A Novel Sudden Fault Prediction Method Based on Hierarchical Structure with GRU Neural Network for X-ray High-voltage Power Supply

Jianlong Zhang, Qiao Li*, Bin Wang and Mengying Cui
School of Electronic Engineering, Xidian University, Xi’an Shaanxi, China

*Email: qli_1230@stu.xidian.edu.cn

Abstract: Considering the sudden fault feature of X-ray high-voltage power supply are mostly reflected in the high-frequency components, a hierarchical structure with time sequential neural network is proposed to predict sudden fault of X-ray high-voltage power supply in this paper. Gate recurrent unit is the basic unit of time sequential neural network. Firstly, multi-wavelet transform is used to decompose the signal of the X-ray high-voltage power signal to obtain the coefficients of each frequency band. Secondly, signal waveform reconstructed for each layer of wavelet coefficients is as the input of the gate recurrent unit network to predict signals at different frequencies. Finally, each gate recurrent unit performs multi-step prediction on the reconstructed waveform in this frequency band for sudden fault and all band prediction outputs are composed to obtain the final prediction results. The simulation experiment shows that our method has better performance than the Kalman filter, recurrent neural network and long-short term memory prediction methods, and can accurately predict sudden fault of X-ray high-voltage power supplies.

1. Introduction
X-ray equipment has been widely used in numerous fields of natural sciences e.g., medical diagnosis, safety inspection and microscopic three-dimensional imaging. The X-ray high-voltage power supply (XHPS) is the primary component of the X-ray system, and its performance determines the service life of X-ray system. The high-frequency and miniaturization of XHPS are the main development trends, which makes XHPS become more complex, and small fault may cause catastrophic accidents. Therefore, X-ray power fault prediction is of great significance to ensure the safe operation of X-ray systems.

X-ray high-voltage power signals are typical time sequence. Support vector machine (SVM) and linear regression are applied to traditional time sequential prediction. For example, Bao Yongsheng [1] et al. introduced SVM into short-term wind speed prediction, which can effectively mine the feature of wind speed sequence. Compared with the combined model of improved fuzzy analytic hierarchy process, it has better prediction effect. Zhang Suxiang [2] et al. successfully performed short-term power load forecasting based on a locally weighted linear regression model. These methods perform well in dealing with stationary signals. But as a typical nonlinear and non-stationary signal, XHPS signal cannot be accurately described by traditional mathematical models. The investigation shows that the time sequential neural network has achieved great results in the research of sequential data in
the fields of recognition, classification and reliability prediction because it considers the correlation between signal at different times [3]. For example, W. Wang [4] et.al successfully used the LSTM network to realize the fault time sequential prediction. Xue Yang [5] et.al proposed an ultra-short-term wind speed prediction method combining convolutional neural networks (CNN) [6] and gate recurrent unit (GRU) [7]. And this method has higher prediction accuracy than the current wind speed prediction methods. However, most of these methods are single-step prediction. They can only predict the value of the signal’s next moment and cannot describe sudden fault. So it is difficult to be directly applied to the field of X-ray high-voltage power fault prediction.

The research shows that the fault feature of the power signal is mainly reflected in the high-frequency component [8]. Wavelet transform realize multi-scale decomposition of signal and obtain local feature of fault signal, which is more conducive to predicting fault. Sequential neural networks perform better on sequence prediction problems. Based on this, we propose a time-frequency domain hierarchical structure based on the GRU time sequential neural network. Firstly, wavelet decomposition is used to obtain wavelet coefficients of a fault signal. Then, reconstructed signals in different frequency bands are as input of time sequential neural network. Finally, all frequency band prediction outputs are added to obtain the final prediction results.

The rest of this paper is organized as follows. Section 2 explains our approach in detail. In Section 3, we describe the experimental setup, simulation and obtained results. And we present our conclusions and discussions in Section 4.

2. Sudden fault prediction of XHPS based on hierarchical structure with GRU neural network

Figure 1 shows the fault prediction model of XHPS designed in this paper. Firstly, the signal waveform of XHPS is decomposed and reconstructed to obtain waveforms at different frequencies. Then each signal waveform is trained and predicted by GRU model, respectively. Finally, the prediction results of each subsequence are added in time domain to obtain the final fault prediction result. Typical fault waveforms are difficult to express with a single discrete timing point, and multi-point prediction can effectively express sudden fault instead. Our work considers the accuracy and actual situation of power fault prediction, does not rely on too much historical data.

Figure 1. The structure diagram of fault prediction Hierarchical Structure with GRU Neural Network.

2.1. Wavelet decomposition

The feature of fault signals in the power supply is mostly reflected in the high-frequency components and we can handle fault feature through time-frequency domain transformation. The common time-frequency domain transformation methods include Fourier transform and wavelet transform [9]. The Fourier transform expands the signal into the sum of sine waves of different frequencies, which does not possess the local time-frequency analysis capability. The wavelet transform can control the width of time domain and frequency domain under a certain time-frequency window, which is suitable
for analyzing the sudden part of the time sequence signal.

The definition of wavelet transform is to perform multi-scale transformation on the input signal $f(t)$ through the scaling translation operation. The definition is described as

$$WT(\alpha, \tau) = \frac{1}{\sqrt{\alpha}} \int_{-\infty}^{\infty} f(t) \varphi(\frac{t-\tau}{\alpha}) \, dt,$$

where $\alpha(\alpha > 0)$ is called the scaling factor and $\tau$ is called the translation factor. The former plays the role of contraction and extension of basic wavelet $\varphi(t)$, and the latter enables the wavelet to shift along the signal.

Through different values of $\alpha$ and $\tau$, each shift analysis can achieve approximation of different frequency signals.

There are many types of base wavelet, among which Daubechies (DB) wavelet [10] and Morlet wavelet [11] are commonly used in power fault prediction. Through analyzing the experimental comparative, we found that the fault feature is more obvious when DB wavelet is used as basic wavelet, so we select DB wavelet as the wavelet function.

Assuming that the discrete time sequence signal of X-ray power supply is $F(x) = \{x_1, x_2, \ldots, x_k\}$ and $k$ is the signal length, the discrete wavelet decomposition and waveform reconstruction [12] can be expressed as equations (2) and (3), i.e.,

$$a_{j,l} = \sum_m h[m-2l]a_{j-1,m}, \quad d_{j,l} = \sum_m g[m-2l]a_{j-1,m},$$

(2)

$$a_{j-1,m} = \sum_l a_{j,l}h[m-2l]+\sum_l d_{j,l}g[m-2l],$$

(3)

where $h[*]$ is a low-pass filter, $g[*]$ is a high-pass filter, $a' = \{a_{j,1}, a_{j,2}, \ldots, a_{j,k}\}$ represents the low-frequency component of the $j$ layer, and $d' = \{d_{j,1}, d_{j,2}, \ldots, d_{j,k}\}$ represents the high-frequency component of the $j$ layer. When $j = 0$, $a_0 = F(x)$.

Figure 2. Wavelet decomposition and reconstruction results of fault power supply. (a) is overload fault, (b) is spike fault.

Common current sudden fault includes overload, spikes and noise, etc. In order to verify the validity of wavelet decomposition and reconstruction for fault signals, we select the resonance current signal of overload and spike as the research object. We perform wavelet decomposition and reconstruction on the current signals to observe their performance at high-frequency. Figure 2 is the current signals for overload and spike fault, respectively. From top to bottom are the input signal and the third to sixth level wavelet high-frequency reconstruction signals. It can be seen that the fault signal is mainly reflected in high-frequency components, and the fault feature becomes more obvious as the frequency increases.

2.2 Time sequential neural network

Time sequential neural network has strong nonlinear dynamic mapping ability and dynamic memory
function, which is suitable for prediction of time sequence. After undergoing the process from recurrent neural network (RNN) [13] to long short-term memory (LSTM) [14], LSTM network has produced different variants. These networks play an important role in the processing of sequence data.

Due to the problems of gradient disappearance and gradient explosion in RNN, it is difficult to establish long-distance dependencies. LSTM improves structure to make up for this shortcoming. The input gate, output gate and forget gate are added in the LSTM model as shown in Figure 3, and the memory ability of these gates is used to solve the long-term dependence problem in the RNN model.

![Figure 3. The diagram of LSTM model.](image)

![Figure 4. The diagram of GRU model](image)

The GRU network combines the forget gate and the input gate into a single update gate, and also mixes the cell state and the hidden state. While reducing the amount of calculation, it did not reduce the performance of the network, so we select GRU as basic unit of time sequential neural network. The model structure diagram of GRU is shown in Figure 4.

The calculation process in the GRU unit structure is shown in Equation (4), (5), (6) and (7), i.e.,

\[ r_t = \sigma(W'_r, [h_{t-1}, x_t]) \]  \hspace{1cm} (4)
\[ s_t = \tanh(W'_r, [r_t * h_{t-1}, x_t]) \]  \hspace{1cm} (5)
\[ z_t = \sigma(W'_z, [h_{t-1}, x_t]) \]  \hspace{1cm} (6)
\[ h_t = (1 - z_t) * h_{t-1} + z_t * s_t \]  \hspace{1cm} (7)

where \( h_{t-1} \) represents the output of the neuron at the previous moment, \( x_t \) represents the input of the \( t \) moment. The reset gate \( r_t \) controls whether the calculation of the candidate state \( s_t \) depends on the state \( h_{t-1} \). The update gate \( z_t \) controls how much information the current state \( h_t \) needs to retain from the historical state \( h_{t-1} \) and how much new information is received from the candidate state \( s_t \).

where \( P(x_t) \) represents prediction result. \( P_r(x_t) \) represents prediction result of the different frequency bands of the GRU network, and \( r \) is specific frequency band.

The connection of a single-layer GRU network is shown in Figure 5. The network consists of an input layer, a hidden layer and an output layer. After feeding time sequence \( \{x_1, x_2, \ldots, x_m\} \), a hidden layer output sequence \( \{h_1, h_2, \ldots, h_m\} \) can be obtained and connected to the fully connected layer, where \( m \) represents the signal values at \( m \) times. The number of neurons in the fully connected layer is equal to the number of power signal to be predicted.

We divide the training and test sets in the same way. The first layer of low frequency signal subsequence is \( F_{ul}(x_t) = \{x_1, x_2, \ldots, x_k\} \). The time sequence \( \{x_1, x_2, \ldots, x_m\} \) at the first \( m \) moments is selected as the input, and the data \( \{x_{m+1}, x_{m+2}, \ldots, x_{m+n}\} \) are taken as the output, where \( n \) represents the number of output. Successively, the number of samples of the power signal is \( k - m - n + 1 \) and the length of each input sample is \( m \). Therefore, the input data in this paper is a 2-D array of \((k - m - n + 1) \times m\), and the output data is a 2-D array of \((k - m - n + 1) \times n\).
Power sequence data of the first $m$ time is used as input, and the method of time-frequency hierarchical and GRU network is used to predict the power signal of $n$ time after the $m$ time.

3. Experimental results and analysis

3.1 Experimental environment
The experimental data in the paper is the current signals of the XHPS resonance circuit board. The switch on the circuit board is a Silicon carbide (SiC) power switch. The current is a resonance current flowing through $C_r$ and $L_r$, and the resonance frequency is 250 KHZ.

In this paper, four methods are selected to predict the resonance current of the power supply, among which the fault type is overload. The comparison methods are Kalman filtering and hierarchical sequential neural network methods, respectively, and RNN, LSTM and GRU are used as the basic unit. A total of 2000 samples (each discrete length is 10000) are collected in this paper. The number of training sets is 1200, the number of test sets is 600, and the number of verification sets is 200. The proposed method is verified under the conditions that PC specifications are CPU pentium-i7 4.00GHz, RAM 16G. In addition, the simulation software is MATLAB 2018b, the deep learning framework is Keras.

The evaluation criterions, i.e., mean absolute percentage error (MAPE) and mean absolute error (MAE) formulated as equations (9) and (10), respectively, are utilized to evaluate the proposed method.

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|,$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|,$$

where $y_i$ is real value and $\hat{y}_i$ is the predicted value.

When multi-step prediction is performed, $\hat{y}_i$ is an array of length $n$. At this time, the predicted value at each time is averaged to calculate MAPE and MAE.

3.2 Experimental analysis
In order to verify the prediction effects of different methods, we compare the single-step prediction and multi-step prediction effects of the four prediction methods, respectively. It is difficult to accurately express the sudden power fault state with a single-step sequence since the pattern of sudden faults usually consist of the states of multiple points. Therefore, the experiment focuses on comparing the prediction performance differences of the four prediction methods under different $m$ and $n$ values.
Table 1 is the single-step prediction results of the four methods when \( m = 20 \) and \( n = 1 \). As the results shown in Table 1, compared with the Kalman filter, the sequential neural network prediction method has a better performance because it makes good use of the correlations of sequential data. Among the three neural network methods, GRU has a highest prediction accuracy.

| prediction model | MAE    | MAPE    |
|------------------|--------|---------|
| Kalman           | 0.5366 | 0.3029  |
| RNN              | 0.2125 | 0.1113  |
| LSTM             | 0.2086 | 0.1097  |
| GRU              | 0.1967 | 0.1020  |

The overload fault of XHPS resonance current need to be described with about 20 discrete points in the time domain, and single-step prediction cannot work. Therefore, this paper adopts a multi-step prediction method to calculate the prediction performance of each method by setting \( m = 10, 20, 30 \) and \( n = 1, 2, 3 \ldots m \), respectively. The experimental results are shown in Figure 6.

We count the time consumption of the three methods as shown in Table 2. It can be seen that the RNN model has the shortest time because of its simple structure, and the GRU model simplifies the LSTM structure, so the time is less than the LSTM.

| prediction model | Time /mins |
|------------------|------------|
| RNN              | 8          |
| LSTM             | 35         |
| GRU              | 23         |

![Figure 6](image1.png)

**Figure 6.** Multi-step prediction results. (a) \( m=10 \), (b) \( m=20 \), (c) \( m=30 \).

![Figure 7](image2.png)

**Figure 7.** Results of RNN, LSTM and GRU.

As it can be seen from Figure 6, the prediction performance of the three neural networks is similar. The initial error is large. The error decreases firstly, then it increases as \( n \) increases. Figure 7 shows the fault prediction results of RNN, LSTM and GRU. Green line represents the actual value of the signal, and red line represents the predicted value. It can be seen that the method proposed in this paper can better predict sudden fault. Generally speaking, the GRU in the temporal neural network achieved
better prediction results.

4. Conclusion
In this paper, wavelet analysis and GRU sequential network are introduced into XHPS sudden fault prediction. The wavelet decomposition is used to process the input signal to highlight the fault component. Using the GRU model to predict the resonance current state can strengthen the subsequent data nodes’ perception of the previous data nodes and is capable of making full use of data. A multi-step prediction method is adopted, and the prediction accuracy and the actual situation are considered at the same time, so the method has a high generalization ability.

Acknowledgments
This work was partially supported by the National Key R&D Program of China (No.2017YFF0104404), the Open fund of Shaanxi Key Laboratory of Integrated and Intelligent Navigation Project (No.SKLIIN-20180209), the National Natural Science Foundation of China (No.61571347, No.61971331)

References
[1] BAO Y and WU Z 2011 9 Short-term wind speed forecasting based on SVM time-series method [J] Electric Power 9.
[2] Zhang S X, Zhao B Z, Wang F Y and Zhang D 2015 Short-term power load forecasting based on big data Proceedings of the Csee(1).
[3] Cho, K, Van Merrienboer, B, Gulcehre, C, Bahdanau, D, Bougares, F and Schwenk, H, et al. 2014 Learning phrase representations using rnn encoder-decoder for statistical machine translation Computer Science.
[4] Wang, X, Wu, J, Liu, C, Yang, H and Niu, W 2018 Exploring lstm based recurrent neural network for failure time series prediction Beijing Hangkong Hangtian Daxue Xuebao Journal of Beijing University of Aeronautics and Astronautics 44(4) 772-784.
[5] Yang, X, Lin, W, Shu, W, Yafei, Z, and Ning, Z 2019 An ultra-short-term wind power forecasting model combined with cnn and gru networks Renewable Energy Resources.
[6] Lecun, Y, Bottou, L, Bengio, Y, and Haffner, P 1998 Gradient-based learning applied to document recognition Proceedings of the IEEE 86(11) 2278-2324.
[7] Cho, K, Van Merrienboer, B, Gulcehre, C, Bahdanau, D, Bougares, F, and Schwenk, H, et al. 2014. Learning phrase representations using rnn encoder-decoder for statistical machine translation Computer Science.
[8] Tan S W, Qin S R and Tang B P 2001 Time-frequency characteristic of wavelet base and its application transient signal detection Journal of Chongqing University(Natural Science Edition).
[9] Xu J and Wu L L 2015 Method of inspection combined with fourier transformation and wavelet transformation Techniques of Automation and Applications.
[10] Lewis, A S and Knowles, G 1991 Vlsi architecture for 2-d daubechies wavelet transform without multipliers Electronics Letters 27(2) 171.
[11] Lin, J, and Liangsheng, Q U 2000 Feature extraction based on morlet wavelet and its application for mechanical fault diagnosis Journal of Sound and Vibration 234(1) 135-148.
[12] Mallat, S, G 1989 A theory for multiresolution signal decomposition: the wavelet representation IEEE Transactions on Pattern Analysis and Machine Intelligence 11(7) 674-693.
[13] Lipton, Z C, Berkowitz, J and Elkan, C 2015 A critical review of recurrent neural networks for sequence learning arXiv preprint arXiv:1506.00019.
[14] Xingjian, S H I, Chen Z, Wang, H., Yeung, D. Y, Wong, W K and Woo, W C 2015 Convolutional LSTM network: A machine learning approach for precipitation nowcasting In Advances in neural information processing systems pp 802-810.