Novel Fault Location Method for Power Systems Based on Attention Mechanism and Double Structure GRU Neural Network

FAN ZHANG\textsuperscript{1}, QUNYING LIU\textsuperscript{1}, (Member, IEEE), YILU LIU\textsuperscript{2}, (Fellow, IEEE), NING TONG\textsuperscript{2}, SHUHENG CHEN\textsuperscript{1}, (Member, IEEE), AND CHANGHUA ZHANG\textsuperscript{1}, (Member, IEEE)

\textsuperscript{1}School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
\textsuperscript{2}Department of Electrical Engineering and Computer Science, Knoxville, TN 37996-2100, USA

Corresponding author: Qunying Liu (lqy1206@126.com)

This work was supported in part by the Natural Science Foundation of China under Grant NSFC51677020, in part by the China Postdoctoral Science Foundation under Grant 2015M572457, in part by the Provincial Key Laboratory of Power Electronics Energy Saving Technology and Equipment under Grant szjj2016-093, and in part by the 2019 Sichuan Province High Tech Project under Grant 19ZDYF0541.

ABSTRACT Fault location is one of the most essential techniques to maintain the stable operation of power systems. A fast and accurate fault location allows operators to restore power grids faster and avoid economic losses. Conventional methods rely on expert knowledge to extract the necessary features (e.g., DWT, DFT). For large systems, more coupling effects of transmission lines require more complex feature engineering, and incomplete features can easily introduce large errors. To overcome this, a deep learning approach without manual feature extraction is introduced to the fault location model under big data application. Towards this end, in the proposed method, the attention mechanism, the Bi-GRU and a dual structure network are applied to analyze the current data from different perspectives. Complete information for the fault features is extracted for the fault location. Based on a time series model and benefit from the ability to internally acquire the information architecture of faulty line, the established model is adaptive to the power grids with very complex topologies. Simulation results indicate that the proposed double-structure model reduces the maximum error and is less affected by noise. In comparison with different structures and different models, the proposed method shows better performance in IEEE 39-bus system.

INDEX TERMS Attention mechanism, Bi-GRU, dual structure network, fault location.

I. INTRODUCTION

The power system is one of the most complex man-made systems in the world. Due to the aging of transmission lines, the limits of their operation are approaching. Besides that, bad weather, changing surroundings and vandalism, etc., cause line faults to occur frequently. Line faults can seriously affect the transient stability of the entire system, resulting in substantial economic losses. Repairing or insulating the line by using a fast and accurate fault location technique is essential to maintain stable operation of the system and restore user power in time.

The fault location technology developed for distribution networks can be divided mainly into three types: (1) methods based on impedance, (2) methods based on traveling wave, and (3) methods based on training [1].

Impedance-based methods [2] have been applied to determine the fault position typically by using voltage and current measurements from one or more terminals to calculate line impedance. The impedance-based method, especially for the one using local-measurements, is simple, fast and independent of communications. The simplest method is only based on reactance by ignoring the influence of fault resistance and load current. However, this method may produce significant errors caused by remote-end current infeed, load impedance, power transmission angle and angle difference between the transmission line and electricity source impedances [3].

Based on the traveling wave method [4], [5], methods using the measuring instruments with high precision and the communication instruments have been generally used to detect the transient signals generated by the fault position. The calculation of correlation between the forward traveling waves and the backward one along a line, or the direct detection of the time the wave reaches the relay location, is conventionally
high precision. However, this type of approaches rely on the accuracy of the measurement and communication of instruments, whose large number of installations mean high cost in a large power grid. Furthermore, because measuring devices are at the terminals of transmission line, the waveform of the faulty traveling wave is severely attenuated after a long-distance propagation [5], [6]. In [4], a novel distributed system is introduced to improve the accuracy of fault location in the multi-terminal power system, by installing self-charged current measuring devices at the mid-point of each line.

With the rapid development of artificial intelligence, machine learning has begun to be applied to various fault location tasks, such as fault location of power systems, fault location of rudders [7], [8], etc. The measurement-based machine learning approach is considered to be a tool for performing soft computing in fault location. In [9], it has been introduced to train several neural networks to distinguish the fault position according to the different fault types. This work is also inspired by them. In some researches [10]–[15], the discrete wavelet transform has been used to extract features, and then machine learning has been used to predict fault position, which requires a lot of expert knowledge to build feature engineering. However, most of them can not capture the time series features in voltage or current signals. In [10], wavelet transform has been used to analyze the photovoltaic power generation system accompanied by high-frequency noises. By extracting useful features, the artificial neural network (ANN) has been used to locate fault lines. However, the method in [10] is not accurate enough for all types of faults. Moreover, due to the complex structure of a large-scale power grid, the prediction results of the model will not always maintain high accuracy, that is, individual samples will have a significant error. Then, deep learning is introduced to extract the time series features of fault current [13]. Considering that different fault lines have different fault currents, the model is difficult to extract features for different fault lines because it can not know which line has fault. The Decision Tree Regression method [15] has also been used to extract the characteristics of the voltage and current signals by comparing the discrete Fourier transform with the discrete wavelet transform. And then, the regression tree has been used to determine the fault position, but the proposed method has been identified that the results under some experimental samples have large error. In this mode, the learning algorithm is responsible for discovering the complicated relationship between the hidden rules and the pattern features [16]. In [17], the discrete Fourier transform has been used to extract the features of the voltage signal. Then the support vector machine (SVM) has been used to determine the fault line and the fault type. Bewley diagrams have been introduced to observe the traveling-wave patterns, and the wavelet coefficients of the voltage of aerial mode have been used to locate the fault. However, the proposed method has only been identified as a three-machine system. The precise fault location is a regression task, which needs high precision to be applied to practice. Therefore, using traditional machine learning algorithms for high-precision regression tasks, especially in large power grids, is conventionally difficult to meet the low error of fault location. In most research, the information of voltage and current changing with time and the influence of network complexity are ignored. Therefore, it is a crucial breakthrough to mine the characteristics of time series from the data generated by large-scale power grid.

In recent years, with the improvement of hardware computing ability and the rapid development of deep learning methods, more research has been focused on deep learning, such as fault diagnose [18], the transient stability analysis [19] and fault location [10], [20], etc. An improved extreme learning machine [21] has been designed to locate faults, but it does not take into account the features of the time series. In [22], signals have been processed through Empirical Mode Decomposition (EMD) and used to train a CNN. Then the CNN’s classification mechanism and linear regression mechanism have been used to achieve fault position on HVDC. The influence of line impedance (i.e. different lines) has been analyzed by the model, but the model can not automatically learn to adapt to different fault lines. Therefore, there is still an urgent need for a breakthrough in fault location of grid based on deep learning. In the existing learning-based methods, feature extraction is often needed first, and then the machine learning methods have been used for fault location. These methods need to discuss the effects of the extracted features, which have also been rarely tested on larger systems.

This work proposes a new fault location classification method, which is inspired by the field of speech recognition. The method can determine fault position and the fault line by extracting and analyzing the line’s data for current measurements. In order to locate and classify faults, the proposed algorithm learns the time series of the current, the fault position and type by mapping their relationships. Tested on the IEEE-39 bus system, the accuracy of the method is verified by different data sampling frequencies, different degrees of noise interference and measurement error. Also, comparison studies are conducted among the proposed method and several existing methods. The main contributions of this work are as follows:

1. Feature extraction-dependent methods require a wealth of prior knowledge. In contrast to this, the RNN-based neural network is introduced to automatically extract the characteristics of the grid data over time series.
2. A dual structure is used in the proposed system framework to simultaneously output the location of the faulty line and the specific fault location of the faulty line. Among them, the location of the fault line is used as a feature for fault location, and such a mechanism can adapt the model to a more complex network.
3. By applying the attention mechanism, the data before and after failure are analyzed, and more subtle changes in different fault locations are captured.
4. The proposed method is tested on the IEEE39 bus system, and different fault types and fault locations are set to prove the robustness of the system.
II. FAULT LOCATION ALGORITHM

In this section, the relationship between the current signal and the fault location, that between the voltage signal and the fault location are elaborated in detail. The changes of voltage and current with regard of different locations of the line are demonstrated by simulation.

![Three-phase Radial network with fault at distance p from the source side.](image)

A typical type of fault is set on the bus. Using the post-fault information of voltage and current, the relationship between the loop impedance, the loop voltage and the loop current under normal conditions can be derived. Considering that three-phase-to-ground fault occurs on the line from the source p as shown in Figure 1, the voltage and current on the line have the following relationship [23]:

\[ V\sub{p}(f) = Z_f I_f \]

\[ V\sub{i}(f) = V\sub{i}(0) - Z_i I_f \]

(1)

According to equation (3), equation(2) is expressed as:

\[ V\sub{i}(f) = V\sub{i}(0) - Z_{ip} \cdot Z_{pp}^{-1} \cdot V\sub{p}(0) \]

(2)

where, \( V\sub{i}(0) \) and \( V\sub{i}(f) \) are the pre-fault normal voltages of bus i in the case of a specific fault; \( Z_i \) is the equivalent impedance of bus i; \( I_f \) and \( Z_f \) are the fault current and the fault impedance, respectively; \( V\sub{p}(f) \) is the fault voltage at the fault position.

By equation (2), \( g(Z_{pp}, Z_{ip}) = \text{Distance from sending end to fault position} \). However, the operation in the actual power grid is complex and varied. It is necessary to consider the interference of more factors for deriving the corresponding impedance. Because the properties of each transmission line are different, the deep learning method is introduced to drive the machine learning to solve the relationship between the fault position and the line voltage and current.

In order to better prove the relationship between the current and the fault position, the changes of bus superimposed current when faults occur at different locations of the IEEE39-bus system are drawn, as shown in Figure 2. Assuming that the fault occurs at 1.0s and is removed at 1.1s on lines B1-39(the transmission line between bus 1 and bus 39), B3-4, B5-6, B7-8, B10-13 and B16-17 (other lines are similar), the fault is set to occur at 5%, 25%, 45%, 65% and 85% of the entire length of the line, respectively. The fault type is single-line-to-ground (other fault types are similar). Four PMUs are set up to collect bus current (as shown in Figure 3). As is indicated, with the change of the fault position, the changing trends of the current form a clear law. However, the variation laws of the fault currents at different fault lines are quite different. In order to draw a universal law for the machine to learn when faults are located at different places, a two-module learning algorithm is designed to learn the specific location of the fault and the position of the faulty line. The current varies with the fault position when the model distinguishes different lines. The details of the specific model will be discussed in Section IV.

III. BI-GRU AND ATTENTION MECHANISM

In this section, the Recurrent Neural Network (RNN) and Bidirectional Gated Recurrent Unit is first introduced (Bi-GRU) [27], where the reasons of employing the GRU and attention mechanism are explained in details.

A. BIDIRECTIONAL GATED RECURRENT UNIT

RNN [24] is a structurally unique artificial neural network. From the structural point of view, the RNN has a self-looping network. When the data is passed through the RNN, the RNN will retain the state information of the current data and calculate it together when next data coming. The output of RNN is obtained by the following equation:

\[ h_t = f(W_{xt} + Hh_{t-1} + b) \]

(3)

where, \( h_t \) is the output of the neural network at time \( t \); \( W, H \) and \( b \) are the parameters to be learned by the network; \( f(\cdot) \) is the nonlinear activation function. During the calculation process, the RNN can generate an output at each time step and establish a cyclic network of loop connections between the hidden units at the current moment and output to the next moment. Such a network can allow RNN to selectively retain the important features of time on the sequence.

However, in the process of training, especially when there is long sequence data, the traditional RNN structure is prone to gradient disappearance and gradient explosion, which results in that some neuron weights can’t be updated, in which case the network may fail to learn. The Long-Short Term Memory (LSTM) [25] has solved this problem and efficiently preserved the characteristics of long sequences. Gated Recurrent Unit [26] (GRU) is a variant of LSTM, which requires fewer parameters than LSTM and can obtain higher accuracy. When handling the massive data from large power grids, the GRU training takes less time to achieve nearly the same accuracy. Because LSTM and GRU can more efficiently acquire features on time series, they are widely adopted in the fields such as machine translation and speech recognition. The architecture of GRU is presented in Figure 4. The internal calculation of GRU is as follows:

\[ z_t = \sigma(W_{xz}x_t + W_{zh}h_{t-1} + b_z) \]

(4a)

\[ r_t = \sigma(W_{rx}x_t + W_{rh}h_{t-1} + b_r) \]

(4b)
In classical cyclic neural networks, the state transmission is one-way from forward to backward. However, in some problems, the output of the current moment is related not only to the previous state but also to the subsequent state. At this time, bidirectional GRU (Bi-GRU) is needed to solve this problem. In this model, the GRU output at \( t \) time depends not only on the information at time \( t - 1 \), but also on the information at time \( t + 1 \) [28]. The architecture of GRU is presented in Figure 5. The algorithm of Bi-GRU is as follows:

1. For \( t = 1 \) to \( T \): Forward pass for the forward hidden layer, storing activations at each timestep.
2. For \( t = T \) to 1: Forward pass for the backward hidden layer, storing activations at each timestep.

In classical cyclic neural networks, the state transmission is one-way from forward to backward. However, in some problems, the output of the current moment is related not only to the previous state but also to the subsequent state. At this time, bidirectional GRU (Bi-GRU) is needed to solve this problem. In this model, the GRU output at \( t \) time depends not only on the information at time \( t - 1 \), but also on the information at time \( t + 1 \) [28]. The architecture of GRU is presented in Figure 5. The algorithm of Bi-GRU is as follows:

1. For \( t = 1 \) to \( T \): Forward pass for the forward hidden layer, storing activations at each timestep.
2. For \( t = T \) to 1: Forward pass for the backward hidden layer, storing activations at each timestep.
3. For all $t$, in any order: Forward pass for the output layer, using the stored activations from both hidden layers.

B. ATTENTION MECHANISM

Attention is a mechanism that mimics the way humans look at things and focus their attention on the critical areas. Through attention mechanisms, more attention is paid on the more significant information in time series in the model. Nowadays, attention mechanism has been used in many fields, such as machine translation [29], semantic segmentation [30] and photovoltaic power generation prediction [31].

When a fault occurs in the power grid (for example, three-phase-to-ground fault), the voltage and current on the transmission line will have a sudden change, and the circuit breaker of the corresponding line can operate fast enough to interrupt the fault within merely 0.1 seconds. We need to explore the changes of voltage and current on the line within 0.1 seconds from the time before the fault to the time after the fault, while the information of position of the fault can’t be extracted from the voltage and current before the fault. So the model should focus more on the changes of voltage and current after the occurrence of the fault. Therefore, the attention mechanism is introduced, which combines the attention mechanism with the model, in order to make the model focused more on the useful information in the time series of voltage and current. The attention mechanism is given by the following formula:

$$e_i = \tanh(W_h h_i + b_h)$$  \hspace{1cm} (5a)

$$\alpha_i = \frac{\exp(e_i)}{\sum_{i=1}^{L} \exp(e_i)}$$

$$h' = \sum_{i=1}^{L} \alpha_i h_i$$  \hspace{1cm} (5b)

where, $h_i$ is a state variable of network output; $W_h$ and $b_h$ are weights and deviations; $\alpha_i$ is calculated by software max. Finally, the weighted state $h_i$ is summed up.

IV. PROPOSED MODEL

A single input dual output structure is used in the proposed model. The proposed framework is indicated by Figure 6. In the input layer, a total of $t$ seconds of bus superimposed current before and after the fault on the $l$ lines are selected as inputs. In order to identify the change of the fault current, it is imperative to intercept the obtained current signal in the time domain. A total of $l$ time series, each sequence be sampled $t$ seconds and the matrix formed by $t \times l$ are fed into the network.

The time series of currents is entered into the Bi-GRU unit where the two layers of Bi-GRU are used. The advantage of using the variant GRU unit of LSTM is to increase the speed of the model of the network while maintaining high precision. When the model is applied to a large grid, the dimensions of the input data increases as that of grid increases. To avoid a very long delay caused by a large amount of input data, the GRU model is employed in this paper with less training parameters. Thus, the training speed of the model can be enhanced significantly. With the Bi-GRU structure, when the input current signal is processed, the current sequence is input in both forward and reverse directions, which allows the model to simultaneously analyze the current information before and after the fault and retain more useful information.

The information in the time series extracted by the Bi-GRU unit is passed by the two parts of the model, which are named the fault location positioning module(FLPM) and the fault line identification module(FLIM). The following describes the specific process of the model and then explains the reasons why such a model is designed. As shown in Figure 6, the right half of the model is fault line identification module(FLIM), whose task is to locate the faulted line. The feature information output by the Bi-GRU is subjected to Reshape module, and then the classifier composed of the fully connected layer and the Softmax layer is inputted for multi-classification, and the probability of each line failure is outputted. The activation function used by the full connection layer is Relu [32].
The Softmax layer formula is as follows:

\[ S_i = \frac{e^{O_i}}{\sum C_j e^{O_j}} \quad (6) \]

where, \( O_i \) is the output value of the \( i \)-th neuron. Here, \( S_i \) is equivalent to the probability of the \( i \)-th class.

Conventionally, the transmission line with the highest probability of failure is selected as the current fault line. As shown in Figure 6, the Attention module, the Bi-GRU module and the probability information of the fault position are calculated by the right half fault line positioning module, and the results are input to the Fully connected layer for regression tasks to determine the exact position on the fault line.

In an actual power grid, the physical properties such as line length and impedance of each transmission line are different. When a fault occurs on different fault lines, even if the fault type and position are the same, the current and voltage fluctuations resulted from them are different. This is why the multiple classifiers for different lines have been trained in [9]. This work has been inspired to design a new network structure, which solves this problem well. The information in the time series extracted by the RNN does not clearly indicate the location of the faulty line, so the classifier cannot obtain the information of the faulted line, and the classifier cannot accurately locate the fault. In order to solve this problem, the designed model adds an internal information channel, which can make the FLPM obtain the information of the faulty line. In order to enable Bi-GRU to obtain the information of the faulty line, the FLIM is designed. The faulty line is determined by the time series features extracted by the Bi-GRU. During training, back propagation corrects the parameters of the Bi-GRU based on the loss function of the faulty line module, allowing the Bi-GRU to extract features containing the information of faulty line. Because the information in the time series is used, the accuracy of the FLIM can easily reach 100%, which makes the FLIM provide an accurate information model without the prediction error caused by the FLIM. By using this designation, the model is capable of obtaining time-series information and faulted line information. Combined with the two features, the accuracy of the classifier is greatly improved. The attention mechanism adds different weights to the time-series features. Since the fault position information is implied in the fluctuation of voltage and current after the fault, the use of dynamic RNN, it can support different time windows), including the 0.1s pre-fault data and 0.1s post-fault data. The sampling time is 0.01s. It is necessary to preprocess the collected data, including standardize data to normal distribution. The input data of current is shown in the following matrix:

\[ I = \begin{bmatrix}
  i_{11} & i_{12} & \ldots & i_{1l} \\
  i_{21} & \ldots & \ldots & i_{2l} \\
  \vdots & \vdots & \ddots & \vdots \\
  i_{t1} & \ldots & \ldots & i_{tl}
\end{bmatrix} \quad (7) \]

Among them, \( i_{tl} \) is the current at \( t \) time on the line \( l \).

The data is used to train the model and save the trained model. When the fault occurs, PMU is used to obtain the current data before and after the fault (the same as the training data), and the trained model is used to predict the fault position and fault line.

### V. MODEL TRAINING

Training samples and test samples are obtained through simulations. The model is trained offline by using training samples and evaluated by test samples. Since the task of the FLIM in the model is a classification, to classify the task of the FLPM is a regression task, so different loss functions are respectively used in the two modules. The cross-entropy shown as equation(8) is used in the fault line positioning as a loss function:

\[ L_{FLIM} = -\sum C_j s_j \log p_j \quad (8) \]

\[ L_{FLPM} = \frac{1}{2} \sum (y - \hat{y})^2 \quad (9) \]

where, \( s_j \) is the output of the softmax layer, \( p_j \) is the probability of the \( j \)-th class, \( y \) is the actual value, and \( \hat{y} \) is the predicted value.
function:
\[
\text{loss}_{\text{fault-line}} = -\sum_{i=1}^{n} \sum_{j=1}^{m} y_{ij} \log \hat{y}_{ij}
\]  

(8)

where, \(n\) is the number of samples; \(m\) is the number of categories; \(y_{ij}\) is the probability that the sample is the \(j\)-th class, and \(\hat{y}_{ij}\) is the probability that the model prediction sample is the \(j\)-th class.

The fault location is based on the mean squared loss function:
\[
\text{loss}_{\text{fault-loc}} = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_{i} - y_{i})^2
\]  

(9)

where, \(n\) is the number of \(n\) samples; \(y_{i}\) is the actual value, and \(\hat{y}_{i}\) is the predicted value of the model.

Both modules in the model use the Adam optimizer [33]. The learning rate is 0.01. To prevent over-fittings of the model, an early stop strategy is adopted. When the error of the verification set is not reduced in three epochs, the training process should be stopped. The optimal model obtained from the training uses the data from the test set to evaluate the model. The parameter settings of the model are shown in Table 1.

| Layer | Number of units |
|-------|-----------------|
| The 1st Bi-GRU | 128 |
| The 2nd Bi-GRU | 128 |
| Fully connected (Fault line part) | 64 |
| The 1st Fully connected (Fault location part) | 128 |
| The 2nd Fully connected (Fault location part) | 128 |

The training model needs to complete two different tasks at the same time, while the main task is fault location (completed by FLPM), so FLIM is used to assist FLPM to complete fault location. Figure 8 shows the relationship between the loss and accuracy of the validation set during training. Seen from the figure 8, part of the loss of FLIM decreases rapidly with the advancement of training, and the accuracy of FLIM quickly reaches 100%, which indentifies that FLIM will not bring burden to model training.

VI. CASE STUDY

The New England 10-machine system [34] is adopted to evaluate the performance of the proposed model. Modeling and simulation work is executed by using PSASP (Power System Analysis Software Package), a commercial power system simulation package developed by Electric Power Research Institute, China.

All numerical simulations are conducted on a computer with an Intel Core i5 CPU working at 3.2 GHz and 8 GB RAM. The model is constructed with Keras [35]. The proposed system is implemented in Python.

A. MODEL EVALUATION

The function to evaluate the error is shown as
\[
E_{\text{err}} = \frac{(L_{\text{predict}} - L_{\text{actual}})}{L_{\text{total}}}
\]  

(10)

where, \(L_{\text{predict}}\) is the predicted fault position; \(L_{\text{actual}}\) is the actual location of the fault; \(L_{\text{total}}\) is the total length of the faulty line and \(E_{\text{err}}\) is the error.

Since the test system is a multi-line system, the physical parameters of every transmission line are different, especially the difference in the line length. For the IEEE-39 bus system, when the fault occurs on a shorter line, the predicted value of the model will range from 0km to 10km. When the fault occurs on a line with a longer length, the predicted value of the model ranges from 0km to 100km. Thus, in the case of a large difference in line length, the target difference in model learning is also relatively large. For different faulted lines, it is difficult for the model to learn from the common change of voltage and current when the fault occurs at different lines. However, this normal change is pervasive in the power grid. If this fact isn’t taken into consideration, it will lead to a large error in the individual samples predicted by the model. If more accurate results are required, the strong fitting of deep learning is depended to set up a deeper network, which will undoubtedly greatly increase the training time. Also, more training samples are needed for convergence purpose, since the different lines have different dimensions. In order to unify the dimensions, it needs to normalize them, that is, calculate the percentage of the distance from the fault point to the sending end in the total distance. For all possible faulty lines, the model’s predicted value is between 0% and 100%, which makes it easier to learn the same pattern of voltage and current changes when two lines with large differences in length suffer the fault. However, if the model error is significant and the error occurs on a longer line, which depends on that the fault position equal to the line length multiplied by the
percentage of the fault position, it means that the longer line will amplify this error. Therefore, in the actual case, it takes the operator more time to find the real fault position. For the evaluation function, setting the predicted value in this way does not cause any error. So the real source of error is the evaluation function.

**B. FAULT SETTING**

In the New England 10-machine system with total 34 lines, single-line-to-ground (a-g), single-line-disconnection (a-d), line-to-line fault (a-b), line-line-to-ground fault (ab-g) and three-phase-to-ground fault (abc-g) are respectively set. The fault position and the fault resistance are also set as shown in table 2. Considering the actual situation, the fastest response speed of the circuit breaker is set as 0.1 seconds after the occurrence of the fault.

**TABLE 2.** Fault setting.

| Fault Setting     | Values                                      |
|-------------------|---------------------------------------------|
| Fault Line        | Line-1 ~ Line-34                            |
| Fault Type        | a-g, a-d, ab-g, ab, abc-g                   |
| Fault Position    | 5% ~ 95%(Step:5%)                           |
| Fault Resistance(Ω)| 0.01, 5, 15, 25, 50                        |

**C. PERFORMANCE OF THE PROPOSED MODEL**

K-fold cross-validation is used to test the performance of the model. During this period, 90% data is used as training set and 10% data as test set. There are 16150 samples in total, 1700 samples (all samples with faults at 20% and 80% of the line) are used as the test set (10% of the total data), and the rest samples (faults at other locations) are used as the training set. The 5-fold cross-validation and even 80% sample training model of training set data are used. Furthermore, 20% of training set data is used as a validation set and is repeated five times. Then, the trained model is used to predict the test set.

The average error and maximum error of the proposed model are 0.087% and 0.413%, respectively. The accuracy of the fault line location is 100%. Meanwhile, the average error and maximum error of different fault types are counted, and the results are shown in Figure 9. The average error and the maximum error of two short-circuit faults are the largest among the five fault types. The proportion of samples in different error ranges to the total sample is shown in the Figure 10. The error of more than 70% samples is less than 0.1%. It is worth pointing out that although the current information of the fault line can not be obtained from the broken-line fault, the fault location can also be obtained from the current of other lines.

**D. EXPERIMENTAL RESULTS OF DIFFERENT STRUCTURES**

According to the model from easy to difficult, the performance of the model under a variety of structures is compared to analyze the importance of each module in the model and the average and maximum errors for all samples are calculated. The results are shown in Table 3, in which the structure of the LSTM-based model extracts time series features for the LSTM unit and then flattens it, using MLP for prediction. The training time of the CNN is very short, however, it can only capture local features and cannot capture the features of time series. The structure of the Bi-GRU Attention-based model is the introduction of attention mechanisms. As shown in the figure11, when the fault occurs at 1.00s, the attention weight between 1.01s and 1.07s is greater than that at other times, which indicates that the attention mechanism helps the model focus more on the current change after the fault.
TABLE 3. Experimental results under different structures.

| Method                      | Mean Error | Max Error | Training Time | Execution Time |
|-----------------------------|------------|-----------|---------------|----------------|
| CNN-based model             | 0.83%      | 4.552%    | 21.7s         | 0.1150s        |
| LSTM-based model            | 0.628%     | 7.854%    | 127.0s        | 0.2431s        |
| GRU-based model             | 0.625%     | 6.321%    | 77.1s         | 0.1491s        |
| BiGRU-based model           | 0.552%     | 3.232%    | 133.5s        | 0.4224s        |
| BiGRU-Attention-based model | 0.45%      | 2.150%    | 144.0s        | 0.4242s        |
| BiGRU-based double structure model | 0.281% | 0.732% | 136.3s | 0.4833s |
| BiGRU-Attention-based double structure model | 0.087% | 0.413% | 222.3s | 0.5193s |

The proposed model yields the best results but its training time is longer. However, in practice, the training time of the model does not affect the deployment of the system. The GPU or TPU is also used to accelerate the model training, which can make the large model for the large system and complete the training faster. The execution time also increases with the size of the model, but this execution time is within an acceptable range. Generally, the relay protection device is triggered within 0.1s ∼ 0.2s after the fault occurs. The model uses data from 0s ∼ 0.1s, so the acceptable decision time is within 0.1s, and the fault location can be obtained before the relay protection is triggered.

The performance of the model is shown in Table 3. As can be seen from the experimental results, the dual structure model can significantly improve the accuracy and reduce the maximum error. The current model parameters are shown in Table 1. The model training time is 222 seconds. As the system becomes larger, the model training time will increase. When using offline training, in most cases, as long as the training time is within an acceptable range, it has little effect on the overall system.

E. EXPERIMENTAL RESULTS OF DIFFERENT MODELS

In order to show the performance of the proposed model, different classifiers are compared. Under the condition of ensuring fairness, the parameters of other classifiers of different types are optimized. Five different classifiers are set here, including the regression tree (RT), the linear regression (LR), the random forest (RF), the support vector regression (SVR) and the backpropagation neural network (BPNN).

The pre-processed sequential data are flattened and inputted into the classifier and the classifier is used to discriminate the data. The results are shown in Table 4. It can be seen that most classifiers can’t achieve the accuracy of the proposed model. Firstly, because these classifiers can’t consider the time series of current signals, they can’t extract the time series characteristics of current signals. Secondly, these classifiers can’t get the position information of the fault line, whereas different fault lines are beneficial in determining the specific fault position.

F. IMPACT OF PMU PLACEMENT RATIO ON PERFORMANCE

In this subsection, the impact of the proportion of PMU deployed by the system on the performance of the model is discussed. When the line fails, it will affect the adjacent lines. In particular, for break-up faults, the current of break-up faults is 0. Therefore, the model can’t be judged by the current of the transmission line, but by that of the adjacent lines, or the lines that are affected severely.

The PMU coverage of different sizes is set to analyze the impact of coverage on the results. The proportion of PMU deployment is the proportion of the number of lines deployed in PMU to the total number of lines. The comparative deployment ratio is 40%, 50%, 60%, 70%, 80%, 90% and 100% respectively. Since the placement of deployment is random, 20 placements are tested randomly for each deployment ratio, and the results are averaged to reduce errors. In Figure 12, the results show that when the PMU coverage is less than 70%, it will have a greater impact on the average error. Furthermore, when the PMU coverage is less than 80%, it will have a greater impact on the maximum error.

G. EFFECT OF NOISE ON PERFORMANCE

Since the signal-to-noise ratio of PMU measurement is not specified in the IEEE standard c37.118 [36], the signal-to-noise ratio of PMU in different regions may be different.
(3) A dual-module architecture of fault line location is added so that the model can use the fault line location to determine the specific fault position.

According to the simulation results, the accuracy of fault location of the proposed model is much improved by comparing with the traditional model, and the maximum error of the prediction results is greatly reduced.

Future work is to optimize the model structure to improve the training speed of the model. Since the fault line discrimination module is added to greatly reduce the prediction error, the fault type discrimination module can also be added to the model to further test the performance improvement of the model. Furthermore, to design the model so that it can handle multiple failures occurring simultaneously, will be much meaningful to the grid security.

REFERENCES

[1] M. Majidi and M. Etezadi-Amoli, “A new fault location technique in smart distribution networks using synchronized/nonsynchronized measurements,” IEEE Trans. Power Del., vol. 33, no. 3, pp. 1358–1368, Jun. 2018.
[2] T. P. S. Bains, T. S. Sidhu, Z. Xu, I. Voloh, and M. R. D. Zadeh, “Impedance-based fault location algorithm for ground faults in series-capacitor-compensated transmission lines,” IEEE Trans. Power Del., vol. 33, no. 1, pp. 189–199, Feb. 2018.
[3] M. Kezunovic, “Smart fault location for smart grids,” IEEE Trans. Smart Grid, vol. 2, no. 1, pp. 11–22, Mar. 2011.
[4] C. Zhang, G. Song, T. Wang, and L. Yang, “Single-ended traveled wave fault location method in DC transmission line based on wave front information,” IEEE Trans. Power Del., vol. 34, no. 5, pp. 2028–2038, Oct. 2019.
[5] J. Ding, X. Wang, Y. Zheng, and L. Li, “Distributed traveling-wave-based fault-location algorithm embedded in multiterminal transmission lines,” IEEE Trans. Power Del., vol. 33, no. 6, pp. 3045–3054, Dec. 2018.
[6] F. V. Lopes, K. M. Dantas, K. M. Silva, and F. B. Costa, “Accurate two-terminal transmission line fault location using traveling waves,” IEEE Trans. Power Del., vol. 33, no. 2, pp. 873–880, Apr. 2018.
[7] B. Chang, R. Yang, C. Guo, S. Ge, and L. Li, “Performance evaluation and prediction of rudders based on machine learning technology,” Proc. Inst. Mech. Eng., G, J. Aerosp. Eng., vol. 233, no. 15, pp. 5746–5757, Dec. 2019, doi: 10.1177/0954410019857380.
[8] B. Chang, R. Yang, C. Guo, S. Ge, and L. Li, “A new application of optimized random forest algorithms in intelligent fault location of rudders,” IEEE Access, vol. 7, pp. 94276–94283, 2019.
[9] R. N. Mahanty and P. B. D. Gupta, “Application of RBF neural network to fault classification and location in transmission lines,” IEE Proc.-Gener., Transmiss. Distrib., vol. 151, no. 2, p. 201, 2004.
[10] I. M. Karmacharya and R. Gokaraju, “Fault location in ungrounded photovoltaic system using wavelets and ANN,” IEEE Trans. Power Del., vol. 33, no. 2, pp. 549–559, Apr. 2018.
[11] A. Abdullah, “Ultrafast transmission line fault detection using a DWT-based ANN,” IEEE Trans. Ind. Appl., vol. 54, no. 2, pp. 1182–1193, Mar. 2018.
[12] D. K. J. S. Jayamaha, N. W. A. Lidula, and A. D. Rajapakse, “Wavelet-multi resolution analysis based ANN architecture for fault detection and localization in DC microgrids,” IEEE Access, vol. 7, pp. 145371–145384, 2019.
[13] J. J. Q. Yu, Y. Hou, A. Y. S. Lam, and V. O. K. Li, “Intelligent fault detection scheme for microgrids with wavelet-based deep neural networks,” IEEE Trans. Smart Grid, vol. 10, no. 2, pp. 1694–1703, Mar. 2019.
[14] M. Mirzaei, B. Vahidi, and S. H. Hosseinian, “Unsynchronised fault-location technique for three-terminal lines,” IET Gener., Transmiss. Distrib., vol. 4, no. 13, pp. 2770–2778, 2019.
[15] A. Swetapadma and A. Yaday, “A novel decision tree regression-based fault distance estimation scheme for transmission lines,” IEEE Trans. Power Del., vol. 32, no. 1, pp. 234–245, Feb. 2017.
[16] M. Farshad and J. Sadegh, “Accurate single-phase fault-location method for transmission lines based on K-Nearest neighbor algorithm using one-end voltage,” IEEE Trans. Power Del., vol. 27, no. 4, pp. 2360–2367, Oct. 2012.

[17] H. Livani and C. Y. Evrenosoglu, “A fault classification and localization method for three-terminal circuits using machine learning,” IEEE Trans. Power Del., vol. 28, no. 4, pp. 2282–2290, Oct. 2013.

[18] K. Chen, J. Hu, and J. He, “Detection and classification of transmission line faults based on unsupervised feature learning and convolutional sparse autoencoder,” IEEE Trans. Smart Grid, vol. 9, no. 3, pp. 1748–1758, May 2018, doi: 10.1109/TSG.2016.2598881.

[19] J. J. Q. Yu, D. J. Hill, A. Y. S. Lam, J. Gu, and V. O. K. Li, “Intelligent time-adaptive transient stability assessment system,” IEEE Trans. Power Syst., vol. 33, no. 1, pp. 1049–1058, Jan. 2018.

[20] W. Li, D. Deka, M. Cherkov, and M. Wang, “Real-time faulted line localization and PMU placement in power systems through convolutional neural networks,” IEEE Trans. Power Syst., vol. 34, no. 6, pp. 4640–4651, Nov. 2019.

[21] Y. Q. Chen, O. Fink, and G. Sansavini, “Combined fault location and classification for power transmission lines fault diagnosis with integrated feature extraction,” IEEE Trans. Ind. Electron., vol. 65, no. 1, pp. 561–569, Jan. 2018.

[22] S. Lan, M.-J. Chen, and D.-Y. Chen, “A novel HVDC double-terminal non-synchronous fault location method based on convolutional neural network,” IEEE Trans. Power Del., vol. 34, no. 3, pp. 848–857, Jun. 2019.

[23] D. Thukaram, H. P. Khinchia, and H. P. Vijaynarasimha, “Artificial neural network and support vector machine approach for locating faults in radial distribution systems,” IEEE Trans. Power Del., vol. 20, no. 2, pp. 710–721, Apr. 2005.

[24] K. Narendra and K. Parthasarathy, “Identification and control of dynamical systems using neural networks,” IEEE Trans. Neural Netw., vol. 1, no. 1, pp. 4–27, Mar. 1990.

[25] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[26] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, “Empirical evaluation of gated recurrent neural networks on sequence modeling,” 2014, arXiv:1412.3555. [Online]. Available: https://arxiv.org/abs/1412.3555

[27] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks,” IEEE Trans. Signal Process., vol. 45, no. 11, pp. 2673–2681, Nov. 1997.

[28] L. Zhang and L. Ma, “Coattention based BiLSTM for answer selection,” in Proc. IEEE Int. Conf. Inf. Automat. (ICIA), Jul. 2017, pp. 1005–1011.

[29] H. Choi, K. Cho, and Y. Bengio, “Fine-grained attention mechanism for neural machine translation,” Neurocomputing, vol. 284, pp. 171–176, Apr. 2018.

[30] G. Lin, A. Milan, C. Shen, and I. Reid, “RefineNet: Multi-path refinement networks for high-resolution semantic segmentation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 1925–1934.

[31] H. Zhou, Y. Zhang, L. Yang, Q. Liu, K. Yan, and Y. Du, “Short-term photovoltaic power forecasting based on long short term memory neural network and attention mechanism,” IEEE Access, vol. 7, pp. 78063–78074, 2019.

[32] C. M. Bishop, “Probability distributions,” in Pattern Recognition and Machine Learning, 2nd ed. New York, NY, USA: Springer, 2006, ch. 2, sec. 4, pp. 113–116.

[33] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in Proc. Int. Conf. Learn. Represent., San Diego, CA, USA, May 2015, pp. 1–15.

[34] A. Pai, Energy Function Analysis for Power System Stability, Boston, MA, USA: Kluwer, 1989.

[35] (2017). Keras: Theano-Based Deep Learning Library. [Online]. Available: http://keras.io/

[36] K. E. Martin, “Synchrophasor measurements under the IEEE standard C37.118.1-2011 with amendment C37.118.1a,” IEEE Trans. Power Del., vol. 30, no. 3, pp. 1514–1522, Jun. 2015.

[37] M. Brown, M. Biswal, S. Brahma, S. J. Ranade, and H. Cao, “Characterizing and quantifying noise in PMU data,” in Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM), Jul. 2016, pp. 1–5.

FAN ZHANG received the B.S. degree in automation from Tianqang University, Tianjin, China, in 2018. He is currently pursuing the M.S. degree with the School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu, China. His research interests include the fault location of power systems and machine learning.

QUNYING LIU (Member, IEEE) received the B.Eng., M.S., and Ph.D. degrees in electrical engineering from Sichuan University, Chengdu, China, in 2000, 2005, and 2008, respectively. She is currently with the School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu. Her current research interests include power system security and stability analysis, and advanced application of PMUs.

YILU LIU (Fellow, IEEE) received the B.S. degree from Xi’an Jiaotong University, Xi’an, China, and the M.S. and Ph.D. degrees from Ohio State University, Columbus, OH, USA, in 1986 and 1989, respectively. She was the Governor’s Chair with the University of Tennessee, Knoxville, TN, USA, and the Oak Ridge National Laboratory (ORNL). She was elected as a member of the National Academy of Engineering, in 2016. She is also the Deputy Director of the DOE/NSF Co-Funded Engineering Research Center. Prior to joining UTK/ORNL, she was a Professor at Virginia Tech. She led the effort to create the North American power grid frequency monitoring network at Virginia Tech, which is now operated at UTK and ORNL as Grid-Eye. Her current research interests include power system wide-area monitoring and control, large interconnection-level dynamic simulations, electromagnetic transient analysis, and power transformer modeling and diagnosis.

NING TONG received the Ph.D. degree from the Huazhong University of Science and Technology, in 2016, where he is conducting postdoctoral research in electrical engineering. His interests include power system protective relaying and smart grid.
SHUHENG CHEN (Member, IEEE) received the B.Eng. and M.Sc. degrees from the Harbin Institute of Technology, Harbin, China, in 1998 and 2000, respectively, and the Ph.D. degree from Sichuan University, Chengdu, China, in 2007, all in electrical engineering.

From September 2013 to September 2014, he was a Visiting Scholar with the Department of Energy Technology, Aalborg University, Aalborg, Denmark. Since 2008, he has been a Lecturer with the School of Energy Science and Engineering, University of Electronic Science and Technology of China, Chengdu. His research interests include distribution system analysis, reconfiguration, and reactive power optimization.

CHANGHUA ZHANG (Member, IEEE) was born in Hubei, China. He received the Diploma degree in vehicle engineering from the Wuhan University of Science and Technology, Wuhan, China, in 1993, the M.Sc. degree in vehicle engineering from the Wuhan University of Technology, Wuhan, in 2004, the Ph.D. degree from the Institute of Automation Chinese Academy of Sciences, in 2007.

From November 2011 to October 2012, he was a Visiting Scholar with Loughborough University, U.K. He is currently an Associate Professor with the School of Mechanical and Electrical Engineering, University of Electronics Science and Technology of China (UESTC). His current research and academic interests include the smart grid, converter control, and microgrids analysis theory.

***