A Prediction Method for Power Transformer State Parameters Based on Grid Long Short-Term Memory Network

Jiejie Dai¹, Hui Song², Xiaoqi Wan³, Yingjie Yan³, Gehao Sheng² and Xiuchen Jiang²

¹Shanghai Shinan Electric Power Supply Company, Xinbei Rd.268, Shanghai, China
²Shanghai Jiao Tong University, Dongchuan Rd.800, Shanghai, China
songeos@163.com

Abstract. Power transformer state parameter prediction analysis can provide strongly technical support for equipment state assessment. The available transformer state parameter prediction models are mainly based on the very limited number of state parameters for analysis and judgment. So, the stability and the intelligence of prediction still need improvement. Based on the large amount of information of transformer equipment state, the data of environmental meteorological and grid operation, this paper proposes a transformer state parameter prediction method using deep mining of complex associations with the grid long short-term memory network (GLSTM). Association relationships captured from GLSTM are used to correct the prediction of state parameters. Finally, the method is applied to the top oil temperature trend prediction of a 500 kV transformer. The results show that the proposed method can mine and analyze the relationship between the influencing factors of equipment state. Compared with the prediction methods without considering the correlation and the traditional methods, the association relation extracted by GLSTM improves the stability of the prediction model and reduces the prediction error.

1. Introduction

Power transformer, a key equipment in power system, decides the stability of electricity. Its status is closely related to the information of its state under the high-voltage electric, thermal and mechanical stress [1]. Therefore, the prediction of the transformer state parameters can understand the development trend of transformer operation in advance, so that various precautionary measures can be taken in advance to avoid tremendous electric accident.

Nowadays, many scholars have made some achievement on the prediction of transformer state trend based on the top oil temperature, the winding temperature, the winding vibration signal, and the concentration of gas dissolved in transformer oil [2-4], etc. However, due to the very limited number of the state parameters used in the evaluation of existing prediction models, it is difficult to explore the inherent laws and relations of information parameters. The scientific rationality of the equipment’s state evaluation with single parameters remains to be studied.

Considering the correlation between parameters, researchers have introduced gray correlation, mutual information and Apriori algorithm to do analysis [5-7]. In recent years, with the introduction of layer-by-layer pre-training and reverse tuning training methods, the learning efficiency and the data analysis capabilities of deep network have substantial improvement. Many scholars have achieved...
remarkable results by applying Long short-term memory (LSTM) networks, which have "memory" advantage, to the modeling and prediction of big data sequence information [8]. The grid long short-term memory (GLSTM) network [9], a multi-dimensional space network generated by LSTM units, can be used to process vector, sequence or higher dimensional data, enabling extraction of information in various dimensions including time and depth [10].

In order to improve the prediction stability and intelligence, we generated a prediction method of transformer state parameters regarding association relation, analyzed the structural characteristics of the grid long short-term memory (GLSTM) network, and created a prediction model based on that network. Driven by the data, the GLSTM network can be used to automatically extract the association relation of the characteristic parameters. It contributes to reduce the subjective influence of the human intervention, and to explore the trend connection between the characteristics and the future status. The results of case studies have shown that the proposed method significantly improved the accuracy of power transformer state parameter prediction.

2. GLSTM network
The concept of GLSTM recurrent neural network was introduced by N Kalchbrenner, I Danihelka and A Graves [10] in 2015. The GLSTM models arrange LSTM blocks into multidimensional grids, so that each grid contains a set of LSTM blocks in multiple dimensions, including the depth dimension. Reducing the problem of gradient disappearance, this architecture introduces a gated linear dependency of adjacent cell states in each dimensional space.

The GLSTM module consists of two dimensions, time length and depth dimensions, represented by Time-LSTM and Depth-LSTM [9]. The computations in each grid are defined as follows.

\[
x_{ij} = [h_{i,t,l}^O, h_{i,t,l}^F] \\
(h_{i,t,l}^O, c_{i,t,l}^O) = \text{Time\_LSTM}(x_{ij}, c_{i,t,l}^O, \Theta^T) \\
(h_{i,t,l}^D, c_{i,t,l}^D) = \text{Depth\_LSTM}(x_{ij}, c_{i,t,l}^D, \Theta^D)
\]

where \(x_{ij}\) is the input of the GLSTM at time \(t\) and layer \(l\). Equation (2) and (3) represent the calculation results of the time LSTM module and the depth LSTM module, respectively. \(c_{i,t,l}^O\) and \(h_{i,t}^O\) (\(j\) represents \(T\) or \(D\)) are cell state and cell output respectively at time \(t\) and layer \(l\) of \(i\)-LSTM \((i\) represents \(T\) or \(D\)), while \(\Theta^i\) denotes all the parameters of \(i\)-LSTM. The detailed calculations in the LSTM in each grid are available in [11].

3. Transformer state parameter prediction model based on GLSTM network
3.1. Related sequence mathematical representation
The single variable sequence of parameter can be represented as \(X^m = \{x_{m,1}, x_{m,2}, \ldots, x_{m,t}, \ldots\}\), where \(x_{m,t}\) is the measured value of the \(m\)-th parameter \(X^m\) at time \(t\). All parameters, such as transformer state, environmental meteorology and grid operation data, are arranged in chronological order to form a time series matrix as shown in equation (4).

\[
X = \begin{bmatrix}
X^1 \\
X^2 \\
\vdots \\
X^r
\end{bmatrix} = \begin{bmatrix}
x_{1,1}, x_{1,2}, \ldots, x_{1,t}, \ldots, x_{1,r} \\
x_{2,1}, x_{2,2}, \ldots, x_{2,t}, \ldots, x_{2,r} \\
\vdots \\
x_{r,1}, x_{r,2}, \ldots, x_{r,t}, \ldots, x_{r,r}
\end{bmatrix}
\]

where \(r\) represents the number of related parameters. Therefore, the transformer operating state parameter prediction can be regarded as a multivariate multi-dimensional prediction problem.
3.2. Prediction method based on correlation mining using GLSTM network

The input characteristic vector includes the following parameters: the on-line monitoring state parameters of the transformer, the grid operating state parameters and the environmental meteorological data of the substation. The correlation of the time dimension of the parameter is determined by the feature extraction of the statistical relationship of sequence data through the GLSTM network. Also, the abstract related features can be extracted from depth-dimensional complex input data. The factors, which have more substantial influence, can be activated to suppress and weaken irrelevant and redundant information. At last, further prediction of the trend of the state parameters will be accomplished through the feed-forward neural network with the parameters obtained by GLSTM mining. Figure 1 shows the transformer state parameter trend prediction model based on GLSTM network. The specific application steps are as follows.

1) Normalize the K-order measurement value with strong time relationship by using the dispersion normalization method to obtain the data representation X.

2) Extracts prediction parameters by GLSTM trained with back propagation through time. \( h_{t-k} \) is a hidden layer expression, including \( h_{t-k}^H \) and \( h_{t-k}^D \). The hidden state of upper layer is the input of the next layer network that forms a predictive feature deep mining structure.

3) Adjust the output layer weight of the feed-forward neural network by treating the correlation between the features in step 2) as the prior knowledge. The feed-forward neural network outputs trend results as an inference prediction layer.

4) Regarding the prediction parameters obtained from the training data, test the state parameters in the future time to verify the prediction accuracy.

Figure 1. Transformer state parameter prediction model based on GLSTM network.
4. Case studies and analysis

The study demonstrated the prediction performance of GLSTM model by monitoring data in a 500 kV power transformer at electric power company in China from March 21, 2010 to December, 2013. Since the current monitoring state of the transformer is not very different from that before a few moments, long-term predictions are more meaningful than short-term one in actual guidance. Therefore, our study focused on the long-term predicting effect during the establishment and verification of the model.

Monitoring data from March 21, 2010 to June 28, 2013 was used as the training samples, and monitoring data from June 29, 2013 was used as the test samples. To evaluate the training performance of the GLSTM model, we use the mean absolute error, $\delta$, and max absolute error, $\max \delta$, to measure the error of the training data.

$$\delta = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{\hat{x}_t - x_t}{x_t} \right| \times 100\%$$  \hspace{1cm} (5)

$$\max \delta = \max \left| \frac{\hat{x}_t - x_t}{x_t} \right|_{(t = 1, 2, \cdots, N)}$$  \hspace{1cm} (6)

where $N$ is the number of test samples. $\hat{x}_t$ is the prediction value and $x_t$ is the measured value. The smaller the error is, the better the prediction result is.

4.1. Model parameters

In order to determine the relation between historical data and the parameters waiting for prediction within 60 days, we did multiple experiments to predict the average absolute error of the top oil temperature under different K-order and hidden layer structure. Figure 2 shows the result.

![Figure 2](image_url)

**Figure 2.** The relationship of prediction error, hidden layer structure and related time order K.

As the order K increased, the average absolute error of the prediction results decreased. After K increases to 33, the error remains essentially constant. Therefore, in order to predict the top oil temperature in the next 60 days, it was necessary to provide the historical data at least one month before. We observed that the hidden layer structure in the network increased from 1 to 4, the correlation order K increased from 1 to 4, and the declining trend of the corresponding error was not obvious. Thereafter, as the value of K increased to around 33, the error was remarkably reduced. As the hidden layer increased from 1 to 3, the error decreased, and the stability error of hidden layers 3 and 4 was almost flat. Based on the prediction accuracy and computation efficiency, Table 1 gave the optimal network model, in which the number of neurons in each layer was determined by using a trial and error method.

| K-order | Hidden layer | Network structure |
|---------|--------------|-------------------|
| 33      | 3            | 297-1200-750-200-60 |
4.2. Association analysis

The single-parameter prediction model without consideration of the association relation can be constructed by LSTM network. The comparison of the GLSTM prediction model reflecting the association is shown in figure 3.

Figure 3 shows that the introduction of association relation between the state parameters obtained by data mining with the GLSTM network can significantly improve the prediction accuracy. Compared with the single sequence prediction method, our method reduced the maximum prediction error from 15% to less than 10%. By extracting the mutual coupling relationship between variables, the GLSTM network was better at tracking the trend of data changes, and the prediction result was more robust.

4.3. Compared with traditional prediction models

In order to verify the prediction of the proposed model, we compared the GLSTM prediction model with the commonly used traditional prediction models, including autoregressive (AR) model, radial basis function neural network (RBFNN), and support vector machine regressive (SVR) model and the multi-parameter grey model (GM).

Calculation shows that the grey correlation degree among the concentration of acetylene dissolved in oil, the average temperature and the load current and the top oil temperature, is larger than 0.5, reflecting a strong correlation. This fact validates that the prediction model is in the multi-parameter GM (1, 4) form. The results of the prediction in the next 60 days by the traditional model and the GLSTM model are shown in Table 2.

Table 2. Prediction errors of different models (in percentage)

| Model | Prediction Error |
|-------|------------------|
| GLSTM | 9.87±2.21        |
| AR    | >50              |
| RBFNN | 28.92±8.39       |
| SVR   | 24.72±7.56       |
| GM    | 16.63±3.44       |

It can be seen from Table 2 that the prediction error of the model with the parameter association relation is significantly lower than that of the model without the correlation relationship, and same for the error fluctuation range. The GLSTM model has more comprehensive coverage information, and more complete extracted correlation features, than the GM model. Therefore the error is relatively small, and there is no subjective threshold judgment in the prediction process. Since the operating state of the transformer is influenced by various conditions such as meteorological conditions and load conditions, using a large amount of historical data for analysis, mining data characteristics and predicting the variation trend of the parameters, can effectively reduce the prediction error.
5. Conclusion

Regarding the problems of stability and low accuracy of the prediction for power transformer state trend, this paper proposes a prediction method for extracting sequence association relation by GLSTM network, and makes conclusions as below:

1) The GLSTM network amply extracts, both in time and depth dimensions, the inherent laws and associations among equipment state information, grid operation and environmental meteorological data. GLSTM extracts the feature information as prior knowledge to the neural network weights, adjusting and correcting the single parameter prediction error.

2) The prediction of the top oil temperature in the analysis of the case shows that the correlation obtained by data mining with the GLSTM network can significantly improve the prediction accuracy. Compared with the single sequence prediction method, the maximum prediction error decreases to less than 10%. Moreover, compared with the traditional AR, SVR, RBFNN and multi-parameter GM models, the GLSTM prediction model error is greatly reduced. Our method can also result a smaller error fluctuation range, and a more stable prediction result.

Acknowledgments
This work was supported in part by The National Natural Science Foundation of China (51477100) and The Science and Technology Program of State Grid Corporation of China (GBY17201700204 Fault Diagnosis and Maintenance of Electrical Equipment based on Big Data Technology).

References
[1] J. J. Kelly and D. P. Myers, “Transformer life extension through proper reinhibiting and preservation of the oil insulation”, IEEE Trans. Ind. Appl., vol. 31, no. 1, pp. 56–60, 1995
[2] R. J. Liao, J. P. Bian, L. J. Yang, S. Grzybowski, Y. Y. Wang and J. Li, “Forecasting dissolved gases content in power transformer oil based on weakening buffer operator and least square support vector machine–Markov”, IET Generation, Transmission & Distribution., Vol. 6, No.2, pp. 142-151, 2012.
[3] J. F. Qin, L. L. Li, C. Zhou, “A GD-SVM Model for Calculating Oil-Immersed Transformer Hot Spot Temperature”, IOP Conference Series Earth and Environmental Science, vol. 223, pp. 012031, 2019.
[4] L. Wang, Z. J. Zhang, H. Long, J. Xu, R. H. Liu, “Wind Turbine Gearbox Failure Identification With Deep Neural Networks”, IEEE Transactions on Industrial Informatics, vol. 13, no. 3, pp. 56–60, June 2017.
[5] R. R. Zheng, J. Y. Zhao and T. T. Zhao, “Prediction of power transformer oil dissolved gas concentration based on modified gray model”, in Electrical and Control Engineering, 2010 International Conference on, 2010, pp. 1499-1502.
[6] J. Huang, X. Yan, “Quality relevant and independent two block monitoring based on mutual information and KPCA”, IEEE Transactions on Industrial Electronics, vol. 64, no. 8, pp. 6518–6527, March 2017.
[7] G. H. Sheng, H. J. Hou, X. C. Jiang, Y. F. Chen, “A Novel Association Rule Mining Method of Big Data for Power Transformers State Parameters Based on Probabilistic Graph Model”, IEEE Transactions on Smart Grid, vol. 9, no. 2, pp. 695–702, 2018.
[8] S. Hochreiter and J. Schmidhuber, “Long short-term memory”, Neural Comput, Vol. 9, No.8, pp. 1735-1780, 1997.
[9] W. N. Hsu, Y. Zhang, J. Glass, “A prioritized grid long short-term memory RNN for speech recognition” IEEE Spoken Language Technology Workshop, pp. 467-473, 2016.
[10] N. Kalchbrenner, I. Danihelka, A. Graves, “Grid long short-term memory” arXiv preprint arXiv:1507.01526, 2015.
[11] W. Kong, Z.Y. Dong, Y. Jia, D.J. Hill, Y. Xu, Y. Zhang, “Short-term residential load forecasting based on lstm recurrent neural network”, IEEE Trans. Smart Grid, Vol. 10, No.1, pp. 841-851, 2019.