A Prediction Model for Time Series of Dissolved Gas Content in Transformer Oil Based on LSTM

Chuye Hu, Yang Zhong, Yiqi Lu, Xiaotong Luo* and Shaorong Wang

State Key Laboratory of Advanced Electromagnetic Engineering and Technology, Huazhong University of Science &Technology, Wuhan 430074, Hubei Province, China

*luoxt1996@163.com

Abstract. By combining with dissolved gas analysis, time series prediction of dissolved gas content in oil provides a basis for transformer fault diagnose and early warning. In the view of that, a prediction model based on long short time memory (LSTM) network for time series of dissolved gas content in oil is proposed, which takes advantage of LSTM network's ability to deal with long-sequence prediction problems. Five characteristic gas concentrations are used as input to the model, and the hyper parameters of the model is optimized by Bayesian optimization algorithm to further improve prediction accuracy, then a LSTM prediction model is constructed. By case study, it is verified that the proposed model can precisely predict time series of dissolved gas content. Compared with gray model, BP neural network and support vector machine, the proposed model has higher prediction accuracy and can better track the trend of time series of dissolved gas content in oil.

1. Introduction

As a core device in the power system, power transformer’s operating state is directly concerned with the safety and stability of the power grid [1]. During the running of transformer, a small amount of dissolved gases in oil will generate because of decomposition of the insulating oil or the solid insulating material. The content value and change trend of dissolved gases in oil closely correlate the fault of transformer. Therefore, dissolved gas analysis [2,3] is currently an internationally recognized method for detection of transformer faults and defects. The content value of each dissolved gas in oil is monitored by an online oil chromatography monitoring device to form historical detection data series. And by predicting the time sequences of dissolved gas content and apply dissolved gas analysis in gas content prediction results, condition maintenance can be arranged in advance to avoid future faults, as well as an important basis for transformer condition assessment can be provided [4].

So far, a lot of researches have been carried out on prediction of dissolved gas content in oil, include statistical prediction method like grey model (GM) [5], artificial intelligence prediction method like support vector machine (SVM) [6] and combined forecasting methods [7]. As a result of the traditional artificial intelligence prediction algorithm’ inadequacy in dealing with long time series, there is a big error in the data prediction results. With the rapid development of machine learning, because of RNN’s stronger adaptability and stability in time-series modeling and high prediction precision, it has now been applied to deal with time series prediction problem. However, "gradient dissipation" problem will happen in traditional RNN network when dealing with long sequence. As a special RNN model, LSTM [8-10] is widely applied to predict long time sequence due to the
introduction of modules with "memory function" into its structure, which can handle long sequence dependencies.

Therefore, a prediction model for time series of dissolved gas content in oil based on LSTM network is proposed. By time series segmentation, time series of 5 gas concentrations are used as input of the proposed prediction model. Then the hyper parameters of the model is optimized by Bayesian optimization algorithm to improve prediction accuracy. According to the optimized parameters, the LSTM prediction model is built. Finally, prediction of dissolved gas content in oil is made by the model. By case study, in comparison with other models, the proposed prediction model has greatly improved the prediction accuracy of single-step prediction and multi-step prediction, can track the trend of dissolved gas concentration more precisely.

2. LSTM recurrent neural networks

In the LSTM model, a gating cell is introduced into the network topology, making the model have “long time memory function”, which is suitable for long-term nonlinear sequence prediction problems. Compared with traditional RNN, hidden layer neurons in LSTM are replaced by memory cells with a gating mechanism. The basic structure of the memory cell [11,12] is as shown in figure 1.

**Figure 1.** The illustration of LSTM memory cell structure.

Memory cell is the key component of LSTM network. The cell input includes the sequence input $x_t$ at time $t$, the memory cell state $C_{t-1}$ at $t-1$ and the hidden layer cell state $h_{t-1}$ at $t-1$. The output contains the memory cell state $C_t$ at $t$ and the hidden layer cell state $h_t$ at $t$, where $C_t$ and $h_t$ each contain the model's long-term and short-term memory information. By controlling the above 3 gates, the reading and modification of the memory cell is realized, as well as the information flow between memory cells is carried out. The formulations are given by (1) to (3).

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f)$$  
$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i)$$  
$$O_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o)$$  

where $f_t$, $i_t$, and $O_t$ represent the state calculation results of the outputs of input gate, forget gate, and output gate, respectively; $W_{fx}$, $W_{fh}$, $W_{ix}$, $W_{ih}$, $W_{ox}$, $W_{oh}$ and $b_f$, $b_i$, $b_o$, respectively represent the weight matrix and offset term of the corresponding gate; $\sigma$ represents sigmoid activation function.

The output results of memory cell at time $t$ are memory cell state $C_t$ and hidden layer state $h_t$. The formulas are as follows:

$$\tilde{C_t} = \tanh(W_c h_{t-1} + W_f x_t + b_c)$$  
$$C_t = f_t \circ C_{t-1} + i_t \circ \tilde{C_t}$$  
$$h_t = O_t \circ \tanh(C_t)$$
where \( \tilde{C}_t \) represents the state input of memory cell at time \( t \), tanh is hyperbolic tangent activation function. \( W_C \) and \( b_C \) represent the state weight matrix and offset term of cell state, respectively. The symbol \( \odot \) means multiplying by element.

3. The prediction model for time series of dissolved gas content in oil based on LSTM

3.1. Data processing

So as to improve the data processing efficiency of sequences, sequence of each gas needs to be normalized separately. The formula is:

\[
\begin{align*}
    x^* &= \frac{x - x_{\min}}{x_{\max} - x_{\min}} \\
    \text{min} &\equiv x_{\min} \text{ and } x_{\max} \text{ are the minimum value and the maximum value of the sequence, } x \text{ is the sample data before conversion, and } x^* \text{ is the sample data after conversion.}
\end{align*}
\]  

After the data is normalized, the continuous time series is divided into input sequences and labels, according to the method of walk-forward division of the time series, where \( L_{\text{seq}} \) is time window length of time series segmentation, as well as the input sequence length into LSTM network. The schematic diagram of walk-forward method is shown in figure 2.

![Figure 2. Schematic diagram of walk-forward method.](image)

3.2. Hyper parameter optimization

The hyper parameters of LSTM prediction network include the input time sequence length \( L_{\text{seq}} \), the number of neurons of hidden layer \( N_{\text{hidden}} \), the hidden layer dimension \( D_{\text{hidden}} \), learning rate \( lr \), batch size of training \( S_{\text{batch}} \) and the training epochs \( Ep \). According to related literature [12], the network using a single LSTM layer has powerful nonlinear fitting ability, and its calculation and time cost are also controlled within a reasonable range. So the hidden layer dimension \( D_{\text{hidden}} \) is set as one. Meanwhile, by setting the training epochs \( Ep \) as 200, the training time cost is controlled and overfitting is avoided.

Hyper parameter optimization methods mainly include grid search algorithm, random search algorithm and Bayesian optimization algorithm and so on. Among them, the grid search algorithm [13] performs an exhaustive search in parameter list and train for each situation to find the optimal parameters, which is time-consuming when searching in high-dimensional space. Random search...
algorithm [13] extracts random parameter combinations from the specified distribution, in which above problem is avoided to a certain extent, but it still has the shortcoming that it cannot use the prior knowledge to achieve the next set of parameters and multiple parameters can easily lead to dimensional explosion. The basic idea of the Bayesian optimization algorithm [14] is to use the prior knowledge to approximate the posterior of the unknown objective function. The distribution further adjusts the hyper parameters, which can significantly improve the search efficiency and accuracy in high-dimensional space. With less iteration, it is still robust in non-convex problems.

In summary, this article will use Bayesian optimization algorithm to optimize and debug the four hyper parameters of $L_{seq}$, $N_{hidden}$, $lr$, and $S_{batch}$ of the LSTM neural network. An open source constrained global optimization tool for hyper parameter optimization called BayesianOptimization[15] is used to obtain above four hyper parameters.

3.3. The realization of LSTM network for dissolved gas content prediction

The generation process of gas in transformer oil is affected by oil pressure and temperature, so time series of its content has nonlinear characteristics. Considering that LSTM network overcomes the problem of "gradient dissipation", it can effectively deal with the problem of nonlinear sequence prediction. The proposed prediction model is based on the time series of five characteristic gases of $H_2$, $CH_4$, $C_2H_6$, CO, CO$_2$. The specific modeling process is shown in figure 3, and the steps are as follows:

![Figure 3. The architecture of the proposed prediction model.](image-url)
1) Collect the online oil chromatography sample data of the transformer, and take the time series of 
\( \text{H}_2 \), \( \text{CH}_4 \), \( \text{C}_2\text{H}_4 \), CO, CO\(_2\) as the model input;
2) Normalize the data of each feature gas and map it to [0, 1];
3) Use walk-forward method to divide the time series into input sequences \( X \) and labels \( y \);
4) Divide samples into training set and test set;
5) Construct the LSTM network model, hyper parameters of LSTM network is optimized by 
Bayesian optimization algorithm and the most suitable combination of \( L_{\text{seq}} \), \( N_{\text{hidden}} \), \( lr \), and \( S_{\text{batch}} \) is 
obtain;
6) The training set are learned in accordance with the back-propagation through time (BPTT) 
algorithm and getting network parameters;
7) Use the trained LSTM network to make prediction of the time series of \( \text{H}_2 \), \( \text{CH}_4 \), \( \text{C}_2\text{H}_4 \), CO, CO\(_2\).

4. Case study

In order to testify the effectiveness of the proposed prediction model, the case of this paper analyzes 
time series data using online monitoring of a 220KV transformer as an example (the time series of 
\( \text{C}_2\text{H}_2 \) and \( \text{C}_2\text{H}_6 \) are ignored because the contents of them are zero). The sample data is a total of 218 
sets since May 4, 2012 to December 7, 2012, of which detection period is 1 day. Use 196 sets of data 
from May 4, 2012 to November 15, 2012 as training set to train the prediction model, and use 22 sets 
of data from November 16, 2012 to December 7, 2012 as test set.

4.1. Experiment setting

The proposed prediction model is composed of a LSTM network and a layer of full-connect linear 
layer. The input layer size and hidden layer neurons \( N_{\text{hidden}} \) of LSTM are 5 and 128, respectively. The 
output layer size is 5. The training epochs \( Ep \) is set as 200. The hyper parameter combination 
obtained by Bayesian optimization algorithm are showed in table 1.

Meanwhile, construct GM model, SVM network and back propagation neural (BPNN) network 
to make prediction of time series of dissolved gas content. Among them, GM model is standard GM(1,1) 
model. The radial kernel function (RBF) is used in the SVM network model, where the kernel 
parameter is 0.001 and the penalty factor \( C \) is 8. BPNN with one hidden layer is constructed, and the 
number of neurons in each layer is 5, 128 and 5 respectively, with learning rate as 0.001 and training 
epoch as 200. All the above models are implemented based on Python language and under Pytorch 
framework.

Table 1. Hyper parameters of the proposed prediction model.

| Hyper parameter | \( L_{\text{seq}} \) | \( N_{\text{hidden}} \) | \( lr \) | \( S_{\text{batch}} \) |
|-----------------|---------------------|----------------------|--------|----------------|
| Value           | 30                  | 128                  | 0.001  | 29             |

4.2. Result analysis

4.2.1 Single-step prediction result analysis. Use training set to train above 4 models, and input a set of 
sequence in test set into the models to get single-step prediction, which is the next day’s gas content. 
The single-step prediction results of the above 4 prediction models in the test set are shown in table 2, 
where \( \epsilon \) means absolute percentage error, and is calculated according to following formula:

\[
\epsilon = \frac{|y_{\text{pred}} - y_{\text{label}}|}{y_{\text{label}}} \times 100\%
\]  

(8)
Table 2. Comparison chart of prediction results of 4 models in single-step prediction.

| Gas  | Real value (µL/L) | LSTM Predicted value(µL/L) | ε(%) | GM Predicted value(µL/L) | ε(%) | BPNN Predicted value(µL/L) | ε(%) | SVM Predicted value(µL/L) | ε(%) |
|------|------------------|----------------------------|------|--------------------------|------|-----------------------------|------|---------------------------|------|
| H₂   | 11.8             | 11.756                     | 0.370| 13.064                   | 10.718| 11.621                      | 1.520| 11.690                    | 0.935|
| CH₄  | 0.52             | 0.522                      | 0.406| 0.532                    | 2.360 | 0.514                       | 1.226| 0.516                     | 0.843|
| C₂H₄ | 10.6             | 10.817                     | 2.047| 11.003                   | 3.803 | 9.549                       | 9.911| 10.996                    | 3.733|
| CO   | 294              | 295.435                    | 0.488| 297.258                  | 1.108 | 276.575                     | 5.927| 297.057                   | 1.040|
| CO₂  | 1440             | 1428.690                   | 0.785| 1453.228                 | 1.335 | 1372.130                    | 4.713| 1406.681                  | 2.314|

Table 2 shows that the single-step prediction results of the proposed LSTM model have the smallest error compared with the real value, which is significantly better than the prediction results of the other 3 models.

4.2.2 Multi-step prediction results analysis. After making single-step prediction, the prediction result is integrated into historical monitoring data, and a new input sequence is constructed by time series segmentation included prediction results of last step of prediction. So that multi-step rolling prediction can be realized by input the new input sequence to the corresponding single-step prediction model. The above 4 models are also used to perform 1-5 step prediction on the test set, which means the next 5 days’ gas contents are predicted. The mean absolute percentage error (MAPE) is used to measure the error of prediction results of 4 models in each gas. The formula is as follows:

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|y_{pred} - y_{label}|}{y_{label}} \times 100\%$$  \hspace{1cm} (9)

where \( n \) is the number of samples. \( y_{pred} \) and \( y_{label} \) are the prediction value and real value of the sample respectively.

The result is shown in table 3. Select the predicted results of C₂H₄ content and compare them with the real value, and make a visualization as shown in figure 4.
Due to the accumulated error effect, the prediction accuracy of the proposed LSTM model decreases in multi-step prediction. From the prediction results shown in table 3 we can see, MAPE of 5 gases predicted by the proposed LSTM model are 1.836%, 1.136%, 2.249%, 1.180% and 1.740% respectively, which are clearly lower than the multi-step prediction results of GM, BPNN and SVM. From the above analysis, it is observed that the predictive performance of the proposed LSTM model in both single-step and multi-step prediction are significantly better than those of the other three models.

| Gas  | LSTM   | GM     | BPNN   | SVM    |
|------|--------|--------|--------|--------|
| H₂   | 1.836% | 10.977%| 2.200% | 3.658% |
| CH₄  | 1.136% | 4.237% | 3.855% | 1.982% |
| C₂H₄ | 2.249% | 3.610% | 13.467%| 3.607% |
| CO   | 1.180% | 4.027% | 4.663% | 3.896% |
| CO₂  | 1.740% | 4.015% | 8.634% | 3.746% |

5. Conclusion
In order to make prediction of time sequence of dissolved gas in oil to achieve early warning of transformer fault, a time series prediction model based on LSTM network for dissolved gas content in oil is proposed. Based on recurrent network structure and gating mechanism of LSTM neural network, the timing correlation of time sequences can be captured. The analysis results of the studied case show that, in contrast of the other 3 prediction models, the proposed prediction model has improved prediction accuracy, and can more effectively track the change trend of dissolved gas content in oil.

Because the amount of sampled dissolved gas content data is quite limited, the accuracy of multi-step prediction is not high. The next step will focus on the application in larger-scale data and the improvement and optimization of LSTM network model to further improve the accuracy of prediction.

References
[1] LI Gang , YU Changhai , LIU Yunpeng , et al 2017 Automation of Electric Power Systems Challenges and Prospects of Fault Prognostic and Health Management for Power Transformer 41(23):156-167.
[2] WANG Ke , LI Jinzhong , ZHANG Shuqi , et al 2016 Proceedings of the CSEE New Features Derived from Dissolved Gas Analysis for Fault Diagnosis of Power Transformers 36(23):6570-6578+6625.
[3] Mirowski P, LeCun Y 2012 IEEE Transactions on Power Delivery Statistical Machine Learning and Dissolved Gas Analysis: A Review 27(4): 1791-1799.
[4] WU Guangning , YAO Mengxi , XIN Dongli , et al 2015. Power System Technology Experimental Study on Oil-Impregnated Paper With Non-Uniform Thermal Aging 39(11):3298-3304.
[5] ZHENG Ruirui , ZHAO Jiying , ZHAO Tingting 2010 International Conference on Electrical & Control Engineering, IEEE Prediction of Power Transformer Oil Dissolved Gas Concentration Based on Modified Gray Model.
[6] Liao R J , Bian J P , Yang L J , et al 2012 Generation, Transmission & Distribution, IET Forecasting dissolved gases content in power transformer oil based on weakening buffer operator and least square support vector machine – Markov 6(2):p.142-151.
[7] Liao R J, Zheng H B , Grzybowski S , et al 2011 Electric Power Applications, IET Fuzzy information granulated particle swarm optimisation-support vector machine regression for the trend forecasting of dissolved gases in oil-filled transformers 5(2):p.230-237.
[8] LIU Yunpeng, XU Ziqiang, DONG Wangying, et al 2019 Proceedings of the CSEE Concentration Prediction of Dissolved Gases in Transformer Oil Based on Empirical Mode Decomposition and Long Short-term Memory Neural Networks 39(13):3998-4008.

[9] ZHU Qiaomu, LI Hongyi, WANG Ziqi1, et al 2017 Power System Technology Short-Term Wind Power Forecasting Based on LSTM 41(12):3797-3802.

[10] DAI Jiejie, SONG Hui, SHENG Gehao, et al 2018 International Conference on the Properties & Applications of Dielectric Materials LSTM networks for the trend prediction of gases dissolved in power transformer insulation oil.

[11] Sundermeyer M, Ney H and Schluter R 2015 IEEE/ACM Transactions on Audio, Speech, and Language Processing From feedforward to recurrent LSTM neural networks for language modelling 23(3):517-529.

[12] ZHOU Jie and XU Wei 2015 the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing End-to-end learning of semantic role labelling using recurrent neural networks 1127-1137.

[13] Bergstra J and Bengio Y 2012 Journal of Machine Learning Research Random search for hyper-parameter optimization 13(1):281-305.

[14] Bergstra J, Bardenet R, Bengio Y, et al. 2011 Conference on Neural Information Processing Systems, Granada, Spain Algorithms for hyper-parameter optimization 2546-2554.

[15] Nogueira F 2014 {Bayesian Optimization}: Open source constrained global optimization tool for {Python}, " https://github.com/fmfn/BayesianOptimization".