2020 IEEE IAS Industrial and Commercial Power System Asia Technical Conference

Prediction Method for Transformer State Based on GRU Network

Xiaomeng Su, Guan Wang,Qingquan Li
Department of Electrical Engineering
Shandong University
Jinan,China
mjymx.su@gmail.com

***Abstract*—Dissolved gases content in power transformer oil can provide important information for transformer condition assessment. The recurrent neural network (RNN)has always performed well in the data labeled in the form of sequence but it has long term dependency and vanishing gradient problem. Gated Recurrent Unit (GRU) is an improved version of RNN which deal with problems in RNN. This paper proposed a new forecasting model based on GRU network. The studies show that the model can effectively predict the change state of oil gas in power transformer. Compared with the Long Short Term Memory (LSTM) networks which is also improved recurrent neural network, the accuracy rate is comparable, and the training time is shorter and the universality is better. When the interval and fluctuation of data are greater, the prediction of the GRU network is also more stable.**

Keywords—power transformer, GRU neural network, state prediction

I. INTRODUCTION

The power transformer is the core equipment of the power grid system, its operating state is related to the stability, safety and economy of the entire power system. The content and change law of dissolved gas in oil-immersed transformer are closely related to the fault of power transformer. A large number of cases have proved that Dissolved Gas Analysis (DGA) can effectively diagnose the potential failure in the power transformer [1-5]. Therefore, process and analyze the oil chromatography data can grasp the health status of the power transformer.

The standard [6] specifies the threshold value of the gas concentration of the transformer fault to ensure the normal operation of the transformer. However, with the deepening of scholars' understanding of the fault mechanism of transformers and continuous practical exploration, researchers have found that sometimes power transformers fail when the gas concentration has not reached the threshold. At the same time, frequent on-line detection cycles will also cause excessive loss of monitoring devices [7], and it is not economical if transformers in normal condition.

Predicting gas change trends can solve the above problems well. In recent years, relevant scholars have proposed many prediction models for dissolved gases in transformer oil. Literature [8] proposed a grey model (GM) prediction method for gas in transformer oil. Reference [9] applies support vector machines (SVM) to the state prediction of transformers. The relationship between time and the volume fraction of dissolved gas in power transformer oil is a typical multi-dimensional time series. Traditional methods only consider the development trend of some kind of gas in the modeling and ignore the correlation analysis between the gases. The prediction is not scientific and poor stability. Many combined forecasting methods [10] avoid the defects of single component gas concentration prediction, But it increases the computational complexity of the data.

Recurrent Neural Network (RNN) is a kind of neural network with special structure. Unlike BP neural networks, RNN can not only receive information transmitted by other neurons, but also their own state with time. Therefore, this network model has a memory function and is good at dealing with long-term sequence problems. With the development of artificial intelligence in recent years, RNN has made many achievements in the modeling and prediction of sequence information [11].

Because of its own structure. During data training, the RNN network will encounter the problems of gradient disappears and gradient explosion [12]. So it cannot train models with long delays. Schmidhuber proposed the long short term memory (LSTM) [13]. Added a gate control structure on the basis of RNN to solve the problem of gradient explosion and gradient disappears. In 2014, Cho et al. Made further improvements to simplify the gate control structure of LSTM, and proposed Gated Recurrent Unit (GRU) [14]. Its prediction accuracy is much higher than RNN network. And its performance is comparable to LSTM, because the structure of the neuron is simpler, guarantees it has a faster training speed and is more suitable for dynamic process modeling.

The GRU network structure is more novel and has better performance. The prediction model established by the GRU network has high accuracy in predicting the state of the transformer and the training time is shorter. Based on the characteristic gas data of a power transformer, this paper proposed a state prediction model of power transformer based on GRU network, analyze the effect of GRU network to predict transformer status.

II. GRU NEURAL NETWORK

The GRU network improves the structure of the network on the basis of RNN, it adds a gating mechanism to control the

978-1-7281-4303-3/20/$31.00 ©2020 IEEE

transmission of information in the neural network. The gating mechanism can choose how much information in the memory unit needs to be retained or discarded, and what new state information needs to be saved in the memory unit.

This allows the gate-controlled recurrent neural network to learn the dependencies with a relatively long time span without the problems of gradient disappearance and gradient explosion. In RNN networks, there is a non-linear relationship between the state of the network and the parameters are shared at each time step, which is the root cause of gradient explosion and gradient disappearance. gated neural network add a linear dependency between states, to avoid the problem of gradient disappearance and gradient explosion.

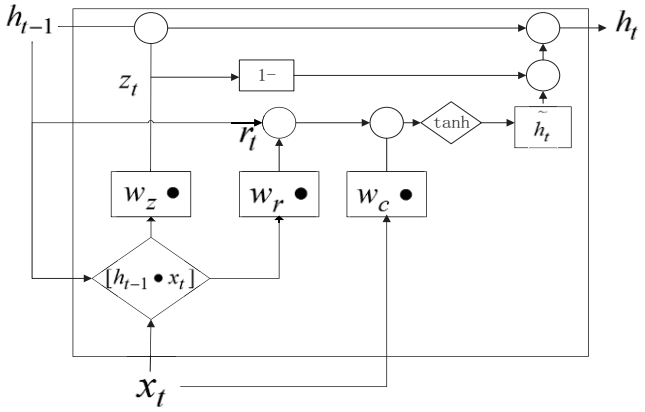

Fig. 1. The illustration of GRU network.

Compared with LSTM network, the GRU network structure is simpler. It combines the input gate and the forget gate which in the LSTM and called update gate. In the GRU network, there is no division between the internal state and the external state of the traditional network structure, but by adding a linear dependency between the current network state and the previous moment network state, to solve the gradient disappearance and gradient explosion. Figure 1 is the structure diagram of GRU. Its function is to memorize and transfer data information, like the cell state in LSTM. For example, in transformer oil gas prediction, it can save important information such as the prominent state of individual gases and the correlation of multiple gases.

In the GRU network, the update gate is used to control how many historical states to keep in the current output state and how many candidate states to keep at the current time. The calculation formula of the update gate is shown in the formula (1).

$$z_t = \sigma(W_z h_{t-1} + W_z x_t + b_z) \tag{1}$$

The output of the update gate is multiplied with the historical state h_{t-1} and the candidate state h_t, h_t multiplied by $1\text{-}z_t$. The output of the network at the current moment is an Equation (2).

$$h_t = z_t \otimes h_{t-1} + (1 - z_t) \otimes \tilde{h}_t \tag{2}$$

The role of the reset gate is to determine whether the candidate state at the current moment depends on the network state at the previous moment and how much it depends on. It can be seen from the figure that the network state at the previous time is first multiplied by the output of the reset gate and then used as a parameter to calculate the candidate state at the current time. The calculation formula of the reset gate is (3).

$$r_t = \sigma(W_r h_{t-1} + W_r x_t + b_r) \tag{3}$$

The value of r_t determines the degree of dependence of the candidate state h_t on the state h_{t-1} at the previous moment. The calculation formula is shown in equation (4).

$$\tilde{h}_t = \tanh[W_c x_t + W_c (r_t \otimes h_{t-1}) + b_c] \tag{4}$$

III. The Realization of Gas Content Prediction in Oil

The entire state prediction process is divided into three stages. In the first stage, the data is preprocessed, and the data is processed into a sequence that can be used for model training. In the second stage, the processed oil chromatography time series data are sent to the GRU model for training to generate a prediction model. At the last stage, predict the date of the validation set, and evaluated the prediction results through various aspects. Prediction process is shown in figure 2.

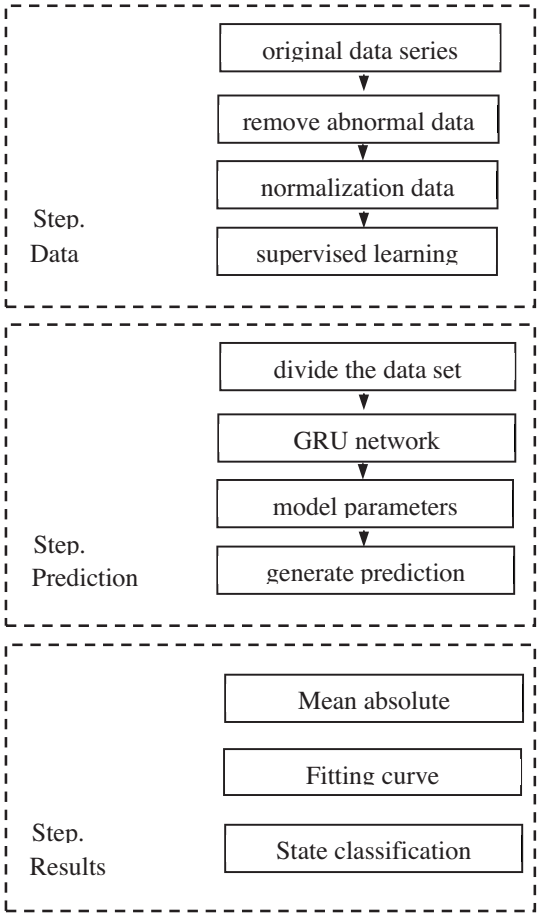

Fig. 2. Training flow chart of GRU.

A. Data Preprocessing

First, eliminate data which unrelated to gas changes in the data set. This step can simplify the entire training process. The load pressure values in the data used in this article all show normal, so deleting such series will not affect the overall forecast. In the original data set, there will be a certain number of abnormal data values, such as continuously zero values, continuously unchanged values, etc., which will affect the establishment of the model and the final prediction results. Through the program settings, abnormal data can be automatically corrected and selected to improve the stability of the model prediction.

The difference in gas content of various data is obvious, such as the difference between the gas content of hydrogen and carbon dioxide. When training the recursive network directly, the sequence with the largest value will affects the proportion of the entire prediction. In order to reduce the influence of the difference in gas volume fractions on the overall prediction in oil chromatography, we adopt the deviation standardization to normalize the input data.

Convert data format to supervised learning format. In this paper, we set the input lag observed value and the output observation value to 3 and 1, and use all the characteristic data at time T-3, T-2, T-1 to predict the value of time T.

B. Prediction Model

Split the data set into training and verification sets. The ratio of the data set in this article is set to 0.67, Training set: validation set is 2:1. The learning of the neural network indicates the current state through a certain index, and uses this index as a benchmark to find the optimal weight parameter. This indicator is called the loss function. This paper uses a loss function based on the principle of the mean absolute error (MAE).

The prediction model consists of a layer of GRU network. In terms of network structure setting, we adapt Adam as optimization algorithm, set batch-size to 20, the number of hidden layers is set to 60， set sigmoid as the activation function.

C. Results Evaluation

In order to evaluate the effectiveness of the prediction model for the prediction of oil chromatography changes, this paper adopts three evaluation indicators to verify the fitting effect of the model on the test data set.

- MAE (Mean Absolute Error), the standard is the most commonly used evaluation index for prediction models. In the experiment, this index is also used as the loss function of model training. The calculation formula is shown in equation (5).

$$MAE(y_i, \hat{y}_i) = \frac{1}{n}\left(\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|\right) \tag{5}$$

- The fitting curve for the test set can most intuitively express the model's ability to predict future data changes. We use the first 596 sets of data (2018.09-2019.06) as the training set, and the last 280 sets of data (2019.06-2019.10) as the validation set. And use this model to fit the curve change of the data after four months.

- According to the standards of the gas generated by the IEEE oil-immersed transformer[15], the risk of the transformer is classified by the following four-level standards. The evaluation criteria are shown in Table I:

TABLE I. OIL AND GAS CONCENTRATION GRADE STANDARD

Stage	*H2*	*CH4*	*CO*	*CO2*
1	<100	<120	<350	<2500
2	100~700	120~400	350~570	2500~4000
3	700~1800	400~1000	570~1400	4000~10000
4	>1800	>1000	>1400	>10000

IV. CASES STUDIES

A. Forecast Result Evaluation

The data set we used is the oil chromatographic data of the main transformer insulation online monitoring device of a substation. The data period is from September 2018 to October 2019. Samples are taken every 12 hours, and there are 1066 sets of oil chromatographic data. The final evaluation results are shown in Table II.

TABLE II. PREDICTION RESULTS OF GASES

Model	*Train set (MAE)*	*Validation set (MAE)*	*Accuracy*
BP	19.59	26.67	83.7%
RNN	1.04	11.36	89.1%
LSTM	0.86	5.3	96.6%
GRU	0.77	5.1	97.1%

After the verification set is selected, the actual data of the 300 sets of data is compared with the fitting curve predicted by the model. From the simulation results, it can be seen that the model has a good prediction effect for the entire verification set. The GRU training set and the MAPE of the verification set are both in Within 5%. The graph of carbon dioxide gas content prediction curve is shown in Figure 3.

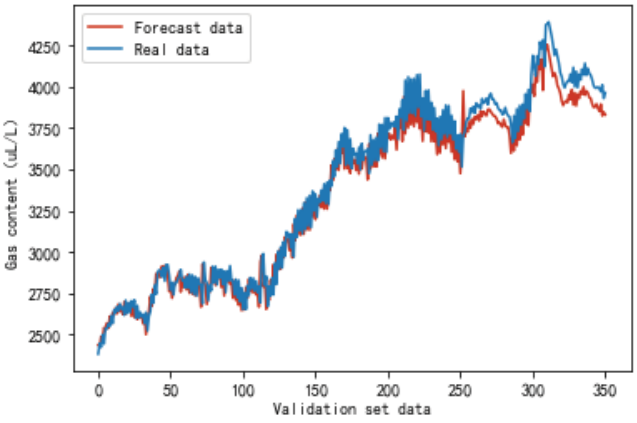

Fig. 3. Training flow chart of GRU.

B. Training Time Comparison

As a variant structure of LSTM, GRU network not only has high prediction accuracy, but also simplifies its own network structure. In order to verify the effect of this advantage in actual engineering, the GRU network is compared with RNN and LSTM, set the same network parameters and use the same training data. The obtained prediction model and test data are shown in Figure 4.

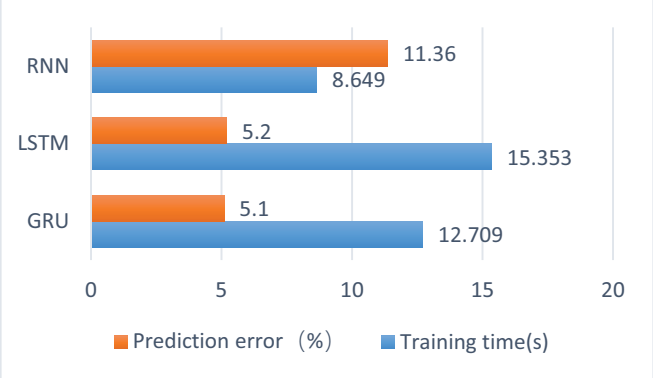

Fig. 4. Training time comparison.

It can be seen from the results that the prediction accuracy of the GRU network and the LSTM network is the same, but the training time of the GRU network is nearly 20% shorter. RNN network training time is very short, but the prediction accuracy is low.

C. Impact of Data Size on Prediction

There is a difference in the amount of transformer data in the actual working environment, so it is necessary to explore the impact of data size on the prediction accuracy. In order to study the impact of data size of the GRU algorithm, the original data set is divided. According to the original time interval, 1066 sets of data in the original data set are divided into data sets with sizes of 100, 800, and 1000, respectively. After the data size changes, some training parameters need to be reset. We set the ratio of training set for validation set of 4:1. The number of neurons is set to 60, and the number of iterations is 40. The results are compared by root mean square error and average percentage error. The test data gas is carbon dioxide gas with obvious changes. The final simulation results are shown in Table III.

TABLE III. PREDICTION RESULTS OF GASES

Data Size	*LSTM*		*GRU*	
	Root mean square error	Percentage error	Root mean square error	Percentage error
100	82.388	4.55	68.099	3.76
800	104.488	3.33	109.634	3.23
1000	124.319	2.73	120.413	2.86

According to Table III, as the scale of data decreases, the prediction accuracy of both networks decreases, and GRU's prediction accuracy for small sample data (100 groups) is better than that of LSTM networks.

D. Effect of Data Interval on Prediction

In order to explore the effect of data interval on model prediction, we use a network with the same parameters to train different sampled data sets. The sampling time of the data is 12 hours, 24 hours and 48 hours, respectively. The simulation results are shown in Figure 5. It can be seen that when the data interval of the training set becomes larger, the GRU network performs better than the LSTM network.

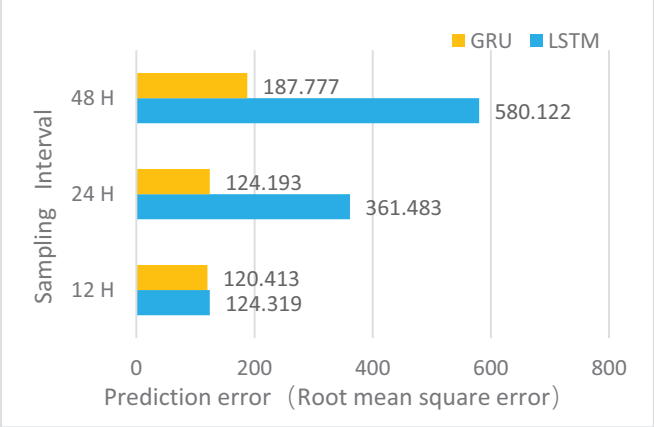

Fig. 5. Forecast accuracy at different sampling intervals.

V. CONCLUSION

Oil chromatography prediction, there are problems such as low prediction accuracy, poor generalization performance, and long model training time. To solve these problems, this paper proposed a prediction model based on GRU network. and compares it with other traditional prediction methods through simulation. The conclusions can be drawn as follows:

1) Base on the actual case analysis the oil chromatography state prediction established by GRU network has the better generalization ability compared with the traditional prediction method and avoids the introduction of the error due to subjective threshold selection.

2) Compared with the LSTM network, it has the same accuracy rate, but the training time is shorter, which is suitable for scenarios that require training time.

3) Because of its own network structure, the prediction accuracy of the GRU network when the data size is small or the data interval is long is better than the LSTM network.

ACKNOWLEDGMENT

This project was supported by National Natural Science Foundation of China(51777117), Natural Science Foundation of China & State Grid Corporation Smart Grid Joint Fund(U1966209).

REFERENCES

[1] M. Duval, “A review of faults detectable by gas-in-oil analysis in transformers”, IEEE Electr. Insul. Mag., Vol. 18, No. 3, pp. 8-17, 2002.

[2] K. P. Oehlmann and L. Mao, "Evaluation of distribution network transformer dissolved gas analysis (DGA) data," PES T&D 2012, Orlando, FL, 2012, pp. 1-4.

[3] S. Bustamante, M. Manana, A. Arroyo, R. Martinez, A. González and J. I. Rodríguez, "Case Study - Calculation of DGA Limit Values and

Sampling Interval in Power Transformers," 2019 6th International Advanced Research Workshop on Transformers (ARWtr), Cordoba, Spain, 2019, pp. 64-68

[4] B. Mohseni, N. Hashemnia and S. Islam, "Condition assessment of power transformer bushing using SFRA and DGA as auxiliary tools," 2016 IEEE International Conference on Power System Technology (POWERCON), Wollongong, NSW, 2016, pp. 1-4.

[5] K. D. PATEKAR and B. CHAUDHRY, "DGA analysis of transformer using Artificial neutral network to improve reliability in Power Transformers," 2019 IEEE 4th International Conference on Condition Assessment Techniques in Electrical Systems (CATCON), Chennai, India, 2019, pp. 1-5..

[6] IEEE Guide for the Interpretation of Gases Generated in Mineral Oil-Immersed Transformers," in IEEE Std C57.104-2019 (Revision of IEEE Std C57.104-2008) , vol., no., pp.1-98, 1 Nov. 2019

[7] 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.

[8] P. S. Huang, M. Kim, M. Hasegawa-Johnson and P. Smaragdis, "Joint optimization of masks and deep recurrent neural networks for monauralsource separation", IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 23, No.12, pp. 2136-2147, 2015.

[9] G. Hinton, O. Simon and T. Yee-Whye, "A fast learning algorithm for deep belief nets", Neural Comput., vol. 18, no. 7, pp. 1527–1554, 2006.

[10] H. Li, Y. Wang, X. Liang, Y. He and Y. Zhao, "Fault prediction of Power Transformer by Association Rules and Markov," 2018 IEEE International Conference on High Voltage Engineering and Application (ICHVE), ATHENS, Greece, 2018, pp. 1-4.

[11] K. Wang, "Characteristic parameters prediction of power transformer based on discrete gray model," 2012 Power Engineering and Automation Conference, Wuhan, 2012, pp. 1-4.

[12] Y. Shu and Y. Xu, "End-to-End Captcha Recognition Using Deep CNN-RNN Network," 2019 IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Chongqing, China, 2019, pp. 54-58.

[13] S. Hochreiter and J. Schmidhuber, "Long short-term memory", Neural Comput, Vol. 9, No.8, pp. 1735-1780, 1997.

[14] M. Pavithra, K. Saruladha and K. Sathyabama, "GRU Based Deep Learning Model for Prognosis Prediction of Disease Progression," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 840-844.

[15] IEEE Approved Draft Guide for the Interpretation of Gases Generated in Oil-Immersed Transformers," in IEEE PC57.104/D6.2, March 2019 , vol., no., pp.1-103, 13 June 2019