Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 879-898
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

Method to Discriminate Collusive Bidding among Power Producers based on Deep Self-Coding Gaussian Mixture Model

Bin Ren*
Shanghai University of Electric Power, Shanghai 200090, China
* Corresponding Author.

Abstract:

With the increasing transaction scale of the power market and the corresponding increase in transaction data, it is urgent to analyze power companies' collusive bidding behavior based on big data. Therefore, this paper combines an indexing system and unsupervised deep-autoencoder Gaussian mixture model to detect collusive bidding. First, this paper introduces an indexing system that includes structure, behavior, and influence indexes. Second, the Gaussian mixture model of the deep autoencoder is proposed according to the characteristics of high-dimensional index data. Then, using the compression network of the deep autoencoder, the hidden representation and reconstruction errors of the unit index data are obtained, and the loss function is constructed based on this data. The estimation network is then used to estimate the density, and the abnormal energy function of the unit is constructed to determine whether collusive bidding behavior occurs. Finally, a numerical example shows that compared with other traditional unsupervised study models, the deep-autoencoder Gaussian mixture model is more efficient and accurate in discriminating the bidding behavior of power producers.

Keywords: *Electricity market, Power suppliers, Collusive bidding, Deep self-coding Gaussian mixture model.*

I. INTRODUCTION

In the electricity market, the collusive bidding behavior of power producers is one of the main forms of market power abuse.[1] In general, power producers collusively bid to obtain excess profits through similar and reverse bids. Traditionally, the main methods used to evaluate collusive bidding are index analysis [2] and post-detection.[3] At present, the power market in China is highly concentrated, and the Herfindahl–Hirschman index (HHI) is obviously higher than the threshold value. [4] Therefore, it is necessary to strengthen the top-level design of power-market credit supervision and predict the hidden risks in the power market in order to detect the collusive bidding behavior of power producers in real time.

At present, the research on power generation companies' collusive bidding can be roughly divided into two categories: research on the criterion of the collusive bidding of power generation companies [5] and research on the method of collusive bidding discrimination [6].

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 879-898
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

As for the research on indicator systems of power-generation companies' collusive bidding, researchers [7] have constructed a relatively ideal evaluation-indicator system of collusive bidding from three perspectives. Systems [8] have also considered the structure, supply and demand, and performance of the market, thus establishing an indicator system for power generation companies' collusive bidding behavior. All these studies introduced an indexing system to determine whether power-generation companies’ engage in collusive bidding, but the correlation of the collusive bidding index was not considered. Studies [9] have explored the structure of the market itself and whether there is a possibility of collusive bidding based on this perspective. Another study [10] applied the Lerner index to market power and evaluated the possibility of market collusive bidding using this index. In addition, the possibility of collusive bidding in an electricity market was gauged based on the overall price level of the market.[11]

As for the research on the discriminant method of power-generation companies' collusive bidding, the literature [12] has considered the superiority of the cloud model in power-market data processing and used fuzzy networks to identify collusive bidding behavior. The fuzzy comprehensive evaluation algorithm was applied to power supervision in order to identify units with collusive bidding, [13] and the collusive bidding behavior of power producers was determined on the basis of a fuzzy game model. [14] Although these studies comprehensively considered the characteristics of power-market data to build an indexing system for power-generation companies' collusive bidding, they did not consider the correlation between the indices and the bidding characteristics of different units. In a previous study, [15] the index database of power producers was constructed, and the pattern-recognition algorithm was adopted to detect illegal bidding behaviors. However, this method cannot self-update and adapt to market changes in real time. Collusive bidding among power producers has been regarded as a binary classification problem, and a supervised learning algorithm has been used to train classifiers to identify the characteristics of the collusive bidding index. [16] However, in practice, there is very few labeled data, so it is difficult to train a supervised learning model with good generalization ability. Therefore, an unsupervised learning model is a better choice because the collusive bidding behavior is actually the abnormal behavior of the generator; units with collusive bidding can be identified by applying the concept of unsupervised learning anomaly detection.

The deep-autoencoding Gaussian mixture model (DAGMM) is an unsupervised anomaly-detection algorithm that has advantages in processing complex and high-dimensional data in the power market. In the current study, an unsupervised deep-autoencoder Gaussian mixture model composed of a compression network and estimation network [17] was used to examine the bidding behavior of power producers. In each iteration, the compressed network could obtain the hidden, characteristic information of the index data of the units in the high-dimensional power market and estimate the density of the units in the low-density area as abnormal units, that is, the units that have colluded bidding, [18] thus allowing collusive bidding discrimination.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 879-898
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

II. INDEXING SYSTEM FOR POWER-GENERATION COLLUSIVE-BIDDING DISCRIMINATION

2.1 Construction Principle of Discriminant Indexing System

A set of scientific indices for determining whether power producers are engaging in collusive bidding is indispensable to the operation of the power market and particularly to supervisory departments that deal with infringements. Under the current environment of power-system reform in China, many conditions are interrelated, including the operation of the power market, the production of electricity, the increasing demand for electricity in developing society, the bidding game between power producers, the inherent security constraints of the power grid, and the external environment. Collusive bidding between power producers is often reflected in the form of several or a series of indicators. Therefore, it is of great practical significance to construct a collusive-bidding discrimination system in order to promote healthy and stable development of the power market.

This paper considers the internal factors of the transaction behavior between power producers and the external factors of the market environment and divides the collusive-bidding indexing system into structure, influence, and behavior indices.

2.2 Structural Indicators

The structure index represents the ability of the unit to participate in collusive bidding and represents the position of the unit in the power market. Structural indicators are composed of the unit market share and key supplier coefficient.

2.2.1 Unit market share

Unit market share refers to the ratio of the total declared generating capacity of the generator set to the total declared generating capacity of all the generator sets in the market, which is calculated as follows:

$$A_{n,t} = \frac{p_{n,t}}{p_t} \times 100\% \tag{1}$$

where $A_{n,t}$ is the unit power generation share of generator set n at time t, $p_{n,t}$ is the declared generating capacity of generator set N at time T, and p_t is the total available power generation capacity in the market at time T.

If the unit market share index is large, the unit possesses high pricing power in the market at that time and can use this pricing power to control the market electricity price in each declaration process.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 879-898
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

2.2.2 Unit critical supplier coefficient

The unit key supplier coefficient represents the percentage of the difference between the declared capacity of all generating sets and the declared capacity of this generating set compared with the total power demand of the market. It can be calculated as follows:

$$KS_{n,t} = \frac{\sum_{m=1}^{M} q_{m,t} - q_{n,t}}{D_t} \tag{2}$$

where $KS_{n,t}$ is the key supplier coefficient of unit n at time t, $q_{m,t}$ represents the declared capacity of unit M at time t, $q_{n,t}$ denotes the declared capacity of unit N to be considered at time t; and D_t represents the power demanded by the whole market at time t.

The unit critical supplier coefficient reflects the position of the generator set in the market. If the key supplier coefficient of the generator set is low, the generator set is in a critical position. In the peak hours of electricity consumption, power resources are tight and the price elasticity is too low to meet the market demand. Even if the declaration of high price is considered as output, the possibility of collusive bidding will also increase.

2.3 Behavioral Indicators

Behavioral indicators represent the following characteristics of the unit-declaration stage: average quotation, high quotation ratio, quotation volatility, volume price index, and high quotation volume price index.

2.3.1 Average price

Average quotation refers to the sum of the product of the declared electricity price of each section of the unit and the declared capacity, divided by the total declared capacity of the unit. It can be calculated as follows:

$$\overline{B}_{n,t} = \sum_{h=1}^{H} \frac{b_{n,t,h} \times q_{n,t,h}}{q_{n,t}} \tag{3}$$

where $\overline{B}_{n,t}$ is the weighted average quotation of generator set n at time t; h is the number of declared sections of the generator set; $h = 1,2,...,H$, $q_{n,t,h}$ is the declared capacity of section H of generator set N at time T; and $b_{n,t,h}$ is the declared electricity price of section H of generator set N at time t.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 879-898
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

The average price can reflect the declared price of the generator sets, which can be used to judge whether there is abnormal declaration behavior of the generator sets and thus collusive bidding through abnormal bidding behavior.

2.3.2 High rate of offer

High quotation ratio refers to the proportion of the high quotation capacity in the capacity declaration of the generator set, which is calculated as follows:

$$R_{n,t}^{high} = \frac{q_{n,t}^{high}}{q_{n,t}} \tag{4}$$

Where $R_{n,t}^{high}$ is the high quotation rate of generator set N at time T and $q_{n,t}^{high}$ is the capacity of generator set N quoted at high price at time t. The high price threshold can be set by experts or adjusted according to the overall market quotation, such as by taking the upper quartile of all units quoted or the clearing electricity price predicted by the system.

A high bidding ratio indicates that the generator set has the intention to raise the bidding behavior in the declaration process, where the larger the index, the higher the intention of the generator set to raise the market clearing price and the greater the possibility of collusive bidding.

2.3.3 Quoted volatility

Quotation volatility is the ratio of the difference between the average quotation of the generator set at a certain moment and the average quotation of the generator set at the previous moment divided by the average quotation of the generator set at this moment, as represented by the following equation:

$$Bid_{n,t} = \frac{\overline{B}_{n,t} - \overline{B}_{n,t-1}}{\overline{B}_{n,t}} \tag{5}$$

where $Bid_{n,t}$ is the quoted volatility of generator set N at time t, and $\overline{B}_{n,t-1}$ is the average quotation of generator set N at time $t-1$.

The index of the generator-set quotation volatility can reflect the change of the generator set quotation. A positive index indicates that the generator set quotation rises, whereas a negative value indicates that the generator set quotation decreases. The higher the price fluctuation of the generator set, the more different the bidding strategy adopted by the generator set and the greater the possibility of collusive bidding.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 879-898
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

2.3.4 Quantity and price index

The volume price index reflects the tendency of generating units to offer high prices and can be calculated as follows:

$$I_{n,t}^{\mathrm{CP}} = \sum_{h=1}^{H} \left(\frac{b_{n,t,h}}{c_{\mathrm{ost}}} \times \frac{q_{n,t,h}}{q_{n,t}} \times 100 \right)^{d} \tag{6}$$

where $I_{n,t}^{\mathrm{CP}}$ is the volumetric price index of generator set N at time T, c_{ost} is the average generation cost of the system, H is the declaration segment in the unit-declaration curve, H is the total number of declaration segments in the unit-declaration curve, and D is a power index with a value of 2–4 and can be adjusted according to the actual market situation.

The higher the volume price index of the generator set, the greater the concentration of the declared capacity of the generator set in the high price region, the more likely it is to raise the market clearing price, and the higher the possibility of collusive bidding behavior.

2.3.5 High quote volume-price index

The high-bid volume-price index can reflect the severity of a unit's high-bid behavior and is represented by the following equation:

$$CPI_{n,t}^{\mathrm{CP}} = \sum_{hi=1}^{HI} \left(\frac{b_{n,t,h}}{c_{\mathrm{ost}}} \times \frac{q_{n,t,hi}}{q_{n,t}^{high}} \times 100 \right)^{d} \tag{7}$$

where $CPI_{n,t}^{\mathrm{CP}}$ is the high-quoted volume-price index of unit N at time t, HI is the high-price declaration segment in the unit-declaration curve, hi is the total number of high-price declaration segments in the unit-declaration curve, and D is a power index in the range of 2–5 and can be adjusted according to the actual market situation. The higher the high-bid volume-price index of the generator, the more serious the high-bid behavior of the generator set, the greater the degree of influence on the market-clearing price when the market is not tense, and the more serious the degree of collusive bidding behavior.

2.4 Impact Indicators

The influence index represents the characteristics of units in the cleaning stage and is composed of winning bid rate, high-price winning bid rate, and marginal unit online rate.

2.4.1 Winning rate

The winning rate refers to the ratio of the bid-winning capacity of the generator set to the declared capacity of the generator set. It can be calculated as follows:

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 879-898
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

$$R_{n,t}^{win} = \frac{Q_{n,t}}{q_{n,t}} \tag{8}$$

where $R_{n,t}^{win}$ is the winning rate of generator set N at time T, and $Q_{n,t}$ is the bid-marked capacity of generator set N at time t.

The higher the winning bid rate, the greater the potential market power of the generator set, and the more likely there is to be collusive bidding behavior of the generator set.

2.4.2 High bid-winning rate

The high bid-winning rate refers to the ratio of the capacity that is bid and won by the generator set to the total bid-winning capacity, and is calculated as follows:

$$R_{n,t}^{hw} = \frac{Q_{n,t}^{high}}{Q_{n,t}} \tag{9}$$

where $R_{n,t}^{hw}$ is the high winning-bid rate of generator set N at time T, and $Q_{n,t}$ is the capacity of the generator set N at time t that is quoted at a high price and won the bid.

The high bid rate reflects the generator set bid and approaches the declared capacity. Therefore, the higher the index, the stronger its ability to manipulate the market price of electricity and the greater the possibility of collusive bidding.

2.4.3 Marginal unit lineup rate

Marginal unit lineup rate refers to the ratio of the number of generator sets becoming marginal units to the number of generator sets reporting within a certain period. It can be calculated as follows:

$$R_{n}^{lim} = \frac{T_{n}^{lim}}{T_{n}} \tag{10}$$

where R_{n}^{lim} is the marginal unit lineup rate of generator set N, T_{n}^{lim} is the number of generator sets n becoming marginal units, and T_{n} is the reporting time of generator set N.

The higher the marginal unit reaches the limit rate, the more times the generator set becomes a marginal unit, and the more abnormal the transaction behavior of the generator set. This abnormal behavior should attract the attention of the market supervision department.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 879-898
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

2.5 Indicator Architecture

A set of collusive bidding discrimination indices for power generation companies is shown in TABLE I.

TABLE I. Evaluation index system of collusive bidding among power producers

LEVEL INDICATORS	SECONDARY INDICATORS	INDEX SYMBOL
STRUCTURAL INDEX	Unit market share	$A_{n,t}$
	Unit critical supplier coefficient	$KS_{n,t}$
BEHAVIORAL INDICATORS	Average price	$\overline{B}_{n,t}$
	High rate of offer	$R_{n,t}^{\mathrm{high}}$
	Quoted volatility	$Bid_{n,t}$
	Quantity and price index	$I_{n,t}^{\mathrm{CP}}$
	High quote volume-price index	$CPI_{n,t}^{\mathrm{CP}}$
INFLUENCE INDEX	Winning rate	$R_{n,t}^{win}$
	High bid-winning rate	$R_{n,t}^{hw}$
	Marginal unit lineup rate	R_{n}^{lim}

III. ESTABLISHMENT OF COLLUSIVE-BIDDING DISCRIMINATION MODEL

In the Gaussian mixture model based on the deep autoencoder, the compression network is composed of a convolutional self-coding network, and the estimation network is a deep neural network, as shown in Figure 1. In the figure, x is the input of the compression network; z_c is the hidden layer of the compression network, i.e., the low-dimensional representation x after compression; and x' is the z_c reconstruction input obtained after the restoration of the compression network; and z_r is the reconstruction error between x and x'. The estimation network on the right of the figure is a multilayer neural network with input z, composed of z_c and z_r. The network calculates the membership of the mixture z by iteratively estimating the parameters of each mixing component in the Gaussian mixture model, setting a threshold based on the calculated abnormal energy to determine whether z is abnormal. In this paper, "abnormal" refers to the collusive bidding behavior of the generator.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 879-898
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

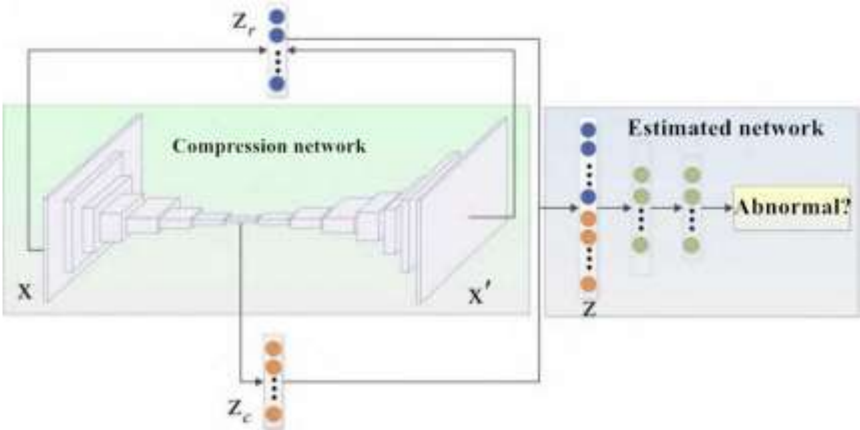

Figure 1. Discrimination method of Gaussian mixture model based on deep autoencoder.

Firstly, the self-coding part of the compression network is used to compress the index data matrix of all units in the input market x into a low-dimensional representation z_c, as shown in Figure 1.

$$z_c = f(x, \theta_e) \tag{11}$$

where x is a matrix $N*M'$, M is the number of selected indicators, N is the total number of units, z_c is a matrix $N*M$, M' is a compressed dimension that can be set artificially, f Is the encoding function, and θ_e is the encoding parameter.

The decoding function can be used to restore z_c to x' as follows:

$$x' = g(z_c, \theta_d) \tag{12}$$

where x' is the matrix of $N*M$, g is the decoding function, and θ_d is the decoding parameter. The encoding and decoding functions are usually nonlinear. The typical encoding and decoding functions are rectilinear ReLU [19] and sigmoidal, respectively.

Meanwhile, the reconstruction error z_r between x and x' is obtained, as shown in Equation (13):

$$z_r = h(x, x') \tag{13}$$

where z_r is a matrix $N*V$, and V depends on the set reconstruction error function, which can be considered as the combination of cosine similarity between x and x' and relative Euclidean distance.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 879-898
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

In this case, the codec parameters θ_e and θ_d can be determined by the reconstruction error to obtain the minimum value of z_r, as shown in Equation (14).

$$\theta_e, \theta_d = \arg\min_{\theta_e, \theta_d} h(x, g(f(x, \theta_e), \theta_d)) \tag{14}$$

where z is formed by z_c and z_r, as shown in Equation (15):

$$z = [z_c, z_r] \tag{15}$$

Then, in the estimation network, z is used as the input, and a multilayer feedforward neural network is used. In the Gaussian mixed distribution, K is the number of Gaussian distributions.

In the training phase where the distribution of the mixed components ϕ, the mixed mean μ, and the mixed covariance $\sum$ are unknown, the estimation network estimates the parameters of the Gaussian model and evaluates the abnormal energy of each unit. Here, the estimation network predicts the mixed membership of each unit $\breve{\gamma}$ through the input z and parameters of the multilayer neural network θ_m, and $\breve{\gamma}$ represents the K dimension vector used for membership prediction of the soft mixing components.

The parameters of the Gaussian mixture model are estimated as follows:

$$\hat{\phi}_k = \sum_{n=1}^{N} \frac{\hat{\phi}_{nk}}{N} \tag{16}$$

$$\hat{\mu}_k = \frac{\sum_{n=1}^{N} \breve{\gamma}_{nk} z_n}{\sum_{n=1}^{N} \breve{\gamma}_{nk}} \tag{17}$$

$$\hat{\Sigma}_k = \frac{\sum_{n=1}^{N} \breve{\gamma}_{nk} (z_n - \hat{\mu}_k)(z_n - \hat{\mu}_k{}^T)}{\sum_{n=1}^{N} \breve{\gamma}_{nk}} \tag{18}$$

where $\breve{\gamma}_n$ represents the membership prediction of z_n; z_n represents the input of the unit n to the estimated network; $\hat{\phi}_k$, $\hat{\mu}_k$, and $\hat{\Sigma}_K$ represent the mixing probability, mean value, and variance, respectively, of the Gaussian model component k in the Gaussian mixture model, where $1 \le k \le K$. In summary, the abnormal energy of unit n can be expressed as follows:

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 879-898
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

$$E(z_n) = -\log\{\sum_{k=1}^{K} \hat{\phi}_k \frac{\exp(-\frac{1}{2}(z_n - \hat{\mu}_k)\hat{\Sigma}_k^{-1}(z_n - \hat{\mu}_k^{T}))}{\sum_{n=1}^{N} \breve{\gamma}_{nk}}\} \tag{19}$$

Considering the above characteristics of the compression and estimation networks, the final optimization objective function of this Gaussian mixture model can be formulated as follows:

$$J(\theta_e, \theta_d, \theta_m) = \frac{1}{N}\sum_{n=1}^{N} L(x_n, x_n') + \frac{\lambda_1}{N}\sum_{n=1}^{N} E(z_n) + \lambda_2 P(\Sigma) \tag{20}$$

where N is the total number of units participating in the training, and λ_2 and λ_1 are used to balance the three components. The first item $\sum_{n=1}^{N} L(x_n, x_n')$ represents the sum of the reconstruction errors of all units formed by using the deep autoencoder. If a good compression network is used, this value will be lower, leaving key data with more characteristic samples. The second item $\sum_{n=1}^{N} E(z_n)$ represents the sum of the abnormal energies of all the units. In order to obtain the best combination through the compression and estimation networks, it is necessary to minimize the sum of the abnormal energy of all units. The third part of the equation is the artificial penalty term, which is applied to avoid calculation difficulties caused by the irreversibility of the matrix.

If a unit y can be identified as a normal unit, then the unit must be close to one or several Gaussian components, so the abnormal energy of the unit will be relatively low. Similarly, it is difficult for the abnormal unit to be close to any Gaussian component analyzed, so the probability of it belonging to all Gaussian components will be small, and the abnormal energy will be high.

During the prediction, the abnormal energy of the unit z_y is first obtained according to equations (11) to (15), and then the abnormal energy z_y, calculated using Equation (19), is used to determine whether the unit y belongs to the abnormal unit by setting the energy threshold ς.

IV THREE STEPS OF COLLUSIVE BIDDING DISCRIMINATION

As shown in Figure 2, the specific steps are as follows:

Step 1: Use the indexing system and calculation method of collusive bidding to calculate the original data and obtain the index data matrix x for all units.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 879-898
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

Step 2: Normalize x and divide x into the training set and test set.

Step 3: Use the DAGMM model to discriminate collusive bidding. Firstly, the low-dimensional representation z_c and reconstruction errors z_r of all units are obtained by using the self-coding part of the compression network. Then, the parameters of the Gaussian mixture model are estimated by using the multilayer neural network in the estimation network. Finally, according to the estimated parameters, the abnormal energy of each unit is calculated to determine whether collusive bidding behavior occurs.

Step 4: Submit the list of power producers suspected of collusive bidding to the relevant regulatory authorities for further investigation. In addition, check the physical MAC addresses and network IP addresses of the transaction declarations as well as the account books.

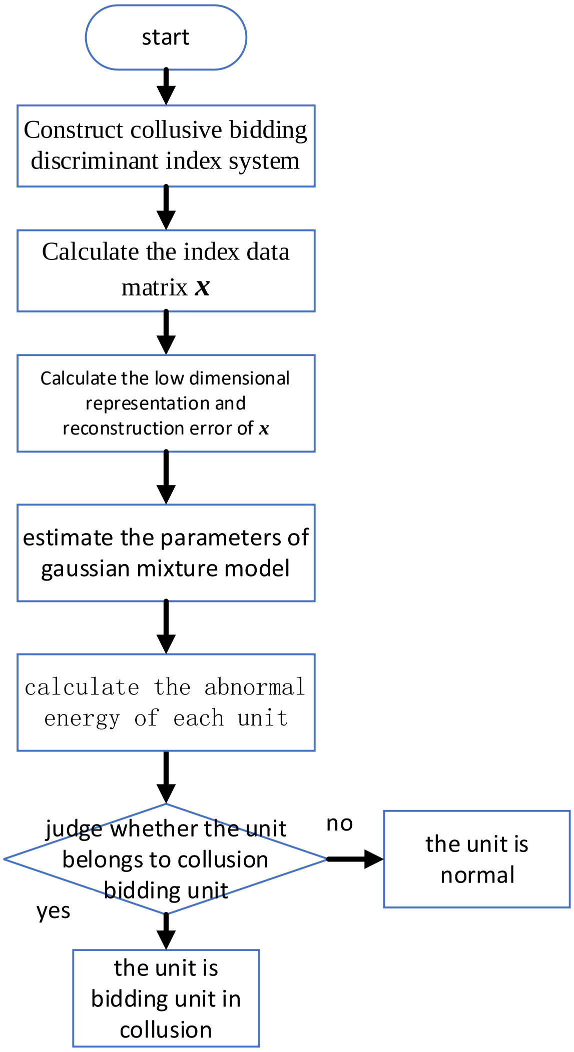

Figure 2 Collusive bidding discrimination steps based on deep-autoencoder Gaussian-mixture model

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 879-898
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

V. SAMPLE CALCULATION AND ANALYSIS

In order to verify the discriminant algorithm proposed in this paper, the three-day trading data of 300 units were selected as the research object based on the spot-trading data of a certain regional power market. The market clears every 15 minutes, and a three-day clearing cycle was selected as a research cycle. According to the unit declaration and transaction data, the indicators in the bidding indexing system of power generation companies were calculated, and in order to eliminate the influence of dimensional and numerical problems, the unit index data was processed using dimensionless min–max. NVIDIA MX350 was used as hardware, and Python 3.8 and TensorFlow 1.16 software were used. Tanh, Sigmoid and Relu were selected as codec functions. λ_1 and λ_2 were set as 0.1 and 0.01, respectively, in Equation (20), and K was set as 16. The parameter optimizer of the whole model was the commonly used Adam optimizer. The initial learning rate was 0.0001, and the number of iterations was 3000. Figure 3 shows the iteration effect; when the number of iterations reached 3000, the loss function nearly stops shrinking. Increasing the number of iterations does not significantly improve the accuracy of the algorithm but results in a huge waste in calculation power.

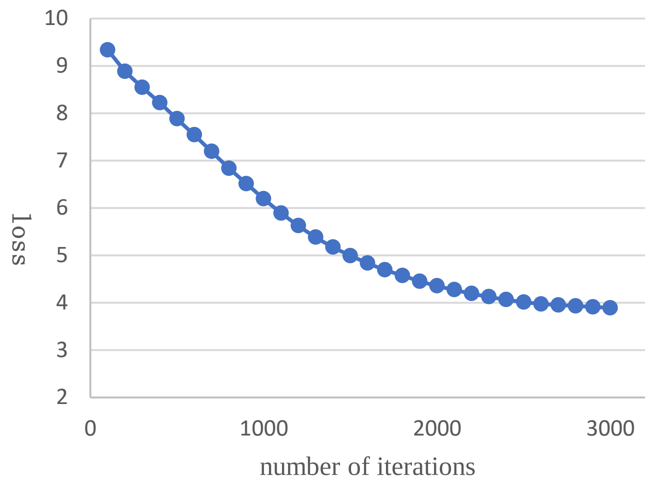

Figure 3 Effect of iteration time

The model was used to distinguish 300 units at a certain transaction time in the simulated regional power market, and the abnormal degree of each unit was calculated, that is, the abnormal energy of the generator unit. After standardizing the abnormal degree of units, the energy threshold ς was set as 0, as shown by the yellow line in Figure 4. As a result, 20 units were identified to be quite likely to collude in bidding behavior, and there is a large deviation from most units in terms of discriminating energy.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 879-898
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

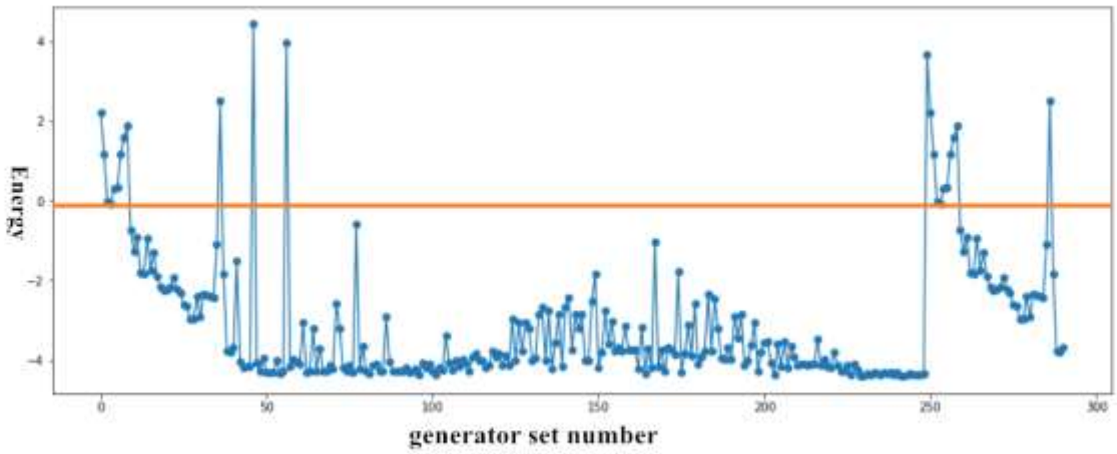

Figure 4 Individual display of unit abnormal degree

As shown in Fig. 4, the horizontal axis represents the unit number, and the vertical axis represents the corresponding abnormal degree of the unit. The figure shows a total of 20 units with abnormal energies over 0, namely those with collusive bidding.

Then, three indicators were chosen to observe the collusive bidding unit from the perspective of two indicators combined. To observe the abnormal distribution of the dots in the figure, the indicators for the average price, winning rate, high quote volume-price were used. The blue dot represents the normal units identified, and the red cross represents the 20 units identified as collusive bidding units.

As shown in Figure 5, the horizontal axis represents the average price of the units, and the vertical axis represents the winning rate of the units. In the figure, most of the units identified to participate in collusive bidding are concentrated in the lower-right side, indicating that these units may use ultra-high prices to raise the price of clearing in the market. At the same time, some units did not report ultra-high prices, but in the case of a certain high price, there is a considerable winning rate to obtain excess profits.

As shown in Figure 6, the horizontal axis represents the average price, and the vertical axis represents the high quote volume-price. The identified collusive bidding units are mainly concentrated in the upper-right corner; they are in the market for high average price, and at the same time, the high price of the unit price index is far greater than those of the other units in the market. Therefore, the units suspected of real-time conspiratorial bidding behavior raise the clearing prices, and the other units submit to the high unit to complete the bid to gain excess profits.

As shown in Figure 7, the horizontal axis represents the unit's high quote volume price index, and the vertical axis represents the unit's winning rate. The identified collusive bidding units are mainly concentrated in the lower-right side; their high quantity and price quotation index exceeded those of the other units in the market. Therefore, they play a role in the high price of the unit price index in the market under high-level conditions and bid-winning rates to obtain extra profits under collusive bidding.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 879-898
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

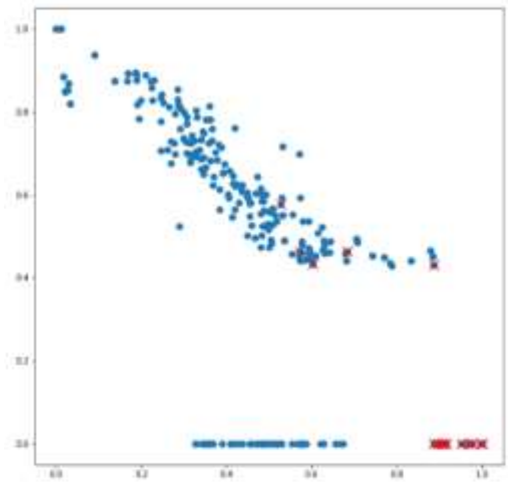

Figure 5. Average bid and winning rate

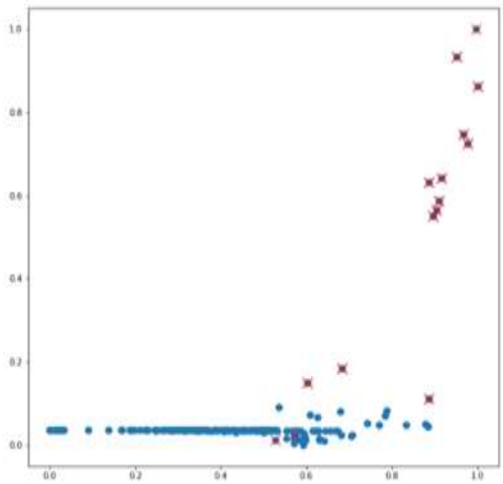

Figure 6. Average bid and High quote volume-price

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 879-898
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

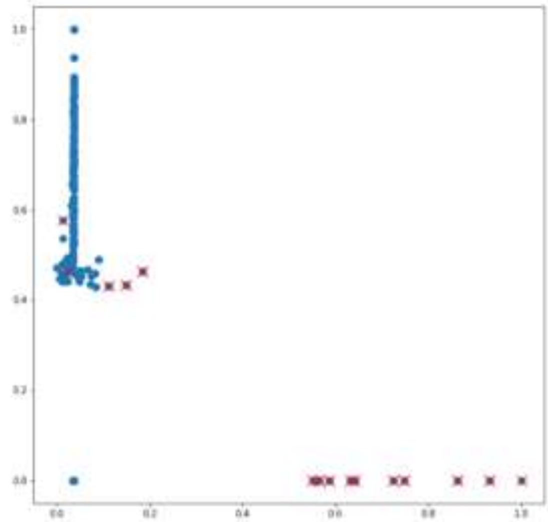

Figure 7. Winning rate and High quote volume-price

It can be clearly seen that the algorithm discriminates groups with different data characteristics from multiple perspectives, and these groups are in the low-density area of the data, which further indicates that the algorithm can simultaneously use data from multiple dimensions to discriminate collusive bidding behavior.

Considering the data scale in this paper, in order to evaluate the rationality of setting K as 16 and the influence of K in the calculation, experiments were conducted by changing K and within the same dataset. It can be observed from Figure 8 that the accuracy of discrimination increases with increasing K. When K exceeds 16, the discrimination performance does not change significantly, and the exponential increase of the calculation amount caused by the increase in K greatly affects the calculation efficiency. Therefore, it is reasonable to set K as 16 in this paper.

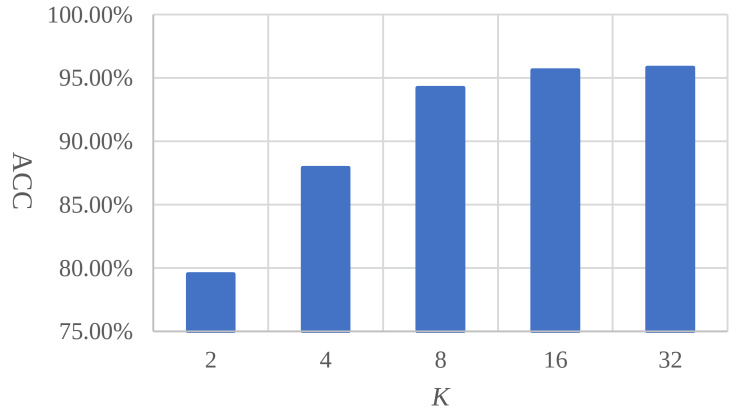

Figure 8. Influence of Gaussian mixture classification K on the results

In order to demonstrate the efficiency of the DAGMM algorithm, it was compared with other supervised

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 879-898
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

and unsupervised algorithms. Generally speaking, when assessing the performance of a classifier algorithm, researchers often consider the comprehensive performance of the classification, commonly using the accuracy rate, recall rate, and F1 score as criteria.

The first step is to present a confusion matrix, which helps to explain the meaning of these metrics. The confusion matrix can describe the classification of a classifier algorithm in detail; it provides a detailed description of whether the classification is correct. First of all, the collusive bidding problem is a classic dichotomy problem. Generally, for classical dichotomies, the confusion matrix is usually a square matrix of 2×2, as shown in TABLE II. Among the results, TP is the real example, which shows that the actual bidding units for collusion are predicted correctly; FP is a false positive example, which indicates that the unit is actually normal but is predicted incorrectly; FN is a false counter example, which indicates that the actual lack of bidding units is predicted incorrectly; and TN is a true counter example, which represents the actual normal unit and is correctly predicted by the classifier.

TABLE II. Dichotomous confusion matrix

REAL FORECASTING	**CONSPIRED BIDDING UNIT (FORECAST RESULTS)**	**THE NORMAL UNIT (FORECAST RESULTS)**
CONSPIRED BIDDING UNIT (TRUE)	TP	FN
NORMAL UNIT (TRUE)	FP	TN

The accuracy rate (ACC) refers to the proportion of the predicted results consistent with the actual situation in the samples of all units and can be calculated as follows:

$$ACC = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{21}$$

The accuracy rate (PRE) refers to the proportion of units that are predicted to be correct in collusive bidding:

$$PRE = \frac{TP}{(TP + FP)} \tag{22}$$

The recall rate (REC) denotes the proportion of units in collusive bidding that is predicted to be correct in the actual situation:

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 879-898
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

$$REC = \frac{TP}{(TP + FN)} \tag{23}$$

The F1 score refers to the comprehensive consideration of accuracy and recall rate. It employs the harmonic average algorithm, which can punish the limit cases of data and consider the balance between the two, as follows:

$$F1 = \frac{2 \times PRE \times REC}{PRE + REC} \tag{24}$$

In order to verify the advantages of the proposed method, it was compared with commonly used unsupervised and supervised algorithms. The unsupervised algorithm uses a 2D convolutional neural network operator, and the supervised algorithm uses a support vector machine.

TABLE III. Algorithm comparison

DETECTION ALGORITHM	CORRECT (%)	RECALL RATE (%)	F1 SCORE (%)
2D CONVOLUTIONAL NEURAL NETWORK OPERATOR	84.1	80.5	73.6
SUPPORT VECTOR MACHINE	85.7	79.3	84.2
GAUSSIAN MIXTURE MODEL DEPTH-AUTOENCODER	95.6	90.1	88.3

As can be seen from TABLE III, several classification algorithms have comparable recognition performance, and the algorithm used in this paper improves the accuracy and recall rate by at least 5% compared with other algorithms. In terms of F1 score, the algorithm used in this paper is significantly better than that of two similar unsupervised algorithms. Compared with the supervised algorithm, the improvement of the algorithm used in this paper is not obvious; however, in practice, it is relatively difficult to train a supervised learning model because there are very few samples that can be artificially marked for collusive bidding. In addition, if the generators in the test set collude to bid through the new bidding method, the supervised learning model cannot easily identify such units. Therefore, collusive bidding discrimination based on unsupervised learning is more practical and has more reference value.

VI. CONCLUSION

In order to improve the supervisory system of China's electricity market, a set of collusive-bidding discrimination methods for power producers in the electricity market was proposed, which was combined with a collusive bidding discrimination indexing system and unsupervised learning algorithm to assist experts in decision-making. The main conclusions are as follows:

1) A relatively ideal indicator system of collusive bidding discrimination among power producers was constructed to improve the early warning performance of the method.

2) An unsupervised learning model was selected because it is difficult to obtain indicators of collusive bidding behavior in sample data from the electricity market. Considering the high dimension and complexity of the indey dataset, a deep autoencoder network DAGMM was proposed to discriminate power producers' collusive bidding behavior in the power market.

3) Compared with other methods, DAGMM performs optimal dimensionality reduction in each iteration, maximizes the retention of the original data, and is more flexible in adjusting the data space, leading to higher accuracy in practical applications.

REFERENCE

[1] Xie J D, Huang X Y, Lu H Z,etc. Market power in electricity market risk prevention method research. Journal of price theory and practice, (12): 2020-53 + 49 162. DOI: 10.19851 / j.carol carroll nki cn11-1010 / f 2020.12.511.
[2] Wang W J, Fan H J. The game analysis of price collusion recognition and simulation. Journal of business research, 2010 (05):49-52. DOI:10.13902/j.carol carroll nkisyyj. 2010.05.020.
[3] Shi W J. Research on false data and detection scheme based on prediction in power system. Nanjing university of posts and telecommunications, 2019. DOI: 10.27251 /, dc nki. GNJDC. 2019.000906.
[4] Zeng D J, Yang L B, Li X G. The role and function of East China Power Grid in China's unified power market System. China Electric Power, 201,54(02):120-126.
[5] Liu D X. Research on long-term dynamic equilibrium decision model of power market based on system dynamics. North China Electric Power University (Beijing),2016.
[6] Sun J Q. Research on effective competition of power Market based on complex System. North China Electric Power University,2013.
[7] Zhang J. Hierarchical evaluation index system of market power in power market. Proceedings of the csee,2006,26(6):123-128
[8] Liu W M, Wu J J, Yang K. Power system technology,2007,31(S2):211-214.
[9] Lin J K, Ni Y X, Wu F L. Review on market power in power market. Power System Technology, 2002, 26(11): 70-76
[10] Zhao J G, Han X S, Cheng S J. Market force analysis considering power and speed constraints of generator sets in power market. Power Grid Technology, 2003, 27(11): 43-47
[11] Ding J W, Shen Y, Kang C Q, et al. Automation of Electric Power Systems,2003, 27(13): 24-29,67
[12] Liu D N, Zhang Q, Li X T, etc. Identification of Potential Hazard Behavior in Power Market based on Cloud Model and Fuzzy Petri Net. Automation of electric power systems,2019,43(2):38-50
[13] Wang J X, Yu Y, Sun X H. Fuzzy Analytic Hierarchy Process Based Reputation Evaluation for Generating Companies. Applied Mechanics and Materials, 2014, 556-562: 6660-6664
[14] Ren Z, Zhu R, Huang W. Power Automation Equipment, 2004, 24(10):7-11.
[15] Liu D N, Research and Application of Credit Management and Service System for power Market. Beijing, North China Electric Power University,2016-10-26.
[16] Zhang H S, Cao Z, Yang C H, etc. Research on Collusion Behavior Identification in power market based on

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 879-898
Article History: Received: 06 April 2022, Revised: 28 April 2022, Accepted: 04 May 2022, Publication: 15 May 2022

Adaboost-DT Algorithm. Electric power engineering technology,2020,39 (2) :152-158.
[17] Yin B C, Wang W T, Wang L C. Journal of Beijing university of technology,2015,41(01):48-59.
[18] Xie C, Wang B B, Zhao S N. Power system technology,2018,42(04):1170-1177. DOI:10.13335/j.1000-3673.pst.2017.1965.
[19] Liu J, Gao F, Luo X. review of deep reinforcement learning based on value function and strategy gradient. Chinese journal of computers,2019,42(06):1406-1438.