

Research Article

Identification of Generating Units That Abuse Market Power in Electricity Spot Market Based on AdaBoost-DT Algorithm

Qian Sun , Yuting Xie , Xiuzhen Hu , En Lu , Yi Wang , and Ning Wang 

Guangdong Power Exchange Center, Guangzhou 510000, China

Correspondence should be addressed to Xiuzhen Hu; huxiuzhen@126.com

Received 3 March 2021; Revised 5 April 2021; Accepted 29 April 2021; Published 13 May 2021

Academic Editor: Yan Gao

Copyright © 2021 Qian Sun et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The identification of generating units that abuse market power is an essential part of risk prevention in a spot market, especially in the early stage of the construction of the spot market. In this study, a model for identifying generating units that abuse market power is designed based on the AdaBoost-DT algorithm. It is targeted at the imbalance between samples of generating units that abuse market power and normal generating units in the spot market. First, the four main methods by which market power is abused by generating units in the spot market are described: collusion, economic withholding, physical withholding, and extreme quotation. Second, the specific characteristics of the four methods are analyzed, and the identification indexes for generating units that abuse market power are established. Thereafter, a sample set of generating units that abuse market power using different methods is constructed. Furthermore, a training set is formed with samples of normal generating units to construct a model based on the AdaBoost-DT algorithm, for identifying generating units that abuse market power. Finally, the spot market data of a certain region are used for an example analysis. The results show that the accuracy of model identification is 97%, which validates the method.

1. Introduction

Since the promulgation of Zhongfa [2015] No. 9 and its relevant supporting documents, China's electricity market has begun to undergo a new round of reform. This concept is proposed to discover prices in the spot market and establish an electricity market with all available trading varieties and perfect functions [1]. In July 2020, China's National Development and Reform Commission and National Energy Administration jointly issued the Notice on Strengthening the Work Related to the Power Spot Market Pilot Continuous Trial Settlement to further promote the development of the spot market. The trading rules of the spot market under the new scenario are still preliminary. Furthermore, during market trading, generating units abuse market power by exploiting the drawbacks in the market rules to obtain high profits. This severely impairs the capability for price discovery in the spot market [2]. Therefore, the establishment of a set of identification methods for units that abuse market power in the spot market and thereby maintain the safe,

stable, and reliable operation of the electricity market is highly significant.

Based on the abuse of market power by generating units, Xue et al. [3] review the problem of market power abuse in terms of research methods and control. Yan et al. [4] design a mechanism to prevent the abuse of market power by generating units, while Chen et al. [5] summarize the fundamental concept of market power, common market power monitoring indexes, and market power mitigation methods. Moreover, the study introduces the implementation status of three typical electricity markets: the United States, the United Kingdom, and Northern Europe. These studies presented the definition of market power, methods to abuse market power, and methods to mitigate market power. However, these works do not identify the abuse of market power by generating units. Li et al. [6] propose various evaluation methods to comprehensively assess market power in electricity markets from different perspectives. Liu et al. [7] consider five types of market indicators (market supply and demand, market structure, bidding strategy, supplier

position, and trading results) and propose a system for evaluating the abuse of market power by generating units. Zhao et al. [8] propose a process for implementing the comprehensive evaluation index system of the electricity market based on the multilevel fuzzy comprehensive evaluation method. In these studies, evaluation indexes based on the characteristics of data on generating units are constructed, as well as the comprehensive evaluation method to identify market power abuse. However, the comprehensive evaluation method is not applicable to large volumes of power transaction data that display high dimensionality. Wu et al. [9] focus on a simplified version of the convex hull pricing model, analyze the potential market manipulation behavior and manipulation ability of market participants in the simplified model, and propose an index to quantify market power. Dai et al. [10] build a multi-leader-follower Stackelberg game based on real-time pricing, model the strategic interaction behavior between multiple electricity retailers and users while simultaneously considering the power load uncertainty of users and the price competition among electricity retailers, which can reduce the real-time electricity price. Dai et al. [11] propose a dynamic pricing scheme based on Stackelberg game for an electric vehicle charging station with a photovoltaic system. These researchers have analyzed market power and real-time electricity price, but they have not identified the abuse of market power. Sun et al. [12] construct a method for identifying market power abuse in cartel-type generating units based on the ranked multivariate Logit model. Liu et al. [13] propose the identification of the abuse of market power in the electricity market based on the cloud model and fuzzy Petri net using the pattern recognition approach. Xu et al. [14] propose an intelligent identification method that is applicable to computer analysis to identify violations by power generation enterprises. It is based on an improved support vector machine and provides a reference for the credit evaluation of electricity market entities. These studies identify the market power abuse behavior of generating units, through intelligent identification algorithms. However, the distribution of generating units that abuse market power and normal generating units in the electricity market is unbalanced. Therefore, the accuracy of these algorithms in identifying data with unbalanced positive and negative samples has not been considered. By considering the characteristics of the spot market (high data dimensionality, large data volume, and unbalanced positive and negative samples), this study adopts the AdaBoost algorithm which has significant advantages in handling the unbalanced sample problem.

The AdaBoost algorithm is a classical integrated learning method that has been used in various fields for recognition problems. In electric power field, Li et al. [15] propose a composite perturbation classification strategy for power quality based on the conditional mutual information average optimal feature selection method and AdaBoost dynamic integrated classifier. Chen et al. [16] address the shortcomings of small current ground fault routing with unbalanced sample data, dimensional catastrophe, and high empirical risk. They propose a new method for small current

ground fault line selection based on sample data processing and the AdaBoost method. These studies validate the advantages of the AdaBoost algorithm in solving the sample data imbalance problem. However, the speed of recognition needs to be improved. Yao et al. [17] propose the AdaBoost-decision tree (AdaBoost-DT) identification method that integrates multiple features to identify partial discharges of a gas-insulated composite apparatus. Zhang et al. [18] construct a new composite DT algorithm. They design and implement a DT-based remote sensing image classification system. The AdaBoost evolving DT algorithm is proposed by Zhao et al. [19]. Their experimental results show that the AdaBoost evolving DT can achieve a high recognition accuracy in a short period of time. Thus, the advantage of the AdaBoost algorithm in solving the unbalanced sample data problem is combined with the advantages of the DT (i.e., small computational volume, high recognition accuracy, and high recognition speed). This is suitable for scenarios where few generating units abuse market power in the spot market and the market data volume is large.

This study develops a method based on the AdaBoost-DT for identifying generating units that abuse market power in the spot market. First, the overall framework for identifying generating units that abuse market power in the spot market is designed. Second, the different means of market power abuse by generating units in the spot market are analyzed. Furthermore, the identification indexes for market power abuse by generating units are constructed to form the sample and training sets of generating units that abuse market power. Thereafter, an AdaBoost-DT model for identifying generating units that abuse market power is developed. Finally, the method is applied to the spot market in a region and compared with other methods to verify its effectiveness.

2. General Rationale for Identifying Generating Units That Abuse Market Power

A method is designed based on the AdaBoost-DT technique, to identify generating units that abuse market power in the spot market by combining multiple means. The general rationale is as follows.

First, the methods by which generating units abuse market power are classified into collusion, physical withholding, economic withholding, and extreme quotation. Additionally, indicators for identifying each of these four means are developed.

Second, the data on the generating units in the spot market are used to calculate the identification indexes and construct the characteristics of generating units that abuse market power. Thereby, a sample of generating units that abuse market power in different ways is formed. Together with the normal generating unit samples, a sample set is formed.

Thereafter, an equal sample weight is assigned to each generating unit sample in the sample set, and the DT model is trained. According to the classification error rate of the DT model, the weight of the samples in the sample set is adjusted, while the weight of the misclassified samples is

increased, and the weight of the correctly classified samples is reduced. The new DT model is trained such that the new model pays more attention to the misclassified samples. Iteration is continued until the misclassified samples are sufficiently few or the iteration terminates when it attains the set value. The weighted voting method is used to combine all the DT models to form the final AdaBoost-DT model for identifying generating units that abuse market power.

Finally, a model for identifying generating units that abuse market power is used to determine such units in the market. Thereafter, the identification results are evaluated.

3. Modalities and Indicators of Market Power Abuse by Generating Units

3.1. Main Methods by Which Generating Units Abuse Market Power. Market power in the electricity market refers to the capability of market members to manipulate the electricity price in the market and maintain it at an abnormal level for a certain period of time with the aim of making a profit. There are several methods by which generating units can abuse market power [20].

3.1.1. Collusion. Collusion refers to the scenario wherein generating units participating in the market conclude an “alliance” through negotiations and contract signing and, subsequently, apply the negotiated quotation strategy to quote high prices to obtain excessive profits. Alternatively, certain generating units quote high prices to increase the market clearing price so that other generating units in the “alliance” can obtain excessive profits and take turns to “be the banker” [21]. When a generating unit has a large market share, the long-term gains that can be obtained through collusion among generating units are relatively substantial, and collusion among generating units is more likely to occur. Additionally, the quotation and its variations by colluding generating units tend to have certain similarities.

3.1.2. Withholding. Generating unit capacity withholding refers to the scenario wherein a unit does not participate in the market with its available capacity (which should be traded in the market by the unit) because of certain deliberate actions of the unit including physical and economic withholding.

(1) Physical Withholding. Physical withholding refers to the intentional underreporting of available generating capacity by the generating unit. This is mainly in the form of false claims that the generating unit is faulty and cannot generate electricity, or that the equipment is undergoing or requires overhauling. The intention is to reduce its own generating capacity and, thereby, its supply to the market. Generating units frequently misreport physical withholding as failure of the unit, to evade market regulation and thereby abuse market power. Therefore, if the target generating unit uses physical withholding, its profits are determined by analyzing the influence of the unavailability of the unit on the market clearing price [15].

(2) Economic Withholding. Economic withholding refers to the unreasonably high price (significantly higher than the cost of power generation) quoted by a generating unit for a part of its capacity that results in nongeneration by that part. While participating in the market, the quoted price of an economically withheld generating unit is frequently close to the maximum market price [22] or significantly higher than its power generation cost and historical quoted price. This results in an increase in the clearing price and, consequently, a high profit. Therefore, it is possible to determine whether a generating unit has abused its market power through economic withholding, by comparing its historical price quotations with its winning bids.

3.1.3. Extreme Quotation. Extreme quotation refers to the following acts: (1) frequently quoting a price that exceeds that of similar generating units and their own historically quoted prices at the time of market quotation and (2) frequently quoting at excessively low prices to ensure that the generating units win the bids. The main characteristic of generating units with extreme quotations is that these ensue a high number of quotations with extremely high or low prices at the time of quotation. It is possible to determine whether generating units have abused market power through extreme quotations by comparing their quoted price levels in the market with winning bids.

3.2. Indicators of Market Power Abuse by Generating Units. To better identify generating units that abuse market power in the spot market, identification indexes are constructed considering three aspects: market structure, market behavior, and market performance. This is based on the characteristics of generating units that abuse market power in different ways and considering the principles of systematicity, scientificity, and operability.

Market structure characteristics mainly reflect the market share and position of the generating units, while market behavior characteristics mainly reflect the behavior of the generating units participating in the market, including the reporting scenario and quotation strategy. Market performance characteristics mainly reflect the performance of the generating units participating in the market, including the winning scenario of the generating units. The specific definitions and formulas of the identifying indicators of market power abuse by the generating units are as follows.

3.2.1. Market Structure Category

(1) Market Share [23]. Market share is defined as the proportion of a generating unit’s generating capacity to the total generating capacity of all the generating units in the market. It is calculated using the following formula:

$$s_i = \frac{q_i}{\sum_{j=1}^N q_j}, \quad (1)$$

where s_i is the market share of the i -th generating unit in the market, N is the total number of generating units in the

market, and q_i is the generating capacity of the i -th generating unit in the market considering the maximum declared capacity of generating unit i 's current offer. The generating unit's market share characteristics are used to reflect whether it has market power. Moreover, when its market share is excessively large, it is the key generating unit in the spot market and has a decisive influence on the clearing price in the market.

(2) *Key Supplier Index* [23]. The key supplier index is the number of generating units that must be utilized to satisfy the market demand. That is, the market is undersupplied when the key supplier does not contribute. It is calculated as follows:

$$\text{OPS}_i = \frac{\sum_{j=1}^N q_j^{\text{bid}} - q_i^{\text{bid}}}{D}, \quad (2)$$

where OPS_i denotes the critical supplier index for generating unit i , q_i^{bid} denotes the declared capacity of generating unit i , q_j^{bid} denotes the declared capacity of generating unit j , D denotes the entire market demand for power at time t , and $\text{OPS}_i < 1$ is the key supplier for that time period when the generating unit has market power. This generating unit may be a key generator for the entire system or for a particular region.

3.2.2. Market Behavior Category

(1) *Weighted Average Quotation* [24]. The average quoted price of a generating unit is defined as the sum of the product of the declared tariff and declared capacity of each segment of the effective quoted segment of the unit in the spot market, divided by the declared capacity. It is calculated as follows:

$$\bar{P}_i = \frac{\sum_{h=1}^Y P_{i,h} \times q_{i,h}}{\sum_{h=1}^Y q_{i,h}}, \quad (3)$$

where \bar{P}_i is the average quoted price of generating unit i , $p_{i,h}$ represents the declared price of the h -th segment of the generating unit, $q_{i,h}$ represents the declared capacity of the h -th segment of the generating unit, and Y represents the total number of declared segments in the generating unit's quotation curve. The weighted average quotation of the generating unit reflects the quotation of the generating unit. If the weighted average quotation is higher, the generating unit is suspected of quoting a higher price.

(2) *Relative Level of Quotations*. The relative level of quoted prices reflects the overall level of the generating unit's quoted price in the market. It is calculated by the following formula:

$$\dot{P}_i = \frac{\bar{P}_i - \sum_{j=1}^N \bar{P}_j / N}{\sum_{j=1}^N \bar{P}_j / N} \times 100\%, \quad (4)$$

where \dot{P}_i is the relative level of quotations for generating unit i and \bar{P}_i is the average quoted price of the unit at present. The closer the quoted price is to the average quoted price of all the generating units in the market, the closer it is to zero. If the quoted price is significantly different from zero, it implies that

it is significantly different from the average quoted price of all the generating units in the market and that the degree of abnormal quoting by the generating unit is high.

(3) *Capacity Pricing Index* [25]. A generating unit's capacity pricing index is the sum of its declared electricity price and the product of the declared capacity multiplier. This index can reflect the declared high price of the generating unit. It is calculated according to the following formula:

$$\text{CPI}_i = \sum_{h=1}^Y (p_{i,h} * q_{i,h} * 100)^3, \quad (5)$$

where CPI_i is the capacity pricing index of generating unit i . This index can reflect the relationship between the quantity and price of the generating unit.

(4) *High Quotation Rate* [25]. High quotation rate is defined as the number of times a generating unit's offer attains the high quotation level in a cycle as a fraction of the number of offers by the generating unit in that cycle. It is calculated by the following formula:

$$R_i = \frac{\text{NUM}_{h,i}}{\text{NUM}_i}, \quad (6)$$

where R_i is the high quotation rate of generating unit i , $\text{NUM}_{h,i}$ is the number of times that an offer by generating unit i attains the high quotation level, and NUM_i is the total number of offers by generating unit i . If the high quotation rate of a generating unit is high, it indicates that it displays abnormal offer behavior with the aim of increasing the market clearing price and gaining excess profit.

(5) *Similarity in High Quotation Rates*. The high quotation rate for a generating unit is calculated as shown in (6). The formula for calculating the similarity in high quotation rate between generating units i and j over S cycles is

$$\text{HB}_{i,j} = \sum_{s=1}^S \sqrt{(R_i - R_j)^2}, \quad (7)$$

where $\text{HB}_{i,j}$ is the similarity in the generating units' high quotation rate. R_i and R_j are the high quotation rates for generating units i and j , respectively, in the s -th cycle. The lower the $\text{HB}_{i,j}$ is, the higher the likelihood of simultaneous high quotations by i and j . The likelihood of collusion between generating units with similar offers can be reflected to a certain extent by the similarity in their high quotation rates over a number of cycles. If $\text{HB}_{i,j}$ is close to zero, it is necessary to conduct focused monitoring of these two generating units.

(6) *Generating Unit Failure Rate*. Generating unit failure rate is defined as the fraction of time that a generating unit is out of service or being overhauled. It is calculated by the following formula:

$$F_i = \frac{T_{\text{fail}}}{T_{\text{all}}} \times 100\%, \quad (8)$$

where F_i denotes the failure rate of the generating unit, T_{fail} denotes the total time of failure or overhaul of the generating

unit within a period of time, and T_{all} denotes the period of time.

(7) *Correlation Coefficient of the Quotation Curve.* The correlation coefficient of the quotation curve of generating unit i reflects the correlation between this curve and the quotation curve of generating unit j . The formula is as follows:

$$R_{i,j} = \frac{\text{Cov}(P_i, P_j)}{\sqrt{\text{Var}[P_i]\text{Var}[P_j]}}, \quad (9)$$

where $R_{i,j}$ denotes the correlation coefficient of the quotation curve of generating units i and j , $\text{Cov}(P_i, P_j)$ is the covariance of the quotation series P_i and P_j , $\text{Var}[P_i]$ is the variance of P_i , and $\text{Var}[P_j]$ is the variance of P_j . The correlation coefficient of the quotation curves can reflect the degree of correlation between the quotations of the generating units: the larger it is, the higher the similarity between the quotations of the generating units is and the higher the possibility of collusion is between these. The generating units with a high correlation coefficient of quotation curve are those with the risk of "collusion."

3.2.3. Market Performance Category

(1) *Rate of Increase in Clearing Prices.* The clearing price escalation rate for a specified generating unit is the actual market clearing price minus the simulated clearing price in a fully competitive market, divided by the simulated clearing price in a fully competitive market. The simulated clearing price is obtained by modifying the declaration behavior of the generating units that abuse market power and, then, recalculating the clearing. The formula is as follows:

$$P_i^* = \frac{P_i^{\text{clear}} - P_i^{\text{clear}*}}{P_i^{\text{clear}*}} \times 100\%, \quad (10)$$

where P_i^* denotes the rate of increase in the generating unit's clearing price. P_i^{clear} denotes the actual market clearing price and $P_i^{\text{clear}*}$ denotes the simulated clearing price in a fully competitive market, both after modifying the offer of generating unit i . The higher the clearing price escalation rate, the higher the risk of market power abuse by the generating unit.

(2) *Rate of Winning Bids* [24]. The rate of winning bids is defined as the proportion of the generating unit's total winning bid to its declared total power. It is calculated by the following formula:

$$\text{WR}_i = \frac{q_i^{\text{win}}}{q_i^{\text{bit}}}, \quad (11)$$

where WR_i , q_i^{win} , and q_i^{bit} are the winning bid rate, total bid winning electricity, and total declared electricity, respectively, of generating unit i .

(3) *Out-of-Merit Capacity Index* [23]. The out-of-merit capacity index is defined as the ratio of the out-of-merit

capacity of the generating unit as a fraction of the unit's actual declared capacity to the market unsuccessful bid (the sum of the available capacity of all the generating units that participated in the bidding system and failed to win the bid) as a fraction of the market declared capacity, in a certain trading period. It is calculated using the following formula:

$$\text{OCI}_i = \frac{q_i^{\text{oc}}/q_i^{\text{bit}}}{Q^{\text{oc}}/Q^{\text{bit}}} \times 100, \quad (12)$$

where OCI_i is the out-of-merit capacity index of generating unit i , q_i^{oc} is the power of generating unit i with unsuccessful bids, q_i^{bit} is the declared capacity of generating unit i , Q^{oc} is the power with unsuccessful bids in the market, and Q^{bit} is the total declared capacity in the market.

As is evident from the definition of the out-of-merit capacity index, its magnitude is primarily related to the ratio of out-of-merit capacity to available capacity of the generating unit and the ratio of the remaining system capacity to available market capacity. In an ideal electricity market, the ratio of each generating unit's out-of-merit capacity to its available capacity should be relatively close to the ratio of the remaining capacity of the system to the available market capacity. Thus, the ideal out-of-merit capacity index should be 100. If the out-of-merit capacity index of a generating unit is less than 100 for a specified time period, it indicates that the unit's offer for that time period is normal. Otherwise, it indicates that the unit's offer for that time period is higher than that of the majority of the units in the system and that the unit may have engaged in collusive bidding behavior.

Depending on the different means by which the generating unit abuses market power (each of which has different characteristics), the specific problems are analyzed on a case-by-case basis, as shown in Table 1.

4. Model Based on AdaBoost-DT Algorithm for Identifying Generating Units That Abuse Market Power

4.1. *Sample and Sample Set of Generating Units That Abuse Market Power.* Based on the different methods by which generating units abuse market power and the indicators used to identify these, the sample of generating units that abuse market power and the sample set used for model training are constructed in the spot market CONTEXT.

First, an indicator set for generating units that abuse market power is constructed. Using spot market data and indicators for identifying market power abuse, the indicator set of generating units that abuse market power is constructed for different methods of abusing market power, as shown in

$$X_i = \{x_{i,j}^m \mid j = 1, 2, \dots, J; m = 1, 2, \dots, M\}, \quad (13)$$

where X_i denotes the set of indicators of the i -th generating unit that abuses market power. $x_{i,j}^m$ denotes the m -th indicator of the j -th risk of generating unit i . J denotes the total number of methods of market power abuse, and M denotes the number of indicators.

TABLE 1: Methods and indicators for identifying abuse of market power by generating units.

Methods of abusing market power	Indicator	Quantitative characteristics of indicator
Collusion	Market share	Large market share
	Key supplier index	Key supplier index <1
	Quotation curve correlation coefficient	Large quotation curve correlation coefficient
	Similarity in high quotation rates	Similarity in the high quote ratio close to zero
	Winning rate	High winning rate
Physical withholding	Clearing price escalation rate	High clearing price escalation rate
	Market share	Large market share
	Key supplier index	Key supplier index <1
	Failure rate	High failure rate
	Out-of-merit capacity index	Large out-of-merit capacity index
Economic withholding	Bid winning rate	Low bid winning rate
	Clearing price escalation rate	High clearing price escalation rate
	Market share	Large market share
	Key supplier index	Key supplier index <1
	High quotation rate	Large high quotation rate
Extreme quotation	Out-of-merit capacity index	Large out-of-merit capacity
	Bid winning rate	Low bid winning rate
	Clearing price escalation rate	High clearing price escalation rate
	Market share	Large market share
	Key supplier index	Key supplier index <1
Extreme quotation	Weighted average quotation	Excessively high or low weighted average quotation
	Relative level of quotations	Relative level of quotations significantly different from zero
	Capacity pricing index	Excessively high or low capacity pricing index
	Clearing price escalation rate	High clearing price escalation rate

Thereafter, a sample of generating units that abuse market power and a sample set are constructed. Based on the set of indicators obtained from (13)–(15), a sample of generating units that abuse market power and a sample set of generating units that abuse market power are created (as shown in (16)).

$$y_i = \begin{cases} j, & \sum_{m=1}^M x_{i,j}^m \cdot \omega_j^m - \gamma_j^* \geq 0, \\ 0, & \sum_{m=1}^M x_{i,j}^m \cdot \omega_j^m - \gamma_j^* \leq 0, \end{cases} \quad (14)$$

$$T_i = \{y_i, X_{i,1}, X_{i,2}, \dots, X_{i,J}\}, \quad (15)$$

$$T = \{T_1, T_2, \dots, T_n, T_{n+1}, \dots, T_N\}, \quad (16)$$

where y_i denotes the label of generating unit i . It takes values in the range $\{0, 1, 2, 3, 4\}$, representing normal, collusive, physically withheld, economically withheld, and extreme quotation generating units, respectively.

ω_j^m denotes the weight of the m -th indicator of the j -th method of market power abuse. γ_j^* is the threshold of the j -th generating unit that abuses market power. All the generating units beyond this threshold are identified as generating units that abuse market power, as determined by experts. T_i denotes the i -th generating unit sample. It includes the labels y_i of different methods of abusing market power and the indicator $X_{i,j}$ of generating units that abuse market power. T denotes the set of samples used for model training. It consists of the samples of generating units that abuse market

power (T_1, T_2, \dots, T_n) and a sample of a normal generating unit (T_{n+1}, \dots, T_N).

4.2. AdaBoost-DT Recognition Model. The fundamental concept underlying the AdaBoost algorithm [26] is as follows: (1) vary the classifier weights based on the misclassified samples by continuously iterating until a sufficiently small error rate is attained and (2) combine the different classifiers of each iteration by a strategy to form the final strong classifier (as shown in Figure 1) by first training weak classifier 1 using the original training set and, then, readjusting the weights of the samples in the training set, which is then used to train weak classifier 2. Iterations continue until a sufficiently marginal error rate is achieved. Finally, the weighted voting method is used to combine the weak classifiers. The AdaBoost algorithm has good generalization capability and practicality [27]. Different weights are assigned to different samples through cyclic training to achieve accurate classification by increasing the focus on difficult samples.

Using DT as a weak classifier, samples of generating units with different labels in the training set are used as inputs to the AdaBoost algorithm for training. Thereafter, the unknown generating unit samples are identified. The specific steps of the algorithm are as follows.

Step 1. Initialize the weight distribution of the training data.

Based on the obtained training set, each generating unit sample in the training set is assigned an identical weight in the first iteration, as shown in (17). The weights of the generating unit samples are updated in each iteration:

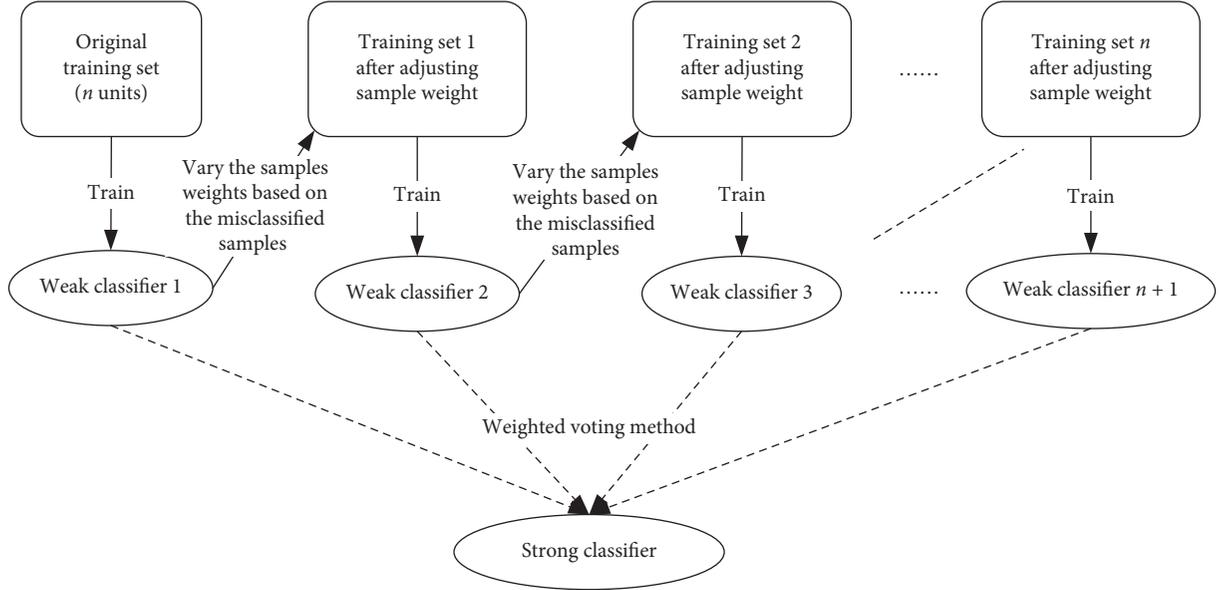


FIGURE 1: AdaBoost algorithm flow.

$$D(1) = \{\omega_{11}, \omega_{12}, \dots, \omega_{1N}\}, \quad (17)$$

where $\omega_{1i} = (1/N)$ is the weight of the first generating unit sample i in the first iteration ($i = 1, 2, \dots, N$).

Step 2. Iterative training of weak classifiers.

- (1) Denote the number of iterations by k ($k = 1, 2, \dots, K$). Set the weight coefficient of each generating unit sample at the k -th iteration as $D(k) = \{\omega_{k1}, \omega_{k2}, \dots, \omega_{kN}\}$, to obtain the k -th weak classifier $G_k(x_i)$.
- (2) Calculate the weighted classification error rate e_k for the weak classifier $G_k(x_i)$ on the training set. Here, the weighted classification error rate represents the sum of the weights of all the generating unit samples that have been misclassified by the current classifier.

$$e_k = P(G_k(x_i) \neq y_i) = \sum_{i=1}^m \omega_{ki} I(G_k(x_i) \neq y_i). \quad (18)$$

- (3) Calculate the weight coefficient of the weak classifier $G_k(x_i)$. It is combined with the weighted error rate e_k to calculate the weight coefficient of the weak classifier using (19). The weighting coefficient indicates the importance of the weak classifier $G_k(x_i)$ in the final classifier. The smaller the error rate, the larger the weight coefficient [28]. The weight coefficient of the k -th weak classifier $G_k(x_i)$ is

$$\alpha_k = \frac{1}{2} \log \frac{1 - e_k}{e_k}. \quad (19)$$

- (4) Update the weight distribution of the training dataset:

$$D(k+1) = \{\omega_{k+1,1}, \omega_{k+1,2}, \dots, \omega_{k+1,N}\}, \quad (20)$$

where $\omega_{k+1,i}$ denotes the weight of the i -th generating unit sample in the $(K+1)$ -st iteration. It is calculated as follows:

$$\omega_{k+1,i} = \frac{\omega_{ki}}{Z_k} \exp(-\alpha_k y_i G_k(x_i)), \quad (21)$$

where Z_k is the normalization factor such that $D(k+1)$ is a probability distribution. It is calculated as

$$Z_k = \sum_{i=1}^m \omega_{ki} \exp(-\alpha_k y_i G_k(x_i)). \quad (22)$$

From the above-stated equation, the weight of correctly classified generating unit samples is reduced according to the k -th iteration of classification, whereas the weight of misclassified generating unit samples increases. Misclassified generating unit samples play a larger role in the next iteration [29]. This enables each generating unit sample to be learned completely through this process.

- (5) Repeat (1)–(4) in Step 2 to obtain a series of weak classifiers and their corresponding weights.

Step 3. Linear combination of weak classifiers using weight parameters.

$$f(x) = \sum_{k=1}^K \alpha_k G_k(x). \quad (23)$$

The continuous function $f(x)$ is transformed into a discrete function using the $\text{sign}()$ function. Thus, the final identification model is

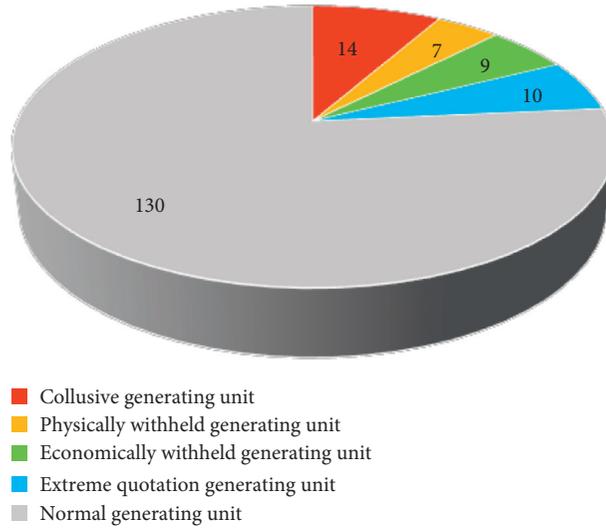


FIGURE 2: Distribution of positive and negative samples.

$$G(x) = \text{sign}(f(x)) = \text{sign}\left(\sum_{k=1}^K \alpha_k G_k(x)\right). \quad (24)$$

Use identification models to identify generating units that abuse market power in the spot market.

5. Analysis of Calculation Examples

The data on the spot market of a region in 2005 (containing information on 170 generating units) are used to validate the method proposed in this study for identifying the anomalous generating units in the spot market. The method is based on AdaBoost and DT techniques. The distribution of positive and negative samples is shown in Figure 2. Among the 170 generating units, the number of units abusing market power accounts for a quarter, including 14 collusive units, 7 physical holding units, 9 economic holding units, and 10 extreme quotation units.

5.1. Samples of Generating Units That Abuse Market Power. Considering the indicators for identifying generating units that abuse market power and the actual scenario of the spot market in a certain region, and based on the relevant data of generating units in the market, the sample of generating units abusing market power and the sample set are constructed. A few samples of the generating units that abuse market power are presented in Tables 2–5.

5.2. Indicators for Model Evaluation. In this study, accuracy, recall, and F-Measure are used to evaluate the model for identifying generating units that abuse market power. Accuracy represents the fraction of genuine positive samples out of all the samples identified as positive, recall is the fraction of all the positive samples that are identified as positive, and F-Measure is a composite index of accuracy and recall. The formulas for the three evaluation indexes are as follows:

$$\begin{aligned} \text{pre} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\ \text{rec} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ F_- &= \frac{\text{pre} \times \text{rec} \times 2}{\text{pre} + \text{rec}}, \end{aligned} \quad (25)$$

where pre denotes the accuracy of the recognition model: the higher the accuracy, the better the model being evaluated; rec denotes the recall of the recognition model: the higher the recall, the better the model being evaluated; F_- denotes the F -measure of the recognition model: the closer it is to one, the better is the model; TP denotes a positive sample that is correctly recognized; FP denotes a negative sample that is recognized as a positive sample; and FN denotes a positive sample that is recognized as a negative sample.

6. Results of Model Identification

Of the sample set of generating units that abuse market power obtained above, 70% is used as the training set to train the recognition model. The remaining (30%) is used as the validation set to verify the accuracy of the model recognition. The accuracy of the classification is low when the number of iterations of the AdaBoost algorithm is relatively marginal. However, when this number is excessively large, it causes overfitting. The validation accuracy increases with the number of iterations. The result is shown in Figure 2. Meanwhile, the time utilized for the iteration increases with the number of iterations.

Figure 3 shows that the accuracy of classification increases with the number of iterations. However, it is essentially stable when the number of iterations is higher than 30. Thus, the number of iterations of the classifier can be set as 30. For the different methods by which generating units can abuse market power, the results of the model evaluation indexes for the AdaBoost-DT identification model are shown in Figure 4.

TABLE 2: Samples of collusive generating units.

Generating unit	Market share	Key supplier index	Quotation curve correlation coefficient	Similarity in high quotation rates	Bid winning rate (%)	Clearing price escalation rate (%)
1	0.0789	0.9816	0.5504	0.5281	71.47	13.17
2	0.0680	0.9880	0.5489	0.5245	52.83	12.50

TABLE 3: Samples of physically withheld generating units.

Generating unit	Market share	Key supplier index	Failure rate (%)	Out-of-merit capacity index	Bid winning rate (%)	Clearing price escalation rate (%)
3	0.0906	0.9733	5.23	197.15	0	6.30
4	0.0525	0.9913	6.76	121.575	38.33	4.27

TABLE 4: Samples of economically withheld generating units.

Generating unit	Market share	Key supplier index	High quotation rate	Out-of-merit capacity index	Bid winning rate (%)	Clearing price escalation rate (%)
5	0.0806	0.9756	0.5	197.1477	0	5.70
6	0.0625	0.9893	0.4	108.4313	45	3.47

TABLE 5: Samples of extreme quotation generating units.

Generating unit	Market share	Key supplier index	Weighted average quotation	Relative level of quotations	Capacity pricing index	Clearing price escalation rate (%)
7	0.0960	0.9725	800	0.4652	59348.39	5.33
8	0.0555	0.9912	268.89	-0.5075	8.70	2.68

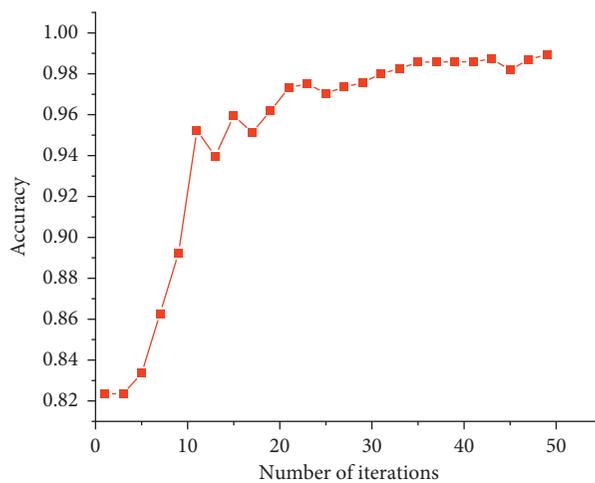


FIGURE 3: Relationship between recognition accuracy and number of iterations.

To validate the model, this study compares the results of identification using DT, SVM [14, 30], and AdaBoost-SVM, as shown in Figure 5.

To better illustrate the relationship between the accuracies of AdaBoost-DT and AdaBoost-SVM recognition with the number of iterations, the assessment of the

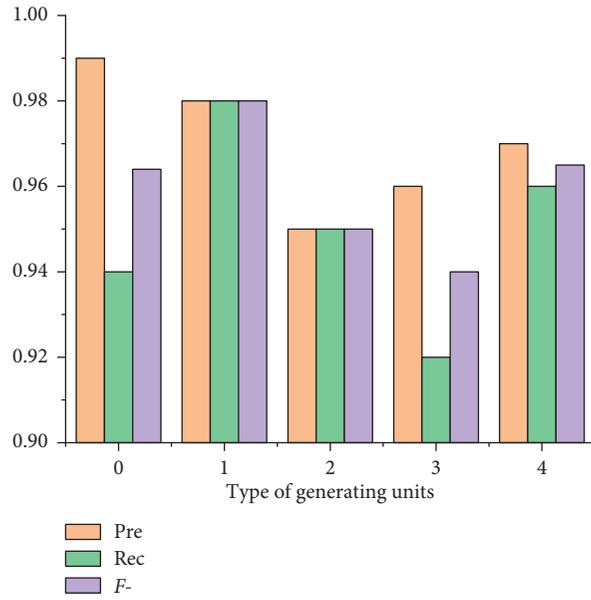
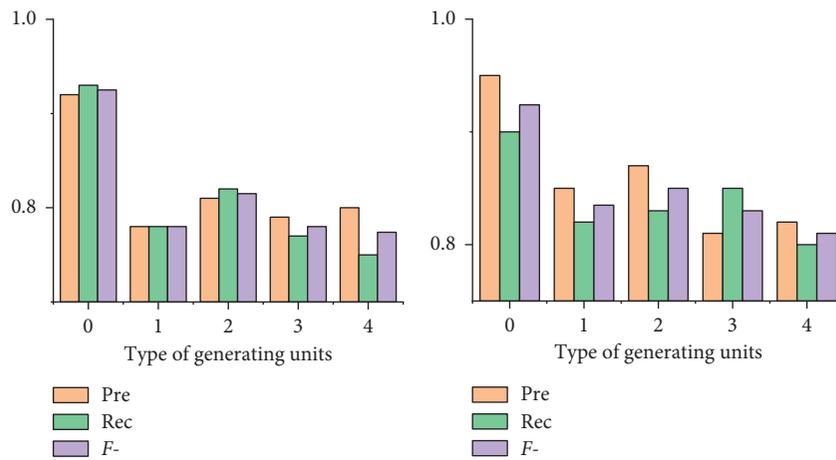
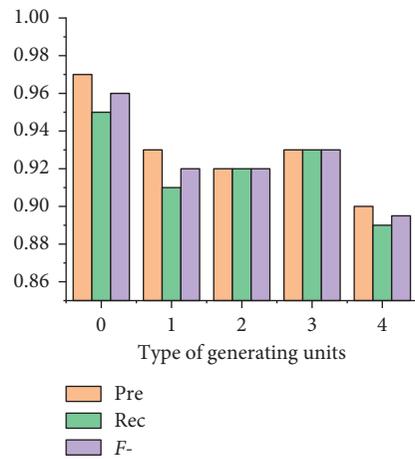


FIGURE 4: Results of the evaluation indicators of the AdaBoost-DT model.



(a)

(b)



(c)

FIGURE 5: Results of the evaluation indicators.

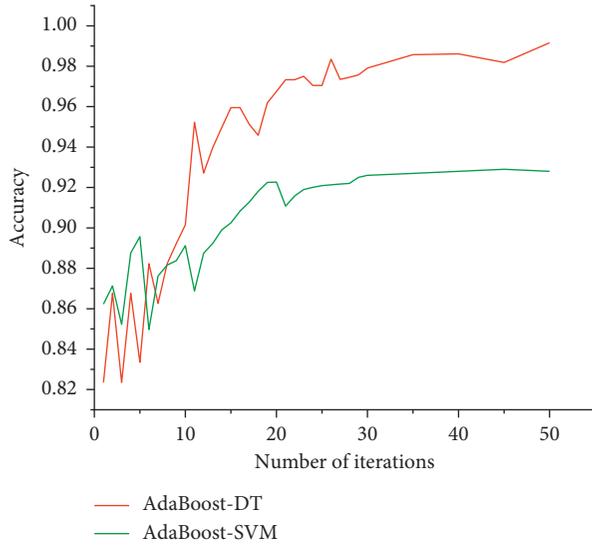


FIGURE 6: Recognition accuracy vs. number of iterations for AdaBoost-DT and AdaBoost-SVM.

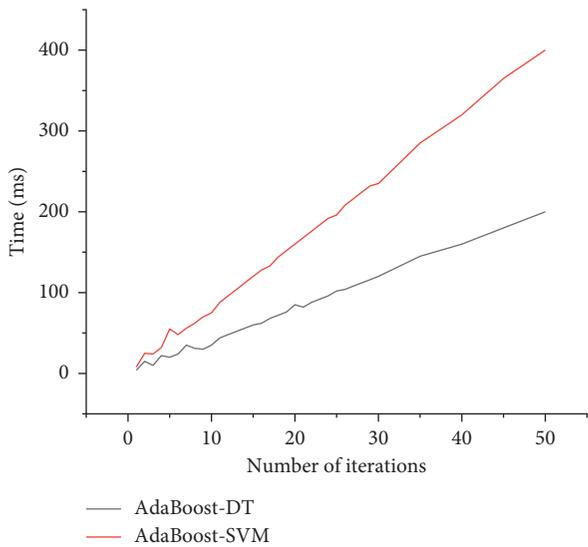


FIGURE 7: Iteration time vs. number of iterations for AdaBoost-DT and AdaBoost-SVM.

recognition results is simplified here. The accuracy of recognition is the mean of the accuracies of the five sets (0, 1, 2, 3, and 4). Figure 6 shows a comparison of the two models based on the relationship between the recognition accuracy and the number of iterations, while Figure 7 shows a comparison based on the relationship between the iteration time and the number of iterations.

Figure 6 shows that the accuracy of identification by AdaBoost-SVM is higher when the number of iterations is less than 10. Furthermore, the accuracy of identification by AdaBoost-DT is significantly higher than that for AdaBoost-SVM when the number of iterations is larger than 10. Meanwhile, it is apparent from Figure 7 that AdaBoost-DT iterates faster. In Figures 4 and 5, the pre, rec, and F -results for each of the five generating units derived from the

TABLE 6: Comparison of identification results of different classifiers.

	Pre	Rec	F -
DT	0.82	0.81	0.815
SVM	0.86	0.84	0.85
AdaBoost-DT	0.97	0.95	0.96
AdaBoost-SVM	0.93	0.92	0.93

different models are averaged to obtain the results of the evaluation indexes of the different models. The specific results are shown in Table 6.

The experiment results demonstrate that the recognition effect of the SVM model is marginally higher than that of the DT model when the AdaBoost algorithm is not used for integration learning. When it is used for integration learning, the recognition accuracy of the AdaBoost-DT model improves from 0.76 to 0.97 relative to the single DT model, and the recognition effect significantly improves. Compared with the single SVM model, the recognition accuracy of the AdaBoost-SVM model improves from 0.82 to 0.93. The recognition effect is also improved. Furthermore, the recognition accuracy of the AdaBoost-DT model is higher than that of AdaBoost-SVM, which indicates the effectiveness of the former for identifying groups that abuse market power.

7. Conclusion

This study analyzes the four methods market power abuse by generating units based on the actual scenario in the spot market, constructs identification indexes of market power abuse by generating units based on three dimensions (market structure, behavior, and performance), and develops an AdaBoost-DT identification model by constructing samples of generating units that abuse market power and sample sets for model training according to the data characteristics of generating units of the spot market. The conclusions are as follows:

- (1) This study makes use of the advantage of the AdaBoost-DT algorithm in unbalanced sample identification. The AdaBoost-DT algorithm displays a high identification accuracy of generating units from unbalanced positive and negative samples in the spot market. Additionally, it is fast and robust; that is, it can rapidly and effectively identify the generating units that abuse market power.
- (2) This study compares the AdaBoost-DT algorithm with the DT algorithm, as well as the AdaBoost-SVM algorithm and the SVM algorithm. The identification accuracies of the DT and SVM algorithms are not as high as those of the AdaBoost-DT and AdaBoost-SVM algorithms. Notably, when the number of iterations is small, the AdaBoost-SVM algorithm has a higher accuracy, but with an increase in the number of iterations, the AdaBoost-DT algorithm shows a better performance.
- (3) With the increase of iterations of the AdaBoost algorithm, the recognition accuracy of abusing market power units becomes increasingly higher, but the

recognition time also lengthens. Therefore, considering the recognition accuracy and recognition time, an appropriate number of iterations is selected.

Data Availability

The data used to support the findings of this study were supplied by Guangdong Power Exchange Center under license and so cannot be made freely available. Requests for access to these data should be made to Yuting Xie (597862406@qq.com).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Science and Technology Project Assistance of China Southern Power Grid Corporation (Project No. GDKJXM20200211).

References

- [1] CPC Central Committee and State Council, *Some Opinions on Further Deepening Power System Reform (ZF [2015] No. 9)*, CPC Central Committee and State Council, Beijing, China, 2018.
- [2] Z. Wu, C. Kang, Q. Xia et al., "Bidding strategy of power suppliers based on game theory," *Power System Automation*, vol. 26, no. 9, pp. 7–11, 2002.
- [3] Y. Xue, T. Li, X. Yin et al., "Research framework of generalized congestion and market power," *Power System Automation*, vol. 34, no. 21, pp. 1–10, 2010.
- [4] M. Yan, Z. Gong, Y. Yang et al., "Design of market operation monitoring mechanism under Yunnan electric power spot market environment," *China Southern Power Grid Technology*, vol. 12, no. 12, pp. 16–22, 2018.
- [5] Q. Chen, J. Yang, Y. Huang et al., "Overview of market power monitoring and mitigation mechanism in foreign power markets," *China Southern Power Grid Technology*, vol. 12, no. 12, pp. 9–15, 2018.
- [6] C. Li, Q. Xia, and Z. Hu, "Evaluation method of power plant market power based on effective competition," *Power Grid Technology*, vol. 30, no. 10, pp. 75–80, 2006.
- [7] D. Liu, R. Li, X. Chen et al., "Power market supervision index and market evaluation system," *Power System Automation*, vol. 2004, no. 9, pp. 16–21, 2004.
- [8] Z. Zhao, S. Yuan, Q. Nie et al., "Continuous reinforcement algorithm and robust economic dispatching-based spot electricity market modeling considering strategic behaviors of wind power producers and other participants," *Journal of Electrical and Computer Engineering*, vol. 2019, Article ID 9406072, 16 pages, 2019.
- [9] C. Wu and J. Sun, "Analysis method of potential market power in electricity market under convex hull pricing mode," *Power System Automation*, vol. 45, no. 6, pp. 101–108, 2021.
- [10] Y. Dai, Y. Gao, Y. Gao, H. Gao, H. Zhu, and L. Li, "A real-time pricing scheme considering load uncertainty and price competition in smart grid market," *Journal of Industrial & Management Optimization*, vol. 16, no. 2, pp. 777–793, 2020.
- [11] Y. Dai, Y. Qi, Li Lu et al., "A dynamic pricing scheme for electric vehicle in photovoltaic charging station based on Stackelberg game considering user satisfaction," *Computers & Industrial Engineering*, vol. 154, no. 10, Article ID 107117, 2021.
- [12] B. Sun, R. Deng, J. Xie et al., "Identification of cartel type unit collusion bidding based on ordered multivariate logit model," *Power System Automation*, vol. 45, no. 6, pp. 109–115, 2021.
- [13] D. Liu, Q. Zhang, X. Li et al., "Identification of potential hazardous behaviors in power market based on cloud model and fuzzy Petri net," *Power System Automation*, vol. 43, no. 2, pp. 25–37, 2019.
- [14] H. Xu, Z. Cheng, H. Zhang et al., "Identification of market power abuse violations by power generation enterprises based on improved support vector machine," *Journal of North China Electric Power University (Natural Tural Science Edition)*, vol. 47, no. 04, pp. 86–95, 2020.
- [15] C. Li, K. Liu, X. Xiao et al., "Power quality composite disturbance classification based on conditional mutual information feature selection method and AdaBoost algorithm," *High Voltage Technology*, vol. 45, no. 2, pp. 579–585, 2019.
- [16] K. Chen, W. Xiaoguang, J. Chen et al., "Small current grounding fault line selection based on sample data processing and AdaBoost," *Chinese Journal of Electrical Engineering*, vol. 34, no. 34, pp. 6228–6237, 2014.
- [17] L. Yao, W. Zheng, Y. Qian et al., "Partial discharge comprehensive feature decision tree recognition method based on AdaBoost," *Power System Protection and Control*, vol. 39, no. 21, pp. 104–114, 2011.
- [18] X. Zhang, Z. Liang, J. Zhang et al., "Application of Adatree algorithm in remote sensing image classification," *Journal of Wuhan University (Information Science Edition)*, vol. 38, no. 12, pp. 1460–1464, 2013.
- [19] M. Zhao, E. Chen, and R. Song, "AdaBoost evolutionary decision tree algorithm based on Ensemble Learning," *Computer Applications and Software*, vol. 2007, no. 3, pp. 1–21, 2007.
- [20] J. Lin, Y. Ni, and F. Wu, "Review of market power in power market," *Power Grid Technology*, vol. 2002, no. 11, pp. 70–76, 2002.
- [21] Q. Yan, C. Qin, M. Nie et al., "Forecasting the electricity demand and market shares in retail electricity market based on system dynamics and Markov chain," *Mathematical Problems in Engineering*, vol. 2018, Article ID 4671850, 11 pages, 2018.
- [22] H. Zhou, F. Zhang, and Z. Han, "Research on economic persistence in power market," *Power System Automation*, vol. 2005, no. 8, pp. 16–20, 2005.
- [23] L. Liu, H. Gao, Y. Wang et al., "Robust optimization model for photovoltaic power producer's bidding decision-making in electricity market," *Mathematical Problems in Engineering*, vol. 2020, Article ID 6109648, 8 pages, 2020.
- [24] J. Xie, C. Lu, S. Lu et al., "Risk assessment of power market based on ordinal relation entropy weight method," *China Power*, vol. 1, no. 12, 2021.
- [25] J. Chen, H. Zhou, Z. Han et al., "Application of volume price index to power supplier quotation analysis," *Journal of Zhejiang University (Engineering Edition)*, vol. 2005, no. 6, pp. 915–920, 2005.
- [26] S. Wu and H. Nagahashi, "Parameterized AdaBoost: introducing a parameter to speed up the training of real AdaBoost," *IEEE Signal Processing Letters*, vol. 21, no. 6, pp. 687–691, 2014.

- [27] P.-B. Zhang and Z.-X. Yang, "A novel AdaBoost framework with robust threshold and structural optimization," *IEEE Transactions on Cybernetics*, vol. 48, no. 1, pp. 64–76, 2018.
- [28] W. Hu, J. Gao, Y. Wang, O. Wu, and S. Maybank, "Online adaboost-based parameterized methods for dynamic distributed network intrusion detection," *IEEE Transactions on Cybernetics*, vol. 44, no. 1, pp. 66–82, 2014.
- [29] F. Xiao, Y. Wang, L. He, H. Wang, W. Li, and Z. Liu, "Motion estimation from surface electromyogram using adaboost regression and average feature values," *IEEE Access*, vol. 7, pp. 13121–13134, 2019.
- [30] Y. Dai and P. Zhao, "A hybrid load forecasting model based on support vector machine with intelligent methods for feature selection and parameter optimization," *Applied Energy*, vol. 279, no. 1, Article ID 115332, 2020.