Research on the Prediction Model of Transformer Bidding

LI Ming1, LIU Yan-hao2, YUAN Yi-ping3, and ZHANG Shi-wen1

1School of mechanical engineering, Xinjiang University, No. 1230, Yan’an Road, Urumqi, China

ABSTRACT. Aiming at the problem of transformer manufacturing enterprises bidding is lacking scientific theoretical guidance and low bid probability, in order to predict the next bid price, based on principal component analysis (PCA) and artificial neural network (ANN) pre-tender estimate forecast model is proposed. The model uses PCA to preprocess the original high dimensional data, select principal components (PC) as the radial basis function (RBF) neural network’s input. PCA eliminates the correlation of the input variables, at the same time of simplifying the structure of ANN, improving the accuracy of the prediction model. The simulation results show the applicability of the pre-tender estimate forecast model.

1 Introduction

The electric power industry in China implement bidding system time is relatively short, which lacks power equipment bidding theory and bidding experience. At the same time, the bidding for electrical equipment involves the combined effects of factors such as raw material prices, product types, product allocation, tendering requirements, and government supervision. Therefore, it is difficult to establish a deterministic model to predict the fluctuation of the next bidding price, resulting in no basic data for reference for the bidding quotation of the power company, but more use of a small amount of experience reduces the probability of successful bids.

At this stage, the pre-tender estimate forecast have focused on using time series[1], fuzzy mathematics[2], gray fuzzy theory[3], game theory[4] to establish model. The forecast model of time series is determined by establishing a functional relationship between variables and time, including the moving average, one or multiple regression[5], autoregressive moving average (ARIMA) [6] and so on, but functional relationship between variables is not often deterministic linear; Fuzzy mathematics method, introduced the fuzzy set theory, fuzzy regression explain the deviation between estimated and actual bids of problems; Grey forecasting method is usually used by bid inviting party to forecast project cost, basic data demand of this method is high, general enterprise can’t get such data; Game model usually assumes that the bidding party rivals are known, and rivals in a state of relative static and isolation, the established model is usually an incomplete information static game model, which assumptions is demanding, and specific implementation is difficult[7].

ANN is a hots pot field of artificial intelligence research and application, which is widely used in all kinds of the field of prediction due to its good nonlinear approximation performance, such as coal production forecast [8], the city built area forecast [9], traffic generation forecast [10], etc. And have achieved satisfactory results. At present, applying ANN to predict the next bid price is very little, as a new method to explore, this paper applies RBF neural network in the field of transformer bid price forecast research. At the same time uses PCA for original multidimensional preprocessing of input variable, select the main component as the input of RBF. The PCA-RBF pre-tender estimate forecast model is constructed, and base on bidding data of a certain type transformer in the State Grid Corporation of China, an empirical study is carried out.

2 The key points of transformer pre-tender estimate forecast

2.1 Determine the pre-tender estimate forecast target

There are many kinds of transformer and voltage level, the same level of transformer will be very different because of the configuration. The classification of transformer is shown in Fig.1.

Fig.1 Transformer Product Classification
110kV voltage transformer is the most widely used transformer. SZ-50000/110kV transformer is the main type of 110kV voltage transformer, its data is representative, therefore, in this paper, we will use SZ-50000/110kV transformer bid price as the forecast target.

2.2 Analysis of factors affecting the pre-tender estimate price

Factors that affect the pre-tender estimate price intricate, depending on the type of influence, factors can be divided into temporal change factors and cost change factors. Temporal changes affected by time, including the raw material price fluctuations, inflation, etc. When the transformer type is determined, cost changes affected by product, including shipping, the number of products, product configuration, etc. Pre-tender estimate influence diagram as shown in Fig.2:

![Bid Price Influence Factors](image)

According to the State Grid Corporation of China’s bidding files and relevant announcement, there are 6 times per annum at present tender, plus a supplementary batch of tender, a total of 7 times. The batch reaction time factors, including fluctuations in raw material prices, inflation, etc. At the present stage, the bidding scope of the transformer is large and the geographical span is wide. Through the combination with the actual production data, in-depth analysis and research, select the following factors, which affect the pre-tender estimate price: batch, area, quantity, radiator configuration mode and delivery mode.

3 Construction of pre-tender estimate forecast model

The traditional neural network model has a lot of shortcomings, such as network learning convergence speed slow, easily trapped in local minimum and difficult to realize the global master. RBF neural network successfully overcome the above shortcomings, with global approximation properties and the best approximation performance, can better dig reveal the actual structure of the complex nonlinear system. But RBF neural network has strict requirements to the input, the input must be numeric, and the accuracy of data is necessary, otherwise the network won’t be able to work on. The PCA algorithm can reduce the dimension of data, simplify the representation of data, and solve problems which caused by the redundancy of experimental data, such as neural network structure complex, the performance of the network decreased, the computation time long, and the calculation precision low. In this paper, the basic data processed by PCA is used as the input of RBF network, and the PCA and RBF pre-tender estimate forecast model is constructed. The prediction model is shown in Fig.3.

![Price Prediction Model](image)

3.1 The pre-tender estimate principal component extraction

Suppose there are m transformer pre-tender estimate influencing factors and n samples, the original data can be expressed in $X_{nxm}$ matrix, namely:

$$X_{nxm} = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix}$$

(1)

$X_{nxm}$ is $X_{nxm}$’s center standardization.

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_{ij}}{s_{ij}}$$

(2)

For $i = 1,2,\ldots, n$; $j = 1,2,\ldots, m$; $\bar{x}_{ij}$, respectively $x_{ij}$’s mean and variance. Establish correlation matrix $R_{nxm}$, and calculating the eigenvalues and eigenvector, namely:

$$R_{nxm} = X_{nxm}^T X_{nxm} / (N - 1)$$

(3)

$X_{nxm}^T$ by $X_{nxm}$ matrix transpose. Calculate the eigenvalues of covariance matrix, and sort the eigenvalues, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$, the corresponding eigenvectors is $u_1, u_2, \cdots, u_m$, for $u_i$ is a $m$ vectors. The variance contribution rate and cumulative variance contribution rate, respectively:

$$\eta_i = 100% \lambda_i / \sum_i \lambda_i$$

(4)

$$\eta_p (p) = \sum_{i=1}^{p} \eta_i$$

(5)

According to principal component extraction method to extract $p$ principal components, principal component for the most part of $m$ is contained in the original variable information $U_{mxp} = [u_1, u_2, \cdots, u_p]$ by principal component corresponds to the feature vector is composed of order, is a main component constitute the main component of p matrix $Z_{nxp}$ as follows:

$$Z_{nxp} = X_{nxm}^* U_{mxp}$$

(6)

After new samples by standardization center, multiply by $U_{mxp}$ can be converted into a form of principal component.

3.2 The neural network model

RBF neural network hidden layer activation function using radial basis function, the radial basis function is Gauss function, the formula is:

$$R(x_p - c) = \exp(-\frac{1}{2\sigma^2}\|x_p - c\|^2)$$

(7)
Type: i-th European norm $\omega_i$ is the center of the Gauss function; $\sigma_i$ is the variance of Gauss function $x_p$ as input vector. At this point, the output of the RBF network model is the output of hidden layer neurons linear weighted summation, the formula is:

$$ y = \sum_{i=1}^{h} \omega_i \exp(-\frac{1}{2\sigma^2_i}||x_p - c_i||^2) \quad (8) $$

Type: $c_i$ as the center of the hidden layer nodes; The hidden layer of $h$ nodes, $\omega_i$, for the hidden layer to output layer connection weights, $i=1, 2, \ldots, h$. Specific, can based on the k-means clustering method calculating basis function center $c_i$, thus solving the variance $\sigma$, when the base function of RBF neural network for Gauss function, the variance of basis function are as follows:

$$ \sigma_i = \frac{\omega_{max}}{2^h} \quad (9) $$

Type: $i=1, 2, \ldots, h$; $\omega_{max}$ is the center of the selected maximum distance. To solve the hidden layer and output layer connection weights between neurons, can be calculated by least square method, the formula is:

$$ \omega = \exp\left(\frac{h}{c_{max}}||x_p - c_i||^2\right) \quad (10) $$

Type: $i=1, 2, \ldots, h$. RBF neural network with input and output error to adjust the parameter weights of $c_i$ and $\omega_i$, then adjust network internal factor, through iterative calculation until the network output of the mean square error reaches the precision requirement, network termination are calculated and output forecast.

4 To verify

The following samples data are SZ-50000/110kV transformer bidding history data, selected from the State Grid Corporation of China from 2015 to 2016, total of 57 samples.

4.1 Basic data processing

The raw data table as shown in Tab.1.

<table>
<thead>
<tr>
<th>Tender package</th>
<th>Demand unit</th>
<th>Engineering voltage grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Anhui Electric Power Corporation</td>
<td>110kV</td>
</tr>
</tbody>
</table>

As the model has strict requirements to the input, input must be numeric, but the basic factors are text data, so it is necessary to perform quantitative data operation. First of all, regional factors passed through by the comparison method to quantify. There are different kinds of configuration and delivery mode transformer in one package, here, based on the numbers of transformer to realize configuration factor and the delivery mode quantity. After treatment operation, constitute a total of seven factors parameters, the parameters of the table as shown in Tab.2.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Batch</th>
<th>region</th>
<th>Quantit y</th>
<th>Radiator arrangement (a body) quantity</th>
<th>Radiator arrangement (horizontal split type)quantity</th>
<th>Delivery (the floors of delivery) quantity</th>
<th>Delivery (delivery equipment foundation) quantity</th>
<th>The base price/ten thousand</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>12</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>388.382</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>12</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>390.123582</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>57</td>
<td>8</td>
<td>10</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>320.5139</td>
</tr>
</tbody>
</table>

Among them: Batches take 1-8 instead of the first batch of 2015 to the first batch of 2016; Area through quantitative into digital instead; Number for the package SZ-500000/110 kV transformer model of total quantity; Radiator arrangement (a body) quantity for radiator one-piece arrangement of the quantities in the package; Radiator arrangement (horizontal split type) quantity for radiator horizontal split type quantities in the package; Delivery (the floors of delivery) quantity for the floors of delivery quantity in the package; Delivery (delivery equipment foundation) quantity for the equipment foundation delivery quantity in the package.

The key operation of data preprocessing is mainly for data normalization, which remove the dimension between different parameter data. According to the formula (1) (2), input is $X_{57x7}$, $X_{57x7}$ is calculated.

4.2 Principal component extraction

There are three extraction methods. First need to define a collection must achieve the cumulative percentage of maximum variance, the ceiling depends on the actual processing problems, can be in the range of 75% ~ 95%, below the upper limit of the cumulative percentage of variance of principal components is to extract the principal component. The second method is to check the variance of gravel figure relative to the number of components, and choose the graphical approximation or eigenvalues greater than 1. The third way is by cross
validation, select small PC number of validation errors. In this paper, the three methods are used to determine the number of PC.

According to the formula (3) calculate the correlation matrix \( R_{7 \times 7} \), and According to (4), (5) calculate the eigenvalues of the matrix \( R_{7 \times 7} \). The total variance explained table and component characteristic value figure, as shown in Tab.3 and Fig.4.

### Table 3. Total Variance Explained Table

<table>
<thead>
<tr>
<th>Component</th>
<th>Initial Eigenvalues</th>
<th>Extraction Sums of Squared Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>% of Variance</td>
</tr>
<tr>
<td>2</td>
<td>1.658</td>
<td>23.680</td>
</tr>
<tr>
<td>3</td>
<td>1.094</td>
<td>15.634</td>
</tr>
<tr>
<td>4</td>
<td>0.870</td>
<td>12.428</td>
</tr>
<tr>
<td>5</td>
<td>0.560</td>
<td>7.998</td>
</tr>
<tr>
<td>6</td>
<td>0.050</td>
<td>0.714</td>
</tr>
<tr>
<td>7</td>
<td>0.011</td>
<td>0.158</td>
</tr>
</tbody>
</table>

Extraction Method: Principal Component Analysis.

First observe the eigenvalue and cumulative variance explain tables and component characteristic value figure, PC number can be determined in accordance with the former two methods for three or four, then use the third way to cross validation, finally found that when take 3 PC for verification of the minimum error, then here in 3 PC for subsequent operations, so in the formula (6) \( p=3 \).

According to the formula (6) can get principal component score values (as shown in Tab.4) as the input into the neural network prediction model.

### Table 4. Principal Component Score

<table>
<thead>
<tr>
<th>Sample</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1.39296</td>
<td>.29020</td>
<td>.74682</td>
</tr>
<tr>
<td>2</td>
<td>-1.39296</td>
<td>.29020</td>
<td>.74682</td>
</tr>
<tr>
<td>3</td>
<td>-1.61047</td>
<td>-.54434</td>
<td>-.83766</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>54</td>
<td>.70037</td>
<td>-.214758</td>
<td>.33687</td>
</tr>
<tr>
<td>55</td>
<td>.59718</td>
<td>.73320</td>
<td>-.03214</td>
</tr>
<tr>
<td>56</td>
<td>.64877</td>
<td>-.70719</td>
<td>.15236</td>
</tr>
<tr>
<td>57</td>
<td>-.90243</td>
<td>.10221</td>
<td>1.40570</td>
</tr>
</tbody>
</table>

**4.3 Neural network training and prediction**

RBF neural network model is established by MATLAB. Base function by Gauss function, form as shown in the formula (7), hidden layer to output layer weighted formula as shown in (8), variance and connection weights to determine as shown in formula (9), (10); according to the actual situation and training experience, Take 3 extension constant sc, and initialize the epsilon learning precision is 0.01, hidden layer nodes of root mean square error principle of minimizing determined according to the experiment many times. Eventually determine the PCA-RBF network structure for 3-46-1.

Take the first 52 samples influence factors and principal components as input of the PCA-RBF pre-tender estimate forecast model, the pre-tender estimate price as output, training model. And predicted respectively 53, 54, 55, 56, 57 samples of the pre-tender price, using samples from 53, 54, 55, 56, 57 of the actual bid price compared with that predicted pre-tender price to verify.

### 4.4 Results contrast

In order to further verify the ability of model in transformer pre-tender estimate forecast, this paper also use the pure RBF network model to forecast. Different kinds of forecasting models of experimental results see the figure below.
The results show in Tab.5. From the table, we can see, in the sample of 53's forecast, the PCA-RBF pre-tender estimate model prediction error is 2.30%, better than pure RBF model; In the sample of 54 and 57's forecast, PCA-RBF pre-tender estimate forecast model prediction accuracy obviously higher than that of RBF pre-tender estimate prediction model, error is 0.45%; In sample 55, 56, PCA-RBF pre-tender estimate forecast model and RBF pre-tender estimate prediction model has the same precision, the error is 0.53%. Through comprehensive comparison, we can see that PCA-RBF pre-tender estimate model achieved good prediction effect, can provide certain data support to bid for the enterprise.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (51365054); The Science and Technology Plan Projects in Xinjiang Uygur Autonomous Region (201512105); Urumqi City Science and Technology Plan Projects (P141010003).

5 Conclusion

Through to the State Grid Corporation of China SZ-500000/110kV transformer research, found that, the PCA-RFB pre-tender estimate forecast model established in this paper achieved good prediction effect. Believe that with the deepening of the research and application, as a kind of advanced artificial intelligence algorithm and forecast new method, the PCA-RBF pre-tender estimate forecast model provides scientific data support for transformer manufacturers.

Some existing prediction methods are mainly concentrated in terms of deterministic prediction, but only provide deterministic prediction results are often difficult to meet pre-tender estimate forecast. How to further analyze the uncertainty of forecasting pre-tender estimate on the basis of the deterministic forecasting results is the direction of future research.

References