Trend Prediction of DC Measuring System Based on LSTM

Ying Pei 1, Lin Niu 1, Haifeng Li 1, Yajin Li 2, *, Dayang Yu 2

1 State Grid of China Technology College, Jinan 250002, China
2 School of electrical engineering, Shandong University, Jinan 250061, China

*Corresponding author e-mail: liuyingnan@sdu.edu.cn

Abstract. The accuracy of DC measurement system directly affects the reliable operation of DC control and protection system. In order to improve the estimation and prediction of DC measurement system operation state, a trend prediction algorithm based on multi-dimensional analysis and long-term memory network is proposed. Based on the analysis of DC measurement principle, the DC measurement status is diagnosed and abnormal is identified by time series trend analysis and anomaly detection. The LSTM is used to construct a multi factor driving current prediction model, and the model is trained and analyzed based on the actual operation data. Compared with the traditional time series prediction model, the results indicate that the proposed method is more accurate, simple and effective, and can be applied to the prediction of driving current.

Keywords: DC measuring system, LSTM, Trend prediction, multi-dimensional analysis.

1. Introduction

The DC measurement system provides input signals for control and protection which measurement accuracy directly affects the safe and stable operation of the HVDC transmission system in the UHV DC transmission project. At present, the failure of the DC optical measurement system is one of the main difficult problems faced by the operation of the DC transmission system of China Southern Power Grid in recent years. When the driving current or voltage measurement has a large deviation, the protection system may act, which will affect the normal operation of the converter station.[1]

The operation status of the DC measurement system is mainly monitored by the personnel's DC measurement data monitoring and threshold judgment methods in the daily operation and maintenance of DC converter stations. It is impossible to predict the operation status of the DC measurement system in advance. And when the DC measurement data does not exceed the alarm value, its abnormality (such as a slow growth trend) may be hidden in the fluctuation of the normal monitoring signal which is difficult to accurately identify it under manual monitoring. Because the DC measurement system has a high acquisition frequency and a large amount of data, simple manual statistical analysis is time-consuming and labor-intensive. Therefore, it is necessary to carry out forecasting and intelligent analysis for the trend of DC measurement data to provide auxiliary basis for the state evaluation of the DC measurement system.
Multi-feature fusion based on artificial intelligence is an effective means to improve the reliability of big data analysis and defect diagnosis [2], which has been used in the operation and maintenance of converter transformers [3], switches [4], arresters [5] and other equipment. However, there is no corresponding research on the prediction and analysis of the trend of DC measurement data. Existing research methods for equipment state prediction include time series-based regression analysis models [6]-[7], support vector machine algorithms [8], and deep learning algorithms [9], etc. However, the regression model based on time series is based on linear model, which is difficult to meet the non-linear characteristics of monitoring state quantity of DC measurement system. Although support vector machines have a certain generalization ability, the model prediction accuracy is difficult to improve with the increase of DC measurement characteristic factors. The prediction method based on deep learning enables the model to have non-linear expression capabilities, which can better adapt to the analysis and trend prediction of multi-dimensional data. Compared with the independent characteristics of the calculation results of other neural networks, the calculation results of each hidden layer of LSTM (Long Short-Term Memory) are related to the current input and the previous state, which has stronger adaptability and advantages in the equipment state timing prediction.

In the above context, this paper studies multi-dimensional time series trend analysis and correlation analysis algorithms based on the online monitoring data sources, and applies LSTM to DC measurement data trend prediction. Multi-dimensional analysis of the DC measurement system status evaluation and trend prediction model is established. The defects of the energy transmission fiber and the remote module in the measurement system are identified. Finally, the effectiveness and feasibility of the proposed method are verified combined with the ±800kV DC measurement system to collect field data.

2. Analysis of DC Measurement Principle

The structure of the DC optical measurement system of the converter station is shown in Figure 1. DC measurement systems generally use photoelectric hybrid methods to collect and transmit signals. The merging unit includes a laser driver and an I/O module. The laser driver supplies power to the remote module through an energy supply fiber, and the I/O module receives data collected by the remote module through the optical fiber.

After the measured state quantity of the primary equipment is stepped down, the analog signal of the state quantity is converted into an optical signal by the remote module, which is transmitted to the merging unit by optical fiber, and the data is transmitted to the control and protection system according to a certain communication protocol.

The causes of abnormal DC measurement data include the abnormality of the measuring device and the abnormality of the DC system itself. The DC measurement data includes drive current, data level and the temperature of the remote module. The overhaul test for the measuring device includes current ratio/voltage ratio check, voltage divider resistance, voltage divider capacitance, laser power, fiber attenuation test, etc., as shown in Table 1. Because periodic test data cannot be obtained online in real time, and it is difficult to meet the needs of online real-time analysis of the system, this paper focusing on analyzing the driving current and data level status based on online monitoring data, and providing decision-making basis for operation and maintenance personnel to inspect.
### Table 1. Multi-source data model

| type of data     | Monitoring data                        | Data Sources      |
|------------------|----------------------------------------|-------------------|
| Online Monitoring| Drive current                          | Laser             |
|                  | Data level                             | Remote module     |
|                  | temperature                            | Remote module     |
|                  | Measurements                           | Merge unit        |
| Maintenance test | Current ratio                          |                   |
|                  | Voltage ratio                          |                   |
|                  | Voltage divider resistance             |                   |
|                  | Voltage divider capacitor             | Field test report |
|                  | Laser power measurement               |                   |
|                  | Fiber attenuation test                |                   |

### 3. STATE EVALUATIONALGORITHM

#### 3.1. Multi-dimensional evaluation algorithm

The DC transmission system is generally equipped with AB two-sided screen cabinets under the same pole, and 4 sets of monitoring devices are configured for the same monitoring point. The key state quantities collected include drive current and data level. This paper uses vertical and horizontal statistical analysis methods, time series trend analysis methods, and outlier abnormality identification methods [10] to comprehensively evaluate the status of the DC measurement system. The specific method is as follows.

- Compare the drive current and data level of the high and low ends of the same pole horizontally. When the deviation is greater than the threshold $\delta_1$, it is judged as abnormal. The $\delta_1$ is a static value, which is the average of the difference between the historical high and low data of the same pole.
- The longitudinal analysis of drive current and data level is divided into the following two steps:
a) Calculate the overall trend change value \( \sum (x_i - x_{i-1}) \) of the sample sequence \( X(x_1, x_2, \ldots, x_n) \), when \( \sum (x_i - x_{i-1}) > \delta_2 \), the warning. Since this time series has a certain trend but not obvious periodicity, a dynamic threshold is configured to detect anomalies in the data. Based on the average value, maximum value and minimum value of the data in the past time period \( T \), let \( \delta_2 = \min \{ x_{\text{max}} - \bar{x}, \bar{x} - x_{\text{min}} \} \).

b) The box plot method is used to detect the abnormal value in the data batch. When the difference between the abnormal value and the mean value exceeds 3 times the standard deviation value exceeds, the data batch is judged as abnormal.

In addition, the outlier abnormality detection is performed by calculating the local outlier factor (LOF) of the data set and comparing the threshold value based on the density of the data set. The specific flow chart is shown in Figure 2.

\[
\begin{align*}
\text{P, O represent the test sample point in the sample set which dimension is m of each set,} \\
O = \{ x_{j1}, x_{j2}, \ldots, x_{jm} \}, P = \{ x_{j1}, x_{j2}, \ldots, x_{jm} \}.
\end{align*}
\]

\[
d(O, P) = d(P, O) = \sqrt{\sum (x_j - x_k)^2}
\]

\[
d_k(O, P) = \max \{|d(O, P), d_k(O)|\}
\]

\[
\rho_k(O) = \frac{\sum d_k(O, P)}{\rho_{\infty}(O)}
\]

\[
\text{LOF}_k(O) = \frac{\sum \rho_k(P)}{k}
\]
\[ d(O,P) \text{ is the Euclidean distance from } P \text{ to } O, \quad d_k(O,P) \text{ is the } k\text{-th reachable distance from } P \text{ to } O, \quad d_k(O) \text{ is the } k\text{-th distance of } O, \quad \rho_k(O) \text{ and } \text{LOF}_k(O) \text{ are the local reachability density and local outlier factor of } O. \]

Identify abnormal measurement points through outlier detection method, and check whether the corresponding energy transmission fiber and remote module are normal.

### 3.2. Prediction of the trend of drive current based on LSTM

The voltage is relatively stable, and the driving current is dynamically adjusted with the operating power of the system in the HVDC system. Therefore, this paper focuses on analyzing the trend state of the driving current. LSTM controls the input of information by introducing a gating mechanism, selectively adding new information and forgetting previously accumulated information in terms of state quantity predictions affected by multiple factors, so LSTM has stronger adaptability and generalization ability in dealing with non-linear and temporal prediction problems.

#### The principle of LSTM

LSTM network model includes forget gate, input gate and output gate. The input gate layer takes the input \( x_t \) of the current layer and the output \( h_{t-1} \) of the hidden unit at the previous moment as input, and the output result \( i_t \) is the information to be updated. The forget gate layer uses \( h_{t-1} \) and \( x_t \) to output \( f\).

\[
i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (5)
\]

\[
f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (6)
\]

Then update the cell state \( C_t \), where \( \tilde{C}_t \) is the candidate value.

\[
\tilde{C}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (7)
\]

\[
C_t = f_t * \tilde{C}_t + i_t * \tilde{C}_t \quad (8)
\]

The output gate layer runs a sigmoid layer to determine which part of the cell state is output. The update of the hidden state \( h_t \) consists of two parts, the first part is the acquisition of \( o_t \), and the second part is composed of \( o_t \) and the activation function.

\[
o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \quad (9)
\]

\[
h_t = o_t * \tanh(C_t) \quad (10)
\]

Wherein, \( \sigma \) and \( \tanh \) are activation functions. \( W_{xi}, W_{xf}, W_{xc}, W_{xo} \) are the weight coefficients from the input layer to the hidden layer. \( W_{hi}, W_{hf}, W_{hc}, W_{ho} \) are the weight coefficients of recursive connections. \( b_i, b_f, b_c, b_o \) are bias factors.

#### Implementation process

This paper fit the relationship between temperature, operating power, data level and drive current to achieve the prediction of drive current based on the LSTM model. The specific training process is shown in Figure 3.
Start

Drive current, data level, temperature, power

Data preprocessing (data elimination, normalization), and divide it into training set, validation set and test set according to the ratio of 3:1:1

Training set
Validation set
Test set

Adjust the number of input nodes, hidden neurons, iterations, and the learning rate, etc.

Model training

MAE?= Minimum stable value

Update the weight coefficient and bias, recalculate

BP neural network training

Model save

Result

Output

Fig. 3 Flow chart of forecast

MAE is selected to evaluate the accuracy of the model during the training process.

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} \left| x_i - y_i \right|
\]  

(11)

Wherein, \( n \) is the number of drive currents, \( x_i \) is the true value of the drive current at time \( t \), and \( y_i \) is the predicted value of the drive current at time \( t \) predicted by the prediction model established based on the time series.

4. Analysis of calculation examples

- Multi-dimensional evaluation

This paper collects the actual DC measurement data of a converter station as a sample set. The bipolar DC transmission system includes 180 measurement points, and the sample set covers the operation and maintenance transcript data from December 2018 to November 2019, the sample interval is 1 hour. The state quantities include drive current, data level, temperature, and active power. The monitoring trend of the DC measurement drive current of the polar 1 is shown in Figure 4.
First, analyze the data. The monitoring value of drive current and data level of the same measuring point does not exceed the threshold 65mA and 70mV, and the quantitative monitoring value of trend does not exceed the dynamic threshold range, and the system status is normal. The difference between the monitoring value of drive current and the average value does not exceed 3 times the standard deviation value, the drive current indicator is normal. The LOF judgment threshold is determined to be 1 through the historical abnormal drive current set. When the value of \( k \) is greater than 50, the detection accuracy of the data reaches 100%, and the detection error rate is 0%. The entire sequence was tested on this basis, and no outlier driving current data was found.

- **Prediction of drive current**
  
  Based on the sample data collection of a converter station, a model and simulation environment based on the LSTM algorithm is built, the data is trained offline, and the trained model is encapsulated and embedded in the early warning system.

  In this model, ADAM optimization algorithm is adopted, the initial value of the model learning rate is 0.01, the input dimension is 4, the output node is 1, the error threshold is \( 2 \times 10^{-2} \), the number of iterations epoch is 80, dropout is 0.2, batch size is 50. When the training error is less than the error threshold, the output drive current value is obtained, and the training ends.

  Take the driving current of a measuring point as an example for analysis, the graph of MAE of the test set and training set during 80 iterations is shown in Figure 5. After training for 40 epochs, the loss remained stable at 1.83%.

![Fig. 4 Monitoring data of high-side drive current of polar 1](image)

**Fig. 4** Monitoring data of high-side drive current of polar 1

![Fig. 5 Graph of MAE of the test set and training set](image)

**Fig. 5** Graph of MAE of the test set and training set

Import all data into the model for prediction and compare it with the real value. At the same time, the ARIMA model is used as a control group to conduct experiments to verify the accuracy of the LSTM prediction model. The prediction and comparison results are shown in Figure 6.
Fig. 6 Forecast results of different models

It can be seen both prediction models can simulate the changing trend of the drive current. The LSTM prediction model can track the change trend of the driving current very well, but there is a small deviation in the fitting of the true curve at the peaks and troughs, and the overall prediction error is 2.6089%. There is a large deviation between the predicted result and the actual value in the ARIMA model when the actual value of the drive current fluctuates and continues to decrease. The MAE evaluation index result of the test set of ARIMA is 36mA, which is much higher than the 9.27mA of LSTM.

The ARIMA model relies on the regularity of the data or the data after the difference, ARIMA is also greatly affected by the random fluctuation data, and ARIMA does not have the ability to "selective forgetting", so the accuracy of the ARIMA model under this data set is not high enough. In contrast, LSTM has lower error and higher accuracy, which can better fit the change trend of the actual drive current result value.

5. Conclusion
A prediction algorithm based on multi-dimensional analysis and LSTM network for the problem of state estimation and trend prediction of DC measurement monitoring quantity is proposed. The following conclusions can be obtained through theoretical research and example analysis.

- The drive current and data level of DC measurement are diagnosed by time-series trend analysis and outlier detection method to assist personnel in decision-making and the status maintenance of the energy transmission fiber and the remote module on the basis of the existing operation and maintenance threshold diagnosis. Algorithm application is simple and effective.
- A multi-input-single-output LSTM driving current prediction model is constructed combined with the operating power, which taking the actual DC measurement data of the converter station as a sample. The MAE is reduced by 33 mA compared with the ARIMA model under the same sample set, which verifies the validity and feasibility of the model.

References
[1] Zhang Yan, Ma Yi. Application Status and Research Progress of DC current measuring device for HVDC transmission system [J]. Electrical Measurement & Instrumentation, 2014, 51(11): 32-39.
[2] PU Tianjiao, QIAO Ji, HAN Xiao, et al. Research and Application of Artificial Intelligence in Operation and Maintenance for Power Equipment [J]. HIGH VOLTAGE ENGINEERING, 2020, 46(2): 369-383.
[3] ZHOU Chunyang, LI Yajin, LIUYingnan, et al. Transformer fault diagnosis method based on Bayesian Interference and multi-dimensional information fusion[J]. Transformer, 2020,57(11):15-20.
[4] ZHANG Cong-cong, WANG Gang, GAO Dong, et al. Partial discharge pattern recognition based on convolutional neural network. Advanced Technology of Electrical Engineering and Energy, 2021, 40(3): 72-80.
[5] Peng Hongxia, Wen Yan, Wang Lei, et al. Differential Operating Maintenance Technology of Power Equipment Based on Two-layer Knowledge Architecture [J]. High Voltage Apparatus,
2019, 55(07): 221-226.

[6] WANG Hongwei, ZHANG Limin, JIANG Jianping, et al. On-Line Monitoring and Analysis System of Lightning Arrester Leakage Current in UHV Substation [J], 2019, (06): 67-72.

[7] LIU Huixin, ZHANG Jiangleong, LIAN Hongsong, et al. Prediction of the Gases Dissolved in Transformer Oil by Sequential Learning [J]. High Voltage Apparatus, 2019, 55(12): 193-199.

[8] PENG Zaixing, WANG Song, YI Lin, et al. Research on Fault Diagnosis of High Voltage Circuit Breaker Opening/Closing Coils Based on SVM Principal Component Analysis [J]. High Voltage Apparatus, 2019, 55(07): 39-46.

[9] SONG H, DAI J J, LUO L G, et al. Power transformer operating state prediction method based on an LSTM network [J]. Energies, 2018, 11(4): 914-928.

[10] HUANG Shuang. Research on fault diagnosis technology of light measurement system in high voltage converter station [D]. Nanjing university of posts and telecommunications, 2020.