue for regulatory authorities, trading centers, market entities and experts and scholars. At the same time, the state's regulatory policy on the legal and compliant operation of the electricity market also requires the supervision of market violations and the strengthening of market operation risk management[1].

Abnormal behavior refers to the behavior that market entities violate trading rules or may bring risks to market operation in the process of participating in the market. The abnormal behavior of market members may lead to the fluctuation, abnormal high and abnormal low of market transaction electricity price, and even bring systemic risks to market operation, such as large-scale losses of market members or the inability of a large number of market members to close transactions. In view of the irregularities in the electricity market, Chinese and foreign scholars have put forward some identification methods and preventive measures. The research on abnormal bidding behavior of market members can be divided into two research fields: abuse of market power abnormal behavior and multi-type abnormal behavior. The research results in the field of abuse of market power by power producers are rich, and the field of multi-type abnormal behavior is developing rapidly.

The identification of abuse of market power in the bidding process of power producers has become the core and focus of current abnormal research. Sun Bo et al. proposed a method for identifying the abuse of market power by power producers based on Lasso-logit model[2]. The model has an accuracy rate of 96.6 % in the application of power spot market in a certain area, which can effectively identify the abuse of market power. Based on AdaBoost-DT algorithm, Sun Qian et al. designed an identification model of power producers abusing market power[3]. The calculation example was based on the actual data of a regional spot market, and the result showed that the model identification accuracy was 97%, which verified the effectiveness of the method. He Ye et al in the power producer is modeled as an agent that selects a high-profit bidding strategy based on historical bidding information, and a monopoly detection model of the electricity market based on the agent bidding simulation framework is proposed[4]. Nan Shang et al classifies and reviews the multi-market entities on the power generation side and the power purchase side, and summarizes the application of regulatory measures in the United States, Northern Europe, the United Kingdom and China[5]. Xie Jingdong et al. designed a semi-supervised support vector machine collusion recognition model for the scenario where collusion samples are scarce[6]. The results of the example prove that the method has high accuracy. Hua Huichun constructed a variational auto-encoder Gaussian mixture model, which was applied to the intelligent early warning of collusion between power suppliers[7]. Sun Bo et al. tackles the problem of recognizing collusive behavior in the bidding process of cartel-type generating units[8]. A methodology based on a ranked multinomial Logit model is introduced. The results of case studies suggest that, when the significance level is below 5%, there is a potential presence of collusive bidding behavior among cartel-type generating units. Current research in identifying market power abuse by power generators is well-established, featuring the proposal of various recognition methods.

The existing references also have research on the identification of multiple types of market entities and multiple types of abnormal bidding behaviors. Xie Jingdong et al proposed an abnormal behavior identification method of market members based on RKLOF algorithm, which improved the accuracy of abnormal identification[9]. A method for identifying potential hazard behaviors in the electricity market based on cloud model and fuzzy Petri net is designed[10]. The results of the example are highly consistent with the results of expert analysis and actual data, which verifies the effectiveness of the method. Yang Mo et al analyzes and models the characteristics of violations in the whole process of power generation, electricity sales, and power user transactions[11]. At the same time, the fuzzy comprehensive evaluation method is used to comprehensively evaluate multiple types of violations. The existing research on the identification of multiple types of market players and multiple types of abnormal bidding behavior is relatively lacking, and it is also one of the hot spots in the research of abnormal behavior of market members.

The contributions of this manuscript are:

(i) Proposed is an enhanced unsupervised machine learning algorithm based on the Isolation Forest algorithm. This algorithm can directly output label values for anomalous behaviors, facilitating the identification of abnormal behaviors in power generation businesses.

(ii) In response to various types of abnormal behaviors exhibited by power generation businesses, a method is proposed. This method involves constructing multiple isolation forests specifically for the detection of bidding data in power generation businesses. The direct output from this approach specifies the particular type of abnormal behavior present.

(iii) The approach presented in this paper has the capability to directly output label values for different types of abnormal behaviors. In contrast, existing methods typically involve initially identifying anomalous points and subsequently combining metric analysis results to manually determine the type of abnormal behavior.

The remaining sections of the paper are organized as follows: Section two discusses the abnormal behaviors present in the electricity market, followed by the introduction of features encoding abnormal bidding behaviors of power generation businesses in section three. Section four delves into the identification process of power generation businesses' abnormal behaviors using multiple isolation forests. Section five presents an analysis of the experimental results, and in the concluding sixth section, the paper is summarized.

TYPES AND MANIFESTATIONS OF ABNORMAL BEHAVIOR OF POWER SUPPLIERS

Extreme Quotation

Power suppliers adopt “extreme quotation” behavior, which is manifested in two forms: extreme high price and extreme low price. Extreme high price behavior means that the power suppliers declare at a high price higher than their own historical quotation level when making market declaration so that they have the opportunity to obtain excess profits. However, this quotation method may raise the market clearing price and disturb the market order. The extreme low-price behavior means that the power suppliers declare at a low price lower than their own historical quotation level when making market declaration so as to achieve the purpose of suppressing competitors, which may make other power producers lose money.

Extreme Declaration of Electricity Quantity

Power suppliers adopt “Extreme Declaration of Electricity Quantity” behavior, which is manifested in two forms: extreme declaring high and low electricity quantity. The extreme low one is mainly manifested in the fact that the power suppliers intentionally declare few electricity quantity, thereby reducing the effective supply of the market, increasing the market price, interfering with the normal operation of the market, and even causing the purchase side to be unable to purchase enough electricity to meet the electricity demand of the electricity side. The extreme high one is mainly manifested in the intentional declaration of more electricity quantity by power suppliers, which may even exceed their own power generation capacity, interfere with the normal clearing of the market, and affect the trading electricity quantity of other market entities.

Complex Anomaly

There is a "complex anomaly" in the bidding behavior of power suppliers, which refers to the abnormal quotation and quantity of power suppliers. For example, power suppliers with large power generation capacity deliberately declare high prices and low quantities at the same time, which may lead to an imbalance between supply and demand in the market and eventually lead to an increase in transaction price. Or some power suppliers deliberately adopt the bidding method of declaring low prices and high quantities, which may lead to the loss of fairness in the market, and even the failure of power generation entities to complete transactions or large-scale losses.

The common abnormal behavior of power producers should also include market collusion. The existing research results have been very mature in identifying market collusion between power suppliers. This paper focuses on the identification of bidding behavior of a single power supplier, that is, whether the bidding behavior to be identified is abnormal compared with the historical bidding behavior, and whether there are extreme quotations, extreme electricity quantity and complex abnormal behaviors.

CHARACTERISTIC CODING OF ABNORMAL BIDDING BEHAVIOR OF SUPPLIERS BASED ON ABNORMAL IDENTIFICATION INDEX

Abnormal Behavior Identification Index

The actual market bidding process is declared in the form of multi-stage quotation. The quotation format of the market subject is $\left[\left(\rho^{bid,1},Q^{bid,1}\right),\cdots,\left(\rho^{bid,N},Q^{bid,N}\right)\right]$, where N the total number of quotation segments is, and $\left(\rho^{bid,i},Q^{bid,i}\right)$ is the quotation and quantity in stage i. The abnormal behavior identification indicators defined in this paper are six in total: weighted average declared electricity price, ratio of the highest price in the segment to the average declared electricity price, ratio of the lowest price in the segment to the average declared electricity price, total declared electricity quantity, ratio of the highest price in the segment to the total declared electricity quantity and ratio of the lowest price in the segment to the total declared electricity quantity.

Weighted average declared electricity price

$$index1(i)=\frac{\sum_{j=1}^{N}\left[\rho^{bid,j}(i)\cdot Q^{bid,j}(i)\right]}{\sum_{j=1}^{N}Q^{bid,j}(i)} \tag{1}$$

In the equation, $index1(i)$ represents the weighted average declared price of the main body of the power suppliers in the i-th market transaction, $\rho^{bid,j}(i)$ represents the j-section quotation of the main body of the power suppliers in the i-th market transaction, and $Q^{bid,j}(i)$ represents the j-section quotation of the main body of the power suppliers in the i-th market transaction. The weighted average declared price is weighted according to the amount of each section, which reflects the overall bidding level of the power supplier.

The ratio of the highest quotation to the average declared price in the section

$$index2(i)=\frac{\max\limits_{1\le k\le N}\left\{\rho^{bid,k}(i)\right\}}{\dfrac{\sum_{j=1}^{N}\left[\rho^{bid,j}(i)\cdot Q^{bid,j}(i)\right]}{\sum_{j=1}^{N}Q^{bid,j}(i)}}=\frac{\max\limits_{1\le k\le N}\left\{\rho^{bid,k}(i)\right\}\cdot\sum_{j=1}^{N}Q^{bid,j}(i)}{\sum_{j=1}^{N}\left[\rho^{bid,j}(i)\cdot Q^{bid,j}(i)\right]} \tag{2}$$

In the equation, $index2(i)$ represents the ratio of the highest quotation in the segment to the average declared price of the power suppliers in the i-th market transaction, which can be used to identify the extreme high-price behavior of the power suppliers.

The ratio of the lowest quotation to the average declared price in the section

$$index3(i)=\frac{\min\limits_{1\le k\le N}\left\{\rho^{bid,k}(i)\right\}}{\frac{\sum\limits_{j=1}^{N}\left[\rho^{bid,j}(i)\cdot Q^{bid,j}(i)\right]}{\sum\limits_{j=1}^{N}Q^{bid,j}(i)}}=\frac{\min\limits_{1\le k\le N}\left\{\rho^{bid,k}(i)\right\}\cdot\sum\limits_{j=1}^{N}Q^{bid,j}(i)}{\sum\limits_{j=1}^{N}\left[\rho^{bid,j}(i)\cdot Q^{bid,j}(i)\right]} \tag{3}$$

In the equation, $index3(i)$ represents the ratio of the minimum quotation to the average declared price of the power suppliers in the i-th market transaction, which can be used to identify the extreme low-price behavior of the power suppliers.

The total amount of declared electricity

$$index4(i)=\sum_{j=1}^{N}Q^{bid,j}(i) \tag{4}$$

In the equation, $index4(i)$ represents the total amount of electricity declared by the power suppliers in the i-th market transaction, which reflects the overall reporting level of the power suppliers.

The ratio of the highest declared amount to the total amount of declared electricity in the section

$$index5(i)=\frac{\max\limits_{1\le k\le N}\left\{Q^{bid,k}(i)\right\}}{\sum\limits_{j=1}^{N}Q^{bid,j}(i)} \tag{5}$$

In the equation, $index5(i)$ represents the ratio of the maximum declared amount to the total amount of declared electricity in the i-th market transaction, which can be used to identify the extreme high-declaring behavior of the power suppliers.

The ratio of the minimum declared amount to the total amount of declared electricity in the section

$$index6(i)=\frac{\min\limits_{1\le k\le N}\left\{Q^{bid,k}(i)\right\}}{\sum\limits_{j=1}^{N}Q^{bid,j}(i)} \tag{6}$$

In the equation, $index6(i)$ represents the ratio of the minimum declared amount to the total amount of electricity declared in the i-th market transaction, which can be used to identify the extreme low-reporting behavior of the power suppliers.

Characteristic Coding of Abnormal Bidding Behavior of Power Suppliers

The numerical feature set of abnormal behavior based on abnormal identification index is shown in Table 1.

Table 1. Metrics and comparisons.

Type of abnormal behavior	Forms and Characteristics
Extreme quotation	The weighted average declared electricity price is abnormal or the weighted average declared electricity price is normal but the quotation in the section is abnormal.
Extreme declaration of electricity quantity	The total amount of declared electricity is abnormal or the total amount of declared electricity is normal, but there is an abnormal amount in the section.
Complex anomaly	The quantity and quotation are abnormal

According to the quotation and quantity characteristics of different types of abnormal bidding behaviors of power suppliers in table 1, combined with the identification index of abnormal behavior, quantitative coding is carried out for different types of abnormal behaviors. Define the special data format for abnormal behavior identification, that is, format $\left[\zeta^1(i),\zeta^2(i),\zeta^3(i),\zeta^4(i),\zeta^5(i),\zeta^6(i)\right]$, and make the abnormal behavior numerical feature set $(Table\_ABFS)$. The storage form is a table, as shown in Table 2:

Table 2. Abnormal behavior data collection.

Abnormal behavior number	Type of abnormal behavior	Abnormal behavior recognition array
A	Extreme quotation	[-1, -1, 1, 1, 1,1]or [-1, 1, -1, 1, 1, 1or [-1, -1, -1, 1, 1, 1]
B	Extreme declaration of electricity quantity	[1, 1, 1, -1, -1, 1]or [1, 1, 1, -1, 1, -1]or [1, 1, 1, -1, -1, -1]
C	Complex anomaly	At least two of ζ^1,ζ^2,ζ^3 are-1 and At least two of ζ^4,ζ^5,ζ^6 are -1.

Note: In the abnormal behavior recognition array of the above table, "1" indicates that there is no anomaly, and "-1" indicates that there is an anomaly.

IDENTIFICATION PROCESS OF ABNORMAL BEHAVIOR OF POWER SUPPLIERS BASED ON MULTIPLE ISOLATED FORESTS

The isolated forest algorithm is a non-parametric algorithm, which does not need to assume the distribution of the data set in advance, nor does it need to train a parametric model[12]. Because of its excellent outlier detection performance, the isolated forest algorithm has been proved to have achieved very good performance in many application fields in engineering practice [13]. In the problem of abnormal behavior identification of power suppliers, the isolated forest algorithm can only detect whether there is an abnormality in the bidding behavior of power suppliers, that is, the model can only output Boolean results with or without abnormalities, and cannot identify which abnormalities are specific. In this paper, the isolated forest algorithm is extended to obtain the multiple isolated forest algorithm, and the identification model of abnormal bidding types of power suppliers based on multiple isolated forests is proposed to realize the identification of different types of abnormal behaviors.

The abnormal bidding type identification model of multiple isolated forests is mainly divided into two parts in the process of anomaly identification: First, it enters the abnormal behavior detection step and adopts the isolated forest discriminant algorithm to detect whether there is any abnormality (corresponding to "Abnormal Behavior Detection in Section 3.1"). If there is no abnormality, the result of "no abnormality" will be output without the specific abnormal behavior identification process. If there is an anomaly, the abnormal behavior identification step will be transferred. Multiple isolated forest identification algorithm was used to identify the specific types of anomalies (corresponding to "Abnormal Behavior Identification in Section 3.2").

After extracting the historical bidding data of power suppliers, the six-dimensional abnormal behavior identification index is calculated for N times of historical bidding data and the historical bidding feature data set of power suppliers is constructed in the format shown in equation (7).

$$X^{train}=\begin{pmatrix} index1(1) & \cdots & index6(1) \\ \vdots & \ddots & \vdots \\ index1(N) & \cdots & index6(N) \end{pmatrix}=\begin{bmatrix} X^{rain\_1} & X^{train\_2} & X^{train\_3} & X^{train\_4} & X^{train\_5} & X^{train\_6} \end{bmatrix} \tag{7}$$

The whole process of abnormal bidding type identification of power suppliers based on multiple isolated forest method is as follows Figure 1:

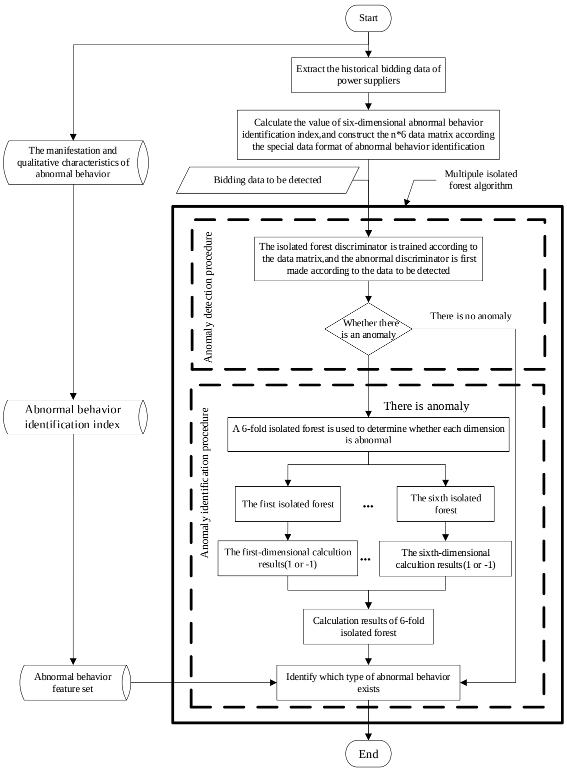

Figure1. Abnormal behavior type identification process of power suppliers based on multiple isolated forest algorithm

Abnormal Behavior Detection Based on Isolated Forest Discriminant Algorithm

Based on the abnormal behavior detection procedure of multiple isolated forest algorithm, the process of using isolated forest discrimination algorithm to detect whether there is abnormal bidding behavior of power suppliers is shown in Figure 2:

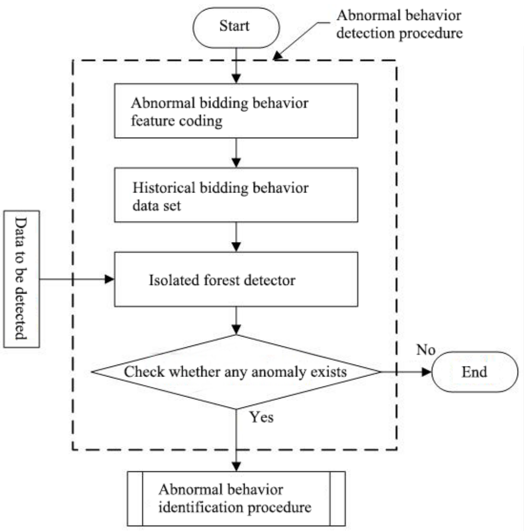

Figure 2. Flow chart of abnormal behavior detection of power suppliers based on isolated forest discriminant algorithm

The isolated forest discriminant algorithm constructs a series of random binary trees for the six-dimensional abnormal behavior identification index feature data $X^{rain\_1}, X^{train\_2}, X^{train\_3}, X^{train\_4}, X^{train\_5}, X^{train\_6}$ in the input training data set X^{train}. Each node of these random binary trees has either two child nodes or leaf nodes. The isolated forest discriminant algorithm divides the data in this range into two branches by randomly taking values in a certain range, and then continues to divide the random values in the two branches. This is repeated until the tree is indivisible or the height of the tree reaches the limit. Since the abnormal points are rare, they will be quickly divided into leaf nodes in the random tree. Therefore, by calculating the path length between the leaf node and the root node, it is possible to quickly determine whether there is abnormal behavior.

The construction process of the isolated forest discriminant algorithm can be roughly divided into two stages. The first stage trains t isolated trees to form an isolated forest. Then, each sample point is substituted into each isolated tree in the forest, the average height is calculated, and the outlier score of each sample point is calculated.

The first step is as follows:

1) Randomly select ψ sample points from the training data X^{train} as the sample subset and put them into the root node of the tree.

2) Randomly specify a dimension (feature) from the six dimensions, and randomly generate a cutting point p in the current node data (the cutting point is generated between the maximum and minimum values of the specified dimension in the current node data).

3) A hyperplane is generated by this cutting point, and then the data space of the current node is divided into two subspaces: the data less than p in the specified dimension is placed on the left child node of the current node, and the data greater than or equal top is placed on the right child node of the current node.

4) In the child node recursively steps 2) and 3), new child nodes are constructed until there is only one data in the child node (can no longer be cut) or the child node has reached a limit height.

5) Loop 1) to 4) until t isolated tree iTrees are generated.

The second step is as follows:

After obtaining t iTrees, iForest training ends, and the generated iForest is used to evaluate the test data X^{test}. For each data point x_i, let it traverse each isolated tree (iTree), calculate the average height $h(x_i)$ of the point in the forest, and normalize the average height of all points. The degree of abnormality is judged by calculating the score of the abnormal value.

Abnormal Behavior Recognition Based on Multiple Isolated Forest Identification Algorithm

On the basis of Section 3.1, based on the abnormal behavior recognition procedure in the multiple isolated forest algorithm, this section identifies the specific abnormal bidding types of power suppliers, and can identify extreme quotations, extreme declaration of electricity quantity and complex anomalies, which further improves the practicability of the model. In the multiple isolated forest algorithm, each isolated forest is defined as an isolated forest identification algorithm.

Mathematical Model of Isolated Forest Identification Algorithm

In the process of anomaly identification, the core parameter of the isolated forest identification algorithm is the anomaly index $s(x,\psi)$, and the expression is shown in equation (8).

$$\begin{cases} s(x,\psi) = 2^{-\frac{E(h(x))}{c(\psi)}} \\ c(\psi) = \begin{cases} 2\cdot\left[\ln(\psi-1)+0.5772156649\right]-\dfrac{2\cdot(\psi-1)}{\psi}, & for\ \ \psi<2 \\ 1 & ,for\ \ \psi=2 \\ 0 & ,otherwise \end{cases} \end{cases} \tag{8}$$

In the equation, n is the size of subsampling, $h(x)$ is the path length of sample x, and $E(h(x))$ is the average value (expected value) of $h(x)$ in the iTrees set.

When $s(x,\psi)$ approaches 1, it indicates that x is an abnormal point (outlier). When $s(x,\psi)$ is far less than 0.5, it indicates that x is a normal point. When all points are near 0.5, there is no obvious anomaly at all points.

The identification process of abnormal bidding types of power suppliers based on multiple isolated forest identification algorithm

When the isolated forest discriminant algorithm detects the existence of anomalies, it enters the abnormal behavior recognition procedure. The specific identification process of abnormal bidding types of power suppliers based on multiple isolated forest identification algorithm is as follows:

S1: The six-fold isolated forest identification algorithm is used to train six-dimensional data $X^{rain\_1}, X^{train\_2}, X^{train\_3}, X^{train\_4}, X^{train\_5}, X^{train\_6}$ of the historical bidding feature data set X^{train} of power suppliers respectively, and the six-fold isolated forest identification algorithm model is obtained, as shown in equation (9).

$$CLF^{j} = iForset^{j}(X^{j}), 1 \leq j \leq 6 \tag{9}$$

In the equation, CLF^{j} represents the j-th isolated forest identification algorithm, and $iForset^{j}$ represents the training function of the j-th isolated forest identification algorithm.

S2: According to the bidding data $X^{test} = \left[\left(\rho^{bid,1}, Q^{bid,1}\right), \cdots, \left(\rho^{bid,N}, Q^{bid,N}\right)\right], 1 < i < N$ to be detected, the six-dimensional abnormal behavior identification index $index^{1}, index^{2}, index^{3}, index^{4}, index^{5}, index^{6}$ is calculated, which is input into the corresponding isolated forest identification algorithm respectively, and the anomaly detection results of each dimension are calculated, as shown in equation (10).

$$y^{j} = CLF^{j}\_predict(index^{j}) \tag{10}$$

In the equation, y^{j} represents the detection result of the j-th isolated forest identification algorithm, and $CLF^{j}\_predict$ represents the prediction function of the j-th isolated forest identification algorithm.

S3: According to the 6-dimensional detection result $\left[y^{1}, y^{2}, y^{3}, y^{4}, y^{5}, y^{6}\right]$ of the 6-fold isolated forest identification algorithm, check Table 2 to identify what type of abnormal behavior exists in the bidding data of current power suppliers.

EXAMPLE ANALYSIS

Aiming at the actual bidding data of power suppliers in the electricity market, the effectiveness of the identification method of abnormal behavior of power suppliers proposed in this paper is verified.

Data Sources

In this paper, the example is analyzed by using the bidding data of the US PJM spot market. PJM began to implement the intraday quotation of power generation resources in November 2017.Therefore, for the quotation data of power suppliers in October 2017 and before, PJM only provides the quotation data with the minimum unit of day, rather than the hourly level. In order to prevent the disclosure of confidential information, the generator identification has been shielded before disclosure, and the shielding code is changed every year. The example of this paper selects the bidding data of PJM electricity market from January 2017 to October 2017 as the data set of this paper. The historical electricity quantity declaration data of power suppliers are shown in Figure 3, and the historical electricity quantity declaration data are shown in Figure 4.

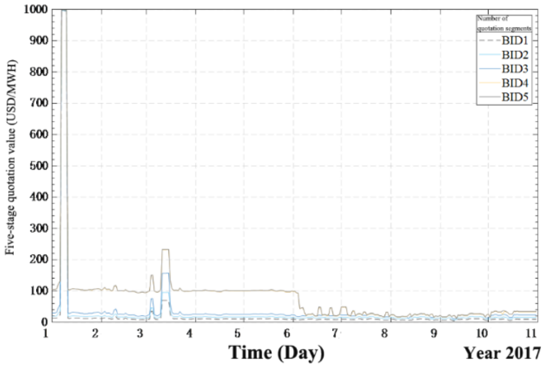

Figure 3. Historical quotation data of power suppliers

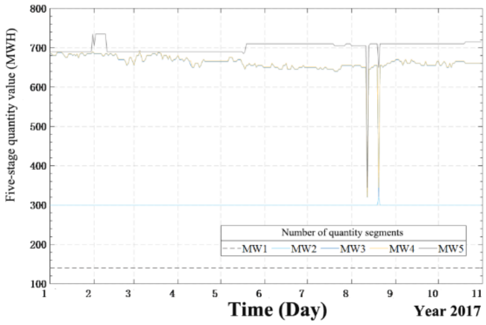

Figure 4. Historical quantity data of power suppliers

Abnormal Behavior Identification Results

Calculation of power suppliers abnormal behavior identification Index

The abnormal behavior identification index was calculated based on the historical bidding data of the power supplier, and the calculation results were shown in Figure 5.

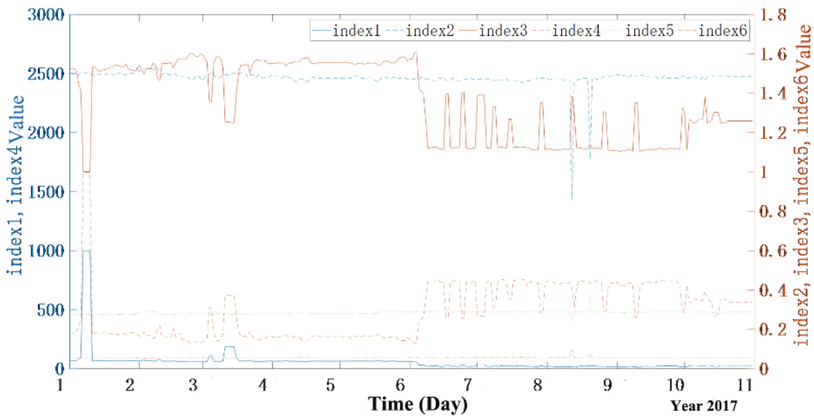

Figure 5. Calculation results of historical abnormal behavior identification index of power suppliers

Taking the quotation data of the power supplier in October 2017 as an example, the calculation results of the abnormal behavior identification index are shown in annexed Table 2.

Detect for abnormal behavior

The train _ test _ split function in the Scikit-learn library is used to divide the training data and the test data. The test _ size is 0.2, that is, the test data accounts for 20 %, and the random _ state is set to 40 (random setting, convenient result reproduction). A total of 243 training data are used to train the abnormal bidding behavior detector of power suppliers based on isolated forest. A total of 61 test data are used as the data to be tested to evaluate the model test results. The distribution of each index in the training set is shown in Figure 6.

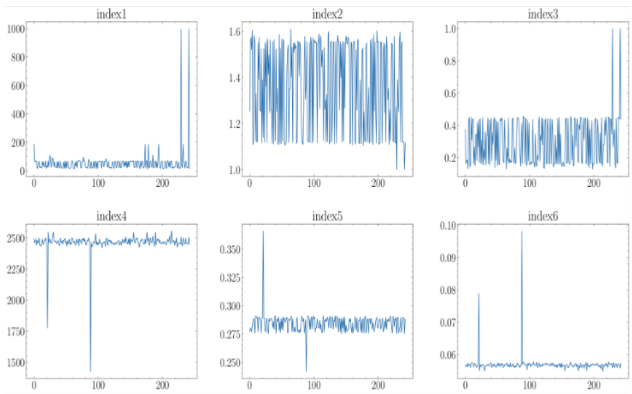

Figure 6. Distribution of six-dimensional abnormal behavior identification indexes in training set

The distribution of each index in the test set is shown in Figure 7.

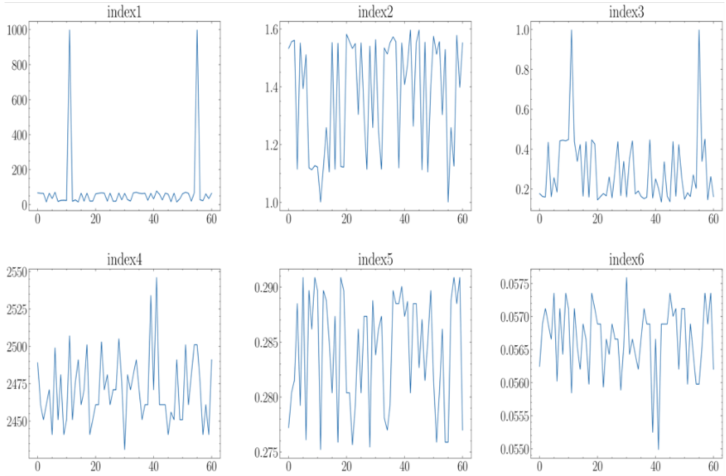

Figure 7. Distribution of six-dimensional abnormal behavior identification index in test set

The test set index is:

[18,112,102,208,92,167,13,187,264,163,227,6,176,292,245,129,260,43,173,194,65,111,23,132,286,125,237,258,77,297,100,289,257,11,3,133,151,139,266,37,156,39,53,275,143,59,215,135,244,175,103,29,113,287,46,9,296,198,146,168,90] (index starting from 0)

In the process of dividing the historical bidding sample data, in order to increase the randomness of the data to improve the generalization ability of the anomaly identification model, the train _ test _ split function is used to randomly divide the training set and the test set, so that the original bidding data sequence arranged incrementally by time is disrupted. Therefore, it is necessary to correspond the index of the test set to the original timestamp one by one to facilitate backtracking and search.

The corresponding time and date are: 1-19,4-23,4-13,7-28,4-3,6-17,1-14,7-7,9-22,6-13,8-16,1-7,6-26,10-20,9-3,5-10,10-18,2-13,6-23,7-14,3-7,4-22,1-24,3-12,10-14,5-6,8-26,9-16,3-19,10-25,4-11,10-17,9-15,1-12,1-4,3-14,6-1,5-20,9-24,2-7,6-6,2-9,2-23,10-3,5-24,3-1,8-4,5-16,9-2,6-14,4-14,1-30,4-24,10-15,2-16,1-10,10-24,7-18,5-27,6-18,4-1 (Example: 1-17 means January 17, 2017, year is omitted).

The parameters of the isolated forest detector are set as shown in Table 3.

Table 3. Parameter settings of isolated forest discriminant algorithm

n_estimators	10000
max_samples	'auto'
random_state	None

The default value of the isolated forest “n _ estimators” parameter is 100. In order to improve the stability and calculation accuracy of the algorithm, the parameter is set to 100000. The experimental results show that the program runs very stably under this parameter setting, and the results of multiple runs do not change. Visualize the results as shown in Figure 8.

Figure 8. Test results of all bidding data to be tested

There are a total of 61 bidding data to be detected. The detection results in Figure 3 show that there are 4 anomalies, and the remaining 57 are not abnormal. According to the sequence number of the detected anomaly is 11, 39, 41, 55 (the sequence number in the program starts from 0), the corresponding index is 6, 37, 39, 9 (test set index). The serial number of abnormal bidding behavior in the test set is traced back, and the date of abnormal quotation data is January 7, 2017, February 7, 2017, February 9, 2017 and January 10, 2017.

Identify Specific Types of Abnormal Behavior

The abnormal behavior detection results show that there are abnormal behaviors in the four biddings in the test bidding data. This section identifies which abnormal behaviors exist in the four biddings.

(1) The multiple isolated forest method is used to train the six isolated forests respectively, and the abnormal behavior types are identified for four abnormal points. The parameter settings of each isolated forest model are the same as Table 4. According to the calculation results of multiple isolated forests, the specific abnormal behavior types are identified by referring to Table 2, as shown in Table 4.

Table 4. Calculation results of abnormal bidding data to be detected based on multiple isolated forests.

Timestamp of abnormal bidding behavior	Six fold isolated forest identification results	Corresponding abnormal behavior type
January 7, 2017	[-1, -1, -1, -1, -1, -1]	Complex anomaly
January 10, 2017	[-1, -1, -1, 1, 1, 1]	Extreme quotation
February 7, 2017	[1, 1, 1, -1, -1, -1]	Extreme electricity quantity
February 9, 2017	[-1, -1, -1, -1, 1, -1]	Complex anomaly

(2) It can be seen from Table 5 that there are four abnormal bidding behaviors in the 61 data of the test set, which are:

(3) On January 7, 2017, there was a complex abnormal behavior, that is, there were abnormalities in both quotation and quantity;

(4) On January 10, 2017, there was an extreme quotation behavior, that is, there was an abnormality in the quotation;

(5) On February 7, 2017, there was an extreme quantity behavior, that is, there was an anomaly in the quantity;

(6) On February 9, 2017, there was a complex abnormal behavior, that is, there were abnormalities in both quotation and quantity.

Verification Analysis of Identification Results

Since the data used in this paper is non-abnormal label data, in order to verify the effectiveness of the proposed algorithm, the identification results of the algorithm are analyzed and verified. Combined with the data of Figure 3, Figure 4, Figure 5 and Figure 7, it can be concluded that:

(7) In terms of quotation, Figure 5 shows that the index1 weighted average declared electricity price index calculated by the bidding data on January 7, 2017 is high, the ratio index of the highest quotation to the average declared electricity price in the index2 section is low, and the ratio index of the lowest quotation to the average declared electricity price in the index3 section is high. The original quotation data in Figure 3 shows that the quotation data of the day is much higher than other data, which is seriously deviated from the historical average. The sub-graph of “index1”, “index2” and “index3” of figure 7 clearly show that the index1 of the bidding data on January 7, 2017 (the 11th point in the graph) is abnormally high, the index2 is obviously low, and the index1 is abnormally high. The above results show that there is an extreme high price quotation on the day. In terms of declaring quantity, Figure 5 shows that the index of the total amount of electricity declared in index4 is high, the ratio index of the highest declaring quantity to the total amount of electricity declared in index5 is low, and the ratio index of the lowest declaring quantity to the total amount of electricity declared in index6 is high. In Figure 7, the index for index4 is significantly higher, the index for index5 is significantly lower, and the index for index6 is significantly higher. Therefore, it is believed that there is an extreme quantity of quotation on the day. In summary, the bidding data on January 7, 2017 are abnormal in terms of quotation quantity, and the identification results of this method are correct.

(8) The quotation data on January 10, 2017 is similar to that on January 7, 2017 (the quotation value is close), and there is also an extreme high quotation. And there is no obvious abnormality in the daily declaration data. Therefore, the bidding data

on January 10, 2017 is abnormal in the quotation, and there is no abnormality in the declaration of quantity. The identification results of this method are correct.

(9) On February 7, 2017, there was an extreme quantity behavior. Figure 4 shows that the fifth section of the daily quantity was significantly higher. Figures 5 and 7 show that the index of index4 in daily quantity is high, the index of index5 is high, and index of index6 is low. There is no abnormality in the quotation, so the bidding data on February 7, 2017 is abnormal in the amount of quotation, and there is no abnormality in the quotation. The identification results of this method are correct.

(10) The quantity data on February 9, 2017 is similar to that on February 7, 2017 (the quantity value is close), and there are extreme cases. The quotation data on the same day also showed abnormalities, and each quotation was relatively high. The quotation data of February 9, 2017 and February 7, 2017 are compared as shown in Table 5. Therefore, the bidding data on February 9, 2017 are abnormal in terms of quotation and quantity, and the identification results of this method are correct.

Table 5. Comparison of the bidding data of the power supplier between 2017-2-7 and 2017-2-9 (USD/ MWH)

(11)	BIDDATE	(12)	BID1	(13)	BID2	(14)	BID3	(15)	BID4	(16)	BID5
(17)	7-Feb-17	(18)	10.04	(19)	15.74	(20)	24.38	(21)	99.57	(22)	100.66
(23)	9-Feb-17	(24)	15.95	(25)	24.58	(26)	39.31	(27)	114.35	(28)	116.23

After analysis and verification, the four abnormal bidding data detected by the identification method in the test set are abnormal, and the abnormal types are consistent with the identification results, which fully demonstrate the feasibility and effectiveness of the proposed method.

CONCLUSION

Firstly, this paper discusses the background and necessity of the research on abnormal behavior recognition of electricity market members, and analyzes and compares the schemes adopted by the existing research. Secondly, the characteristics of three types of abnormal bidding behaviors, including extreme quotation, extreme quantity and compound anomaly, are proposed, and the six-dimensional abnormal behavior identification index constructed in this paper is proposed. Based on this index, feature coding is carried out for different types of abnormal behaviors, and the numerical feature set of abnormal behaviors is obtained. Then, based on the traditional isolated forest algorithm, this paper proposes an abnormal bidding type identification model for power producers based on multiple isolated forests. The principle and identification process of the identification model are introduced in detail, which solves the problem that the traditional isolated forest can only judge whether there is an anomaly and cannot identify which type it belongs to. In the analysis part of the example, the historical bidding data of a power supplier in the PJM spot market is used as the initial data to verify the effectiveness of the identification model in this paper. The results of the example show that the identification model proposed in this paper can also identify the type of anomaly on the basis of judging whether there is an anomaly in the bidding data. The identification results are analyzed and verified again, and the verification results fully prove the feasibility and effectiveness of the proposed method. Finally, this paper analyzes the application scenarios of the identification model, and simply demonstrates the superiority and scalability of the method.

This paper innovatively proposes an abnormal behavior identification model of power generators based on multiple isolated forest algorithm, which solves the problem that existing methods can only detect whether there is an anomaly, but can not identify the specific type of anomaly. The framework and idea of “anomaly behavior feature coding- multiple isolated forest anomaly identification- combine identification results and table to locate the abnormal type” improve the operation efficiency and deployment cost of the abnormal behavior identification model. It provides a new method and thought to solve the problem of abnormal behavior identification of market members.

ACKNOWLEDGEMENTS

This research was funded by the State Grid Jiangsu Electric Power Company - Science and Technology Project “Key technology research on situation equilibrium analysis and market bidding risk trend in provincial power market” grant number (No. SGJS0000JYJS2100513).

REFRENCES

[1] Razmi, Buygi, Esmalifalak M. Collusion strategy investigation and detection for generation units in electricity market using supervised learning paradigm. IEEE system journal, 2021,15:146-157.

[2] Sun Bo, Deng Ruilin, Ren Bin, et al. Identification method of market power abuse of generators based on lasso-logit model in spot market. Energy, 2022. 238: 121634.

[3] Sun Qian, Xie Yuting, Hu Xiuzhen, et al. Identification of generating units that abuse market power in electricity spot market based on adaboost-dt algorithm. Mathematical problems in engineering, 2021, 2021: 1-13.
[4] He Ye, Guo Siming, Wang Yu, et al. An agent-based bidding simulation framework to recognize monopoly behavior in power markets. Energies, 2023; 16(1): 434.
[5] Nan Shang, Yi Ding, Wenqi Cui. Review of market power assessment and mitigation in reshaping of power systems. Journal of modern power systems and clean energy,2022,10(05):1067-1084.
[6] Xie Jingdong, Lu Siwei, Huang Xiying, etc. Collusion identification of electricity spot market based on semi-supervised support vector machine. China southern power grid technology, 2022, 16(05): 123-133. DOI: 10.13648/j.cnki.issn1674-0629.2022.05.015.
[7] Hua Huichun, Deng Bin, Liu Zhe, etc. Intelligent early warning of collusion in power generation enterprises based on variational auto-encoder Gaussian mixture model. Power system automation, 2022, 46(04): 188-196.
[8] Sun Bo, Deng Ruilin, Xie Jingdong, et al. Collusion bidding identification of cartel units based on ordered multivariate Logit model. Power system automation, 2021, 45(06): 109-115.
[9] Xie Jingdong, Lu Haozhe, Lu Chixin, et al. Anomaly identification electricity market based on phased outlier detection. Science and technology and engineering, 2021, 21(09): 3633-3641.
[10] Li Dunnan, Zhang Qian, Li Xiaotong, et al. Potential hazard behavior identification in electricity market based on cloud model and fuzzy Petri net. Power system automation, 2019, 43(02): 25-33.
[11] Yang Mo. Modeling analysis and application of trading irregularities in electricity market. North China Electric Power University (Beijing), 2018.
[12] Ling Li, Cheng Zhangyu, Zou Chengming. An outlier detection method combining isolated forest and local outlier factor. Computer applications and software, 2022, 39(12): 278-283.
[13] Fei Tony Liu, Kai Ming Ting, Zhi-Hua Zhou. “Isolation forest.” Data Mining, 2008. ICDM’08. Eighth IEEE International Conference on.