Deep Belief Network Based Faulty Feeder Detection of Single-Phase Ground Fault

HONGBO WEI¹, HUA WEI², ZHONGLIANG LYU², XIAOQING BAI² AND JUNYANG TIAN¹

¹Power Dispatch and Control Center of Guangxi Power Grid Co., Ltd., China Southern Power Grid, Nanning, 530023, China.
²Guangxi Key Laboratory of Power System Optimization and Energy Technology, Guangxi University, Nanning 530004, China.

Corresponding author: Hua Wei (e-mail: weihuagu@163.com).

This work was supported in part by the National Natural Science Foundation of China under Grant 51967002; and in part by the Science and Technology Projects of China Southern Power Grid Co., Ltd. under Grant GXKJXM20190619, Grant GXKJXM20190680.

ABSTRACT This paper proposes a faulty feeder detection method based on Deep Belief Network (DBN) of deep learning theory for the neutral non-effectively grounded systems. It consists of two steps: firstly, a DBN-based faulty feeder detection model is built with feeder current, power and power factor as input feature parameters. Then, the input feature data are obtained during the single-phase ground fault from the master station of power dispatching system, which will construct a training set. By unsupervised pre-training and supervised fine-tuning, the proposed model obtains the mapping relationship between raw data and fault characteristics and realizes the faulty feeder detection. The advantage of the proposed method is using millisecond-level data in power dispatching system directly. Moreover, the sampling device does not need to install, which significantly reduces the construction costs and is of strong adaptability. The analyzed result using the ground fault data of an actual substation for more than two years shows that the proposed method has a better performance than SVM and BP neural network, and the accuracy is up to 94.7%. The proposed method has been implemented in Lipu Power Grid, Guangxi, China with excellent application effect and extensive application prospects.

INDEX TERMS Deep belief network, faulty feeder detection, power dispatching system data, single-phase ground fault.

I. INTRODUCTION

The neutral non-effectively grounded system is widely used in 3 ~ 60kV voltage level power grid in China. A distinguishing feature of this systems is the ability to continue operating for 1 ~ 2 hours when single-phase ground fault occurs [1]. However, because of its low magnitude of the ground fault current, the single-phase ground fault is very difficult to detect. Operation with fault easily leads to over voltage, resulting in multiple accidents, such as insulation breakdown and TV high voltage fuse breaking. Statistics show that single-phase ground fault is the most common fault type in neutral non-effectively grounded systems [2], accounting for more than 80% of distribution network faults [3]. Therefore, the faulty feeder must be quickly determined and cut off.

There are various reasons for single-phase ground fault, such as equipment aging, branches press on distribution line, etc. During a ground fault, the steady state signal is weak, and the transient signal is coupling and complex, which makes it difficult to extract the fault characteristics accurately. Therefore, the faulty feeder detection of single-phase ground fault problem still open and is a hot topic for research [4], [5].

In the past decade, many methods [6] have been developed to improve the detection of the faulty feeder in a neutral non-effectively grounded distribution system. According to the type of signal used, the existing methods can be divided into three classes: steady-state method, transient method and machine learning method.

The steady-state method mainly includes zero-sequence current specific amplitude and specific phase method, harmonic method, residual flow increment method, etc. In most cases, these methods can solve the single-phase ground fault of a neutral ungrounded system and neutral via resistance grounded system very well. However, there are some defects in principle when applying these to resonance grounded system, which can easily lead to misjudgment [6].
Transient method, such as the wavelet method [1], [7], [8], traveling wave method [9], [10], etc., uses the transient signal during the ground fault to identify the faulty feeder and is used more widely because it is not easily affected by the arc suppression coil. However, limited by the short duration of the transient process, it is difficult to obtain the information of the fault instant effectively. Accordingly, its actual application effect is not ideal.

Machine learning methods usually based on artificial intelligence and integrating various fault information or identification techniques. These methods mainly involve ANN (Artificial Neural Network), SVM (Support Vector Machine), etc. [11], [12]. The self-learning ability of these methods can effectively solve the faulty feeder detection problem under various working conditions. However, the ability of these shallow learning methods to extract fault characteristics is limited, and the generalization performance is restricted. When the number of feeders is large, the dimension of this algorithm is very high, and the convergence speed is slow or even not convergent. Therefore, it is difficult to obtain satisfactory results in actual application.

In the methods mentioned above, whether based on single fault characteristic or multiple fault feature fusion, in essence, the detection process can be summarized as ‘raw data - signal processing - extract fault characteristic - determine faulty feeder.’ This kind of signal processing based characteristic extraction method usually requires an in-depth understanding of signal characteristics to extract typical fault characteristics effectively [13]. When a single-phase ground fault occurs, the operating conditions of the system are complicated, the grounding arc is unstable, the fault information is coupled severely and the change of fault current is difficult to measure [14]. Thus, it is challenging to extract the typical characteristics that correctly reflect the fault condition, which leads to a reduced application performance for most detection methods.

Deep learning, as an important branch of artificial intelligence, has been widely concerned since Geoffrey Hinton et al proposed it. Its application in the field of fault diagnosis has become an important research topic, which provides a new way to solve the faulty feeder detection problem. With the support of deep network models and large data, deep learning shows excellent performance in fitting prediction, characteristic extraction and category determination through unsupervised greedy layer-wise training [7], [15], [16]. As one of the most effective deep learning methods, Deep Belief Network (DBN) has achieved remarkable application effects in the fields of speech detection [17], [18], image recognition [19]–[21], and mechanical fault diagnosis [13], [22], [23]. However, its research in the field of faulty feeder detection of neutral non-effectively grounded systems is still blank.

This paper proposes a DBN-based faulty feeder detection method of single-phase ground fault. Firstly, using the feeder operation data collected by the power dispatching system to train the detection model. With the help of excellent characteristic extraction performance of multiple hidden layer DBN, the highly essential characteristics in the data are obtained. Then, a small number of labeled data are used to fine-tune the model to improve the generalization performance of the model. Finally, a complete faulty feeder detection model is formed, which simplifies the identification process to ‘raw data - DBN - determine faulty feeder.’ Combined with the power dispatching system proposed in [24], a faulty feeder detection system based on the proposed method has been realized in Lipu Power Grid, Guangxi, China. The analysis of actual power grid data shows that the proposed method can adaptively extract fault characteristics from the raw data and identify the faulty feeder of single-phase ground fault, which is of high practicability.

The main contributions of this paper are as follows:

1) This paper presents the problems of the existing faulty feeder detection method, and proposes a new one based on DBN with the data of an actual power dispatching system. Also, the principles of DBN and modeling process of the proposed detection method are described in detail;

2) This paper discusses the influence of activation function, training samples and network structure on the accuracy of faulty feeder detection. Through comparison, it is further shown that the faulty feeder detection method based on DBN has higher accuracy than traditional artificial intelligence algorithm.

The rest of the paper is organized as follows: Section II describes the theory, architecture and training process of DBN. In Section III, the faulty line detection based on DBN is proposed, including data acquisition and procession, construction of training set, as well as the modeling of faulty feeder detection. Based on data of an actual power dispatching system, Section IV discusses the influence of activation function, training samples and network structure on accuracy of faulty feeder detection, respectively. Moreover, the detection performance of traditional artificial intelligence algorithms and the proposed method is compared. Section V presents the application effect of a faulty feeder detection system based on the proposed method. At last, Section VI concludes the paper.

II. DEEP BELIEF NETWORK

DBN is a probability generation model proposed by Geoffrey Hinton in 2006 [25]. Its structure can be regarded as a multi-layer neural network composed of multiple Restricted Boltzmann Machines (RBMs). A well-trained DBN can adaptively extract the essential characteristics of raw data to achieve classification recognition [16].

A. RESTRICTED BOLTZMANN MACHINE

Restricted Boltzmann Machine consists of two network layers, and its structure is shown in Fig.1. The visual layer $v$ is used to input data samples, and the hidden layer $h$ is used to extract the characteristics of the data input. $v_i$ and $h_j$ are binary variables, $v_i \in \{0, 1\}$, $h_j \in \{0, 1\}$. $v = (v_1, v_2, ..., v_n)$,
$h=(h_1, h_2, ..., h_m)$. And the two layers are connected by weight matrix $w$. Reference [26] theoretically proves that RBM can fit any discrete distribution when the number of hidden layer neurons is large enough.

RBM is an energy model [26]. The energy of an RBM is defined as follows:

$$E(v, h|\theta) = -\sum_{i=1}^{n} b_i v_i - \sum_{j=1}^{m} c_j h_j - \sum_{i=1}^{n} \sum_{j=1}^{m} v_i w_{ij} h_j$$  \hspace{1cm} (1)

where $v_i$ and $b_i$ denote the state and bias of the $i$-th neuron of the visible layer, respectively; $h_j$ and $c_j$ denote the state and bias of the $j$-th neuron of the hidden layer, respectively; $w_{ij}$ is the connection weight between the $i$-th neuron of the visible layer and the $j$-th neuron of hidden layer; $\theta$ represents the RBM parameter, $\theta = \{w, b, c\}$.

From function (1), the joint probability distribution of the visible layer and hidden layer is as follows:

$$P(v, h|\theta) = \frac{e^{-E(v, h|\theta)}}{Z(\theta)}$$  \hspace{1cm} (2)

where $Z(\theta) = \sum_v \sum_h e^{-E(v, h|\theta)}$ is the normalization factor (or partition function).

In a practical problems, the focus is on the distribution $P(v|\theta)$ of $v$, that is, the marginal distribution of the joint probability distribution (also known as the likelihood function), expressed as:

$$P(v|\theta) = \sum_h P(v, h|\theta) = \frac{\sum_h e^{-E(v, h|\theta)}}{Z(\theta)}$$  \hspace{1cm} (3)

Considering that there is no connection between neurons in each layer of RBM, the states of neurons in the hidden layer are independent of each other when the state of the visual layer is given. The activation probability of the $j$-th neuron in the hidden layer is as follows:

$$P(h_j=1|v, \theta) = \sigma(-c_j - \sum_{i=1}^{n} v_i w_{ij})$$  \hspace{1cm} (4)

Similarly, when the state of the hidden layer is given, the state of each neuron in the visible layer is also independent of each other. The activation probability of the $i$-th neuron is:

$$P(v_i=1|h, \theta) = \sigma(-b_i - \sum_{j=1}^{m} w_{ij} h_j)$$  \hspace{1cm} (5)

where $\sigma(z)$ is called the activation function. Since ReLU (Rectified Linear Unit) has a better convergence characteristics [27], this paper uses the ReLU function to activate the detection model, and it is defined as follows:

$$\sigma(z)=\max(0, z)$$  \hspace{1cm} (6)

The purpose of RBM training is to obtain a parameter $\theta$ to fit the training sample. Assuming the number of training set samples is $S$, the parameters of RBM can be obtained by maximizing the logarithmic likelihood function of RBM on the training set [28]. The method is as follows:

$$\max L(\theta) = \max \sum_{s=1}^{S} \log P(v^s|h, \theta)$$  \hspace{1cm} (7)

To solve function (7), a stochastic gradient ascending method is usually used [15], and its expression is shown as follows:

$$\frac{\partial L(\theta)}{\partial \theta} = \sum_{s=1}^{S} \left( \left\langle \frac{\partial E(v^s, h, \theta)}{\partial h} \right\rangle_{P_{h|v^s, \theta}} - \left\langle \frac{\partial E(v^s, h, \theta)}{\partial v} \right\rangle_{P_{v|h}} \right)$$  \hspace{1cm} (8)

where $P(h|v^s, \theta)$ represents the probability distribution of the hidden layer when the visible layer is $v^s$, and $\langle \cdot \rangle_P$ represents the mathematical expectation of the distribution $P$. Due to the existence of $Z(\theta)$, equation (8) is hard to solve. To solve this problem, Hinton proposed the Contrast Divergence (CD) algorithm [29], and its principle is as follows:

Firstly, assigning the unlabeled training sample $(x)$ to the visible layer $(v)$ and mapping the visual layer to the hidden layer by equation (4). After one-step Gibbs sample, the binary state of the neurons in the hidden layer are determined. Then, by equation (5), a reconstruction of the visible layer $v_2$ is obtained. Finally, calculating the reconstruction error and adjusting the parameter $\theta$ until the learning of RBM is completed.

### B. DBN TRAINING

DBN is composed of multiple RBMs. The output of the previous RBM is as the input of the next RBM. The structure is shown in Fig.2.

DBN training can be divided into two steps: unsupervised pre-training and supervised fine-tuning. The pre-training of DBN can be regarded as training of multiple RBMs. In the first, taking a training sample $(x)$ in the training set as input, and assigning it to the visual layer of the DBN and following, using the CD algorithm to perform unsupervised training on the first RBM. After completing training, the output of the first RBM is used as the input of the second RBM. The rest RBM can be trained in the same manner until all RBM is trained.

In the pre-training, each RBM only ensures the weights in its layer are optimally mapped to the characteristic vector of the layer, and it cannot guarantee the entire network is optimally mapped to the characteristic vector of the layer. That results in the error of the previous RBM transfer to the
As shown in Fig. 3, the sampled data will be uploaded to the master station of power dispatching system only when the variation of sampled data exceeds the threshold value of the remote terminal. Otherwise it will not be uploaded. In other words, not all features have values when a time section is chosen. For example, Table 1 shows a sampled data of \( I_a \) and \( I_c \) in the same 10s.

|             | \( I_a \) | \( I_c \) |
|-------------|-----------|-----------|
| Time Values |           |           |
| 00:05:00.921| 19.924    | 19.244    |
| 00:05:04.359| 20.060    | 19.040    |
| 00:05:08.199| 19.652    | 18.836    |
| 00:05:09.424| 19.516    | 18.632    |

To solve this problem, according to the transmission principle of 104 protocol, we can read the value of the previous time from the master station to supplement. Taking the sampled data of \( I_a \) and \( I_c \) in Table 1 as an example, after sorting, the results are shown in Table 2.

|             | \( I_a \) | \( I_c \) |
|-------------|-----------|-----------|
| Time Values |           |           |
| 00:05:00.921| 19.924    | ...       |
| 00:05:02.174| 19.924    | 19.244    |
| 00:05:03.461| 19.924    | 19.040    |
| 00:05:04.359| 20.060    | 19.040    |
| 00:05:07.057| 20.060    | 18.836    |
| 00:05:08.199| 19.652    | 18.836    |
| 00:05:09.424| 19.516    | 18.632    |

B. FEATURE DATA SET

The required features of the proposed model and its meaning are shown in Table 3. Since most feeders of distribution network are only equipped with A and C-phase current transformers, the proposed model only considers A and C-phase currents, and does not discuss the B-phase current.

| Symbol      | Quantity       |
|-------------|----------------|
| \( I_a \)   | A-phase current|
| \( I_c \)   | C-phase current|
| \( P \)     | Active power   |
| \( Q \)     | Reactive power |

\[ \cos \varphi \] Power factor

\( T \in [t_a, t_c] \) represents a time windows of single-phase ground fault, where \( t_a \) is the start time of the ground fault. For a ground fault that has ended, \( t_c \) is the end time of fault; for

---

**III. FAULTY FEEDER DETECTION MODEL AND IMPLEMENTATION PROCESS**

A. DATA ACQUISITION AND PROCESSION

The communication between remote terminal and master station of power dispatching system is based on the IEC60870-5-104 (104) protocol [30]. The transmission process is shown in Fig. 3.

**TABLE 1.** sampled data of \( I_a \) and \( I_c \) in the same 10s

| Time Values | \( I_a \) | \( I_c \) |
|-------------|-----------|-----------|
| 00:05:00.921| 19.924    | 19.244    |
| 00:05:04.359| 20.060    | 19.040    |
| 00:05:08.199| 19.652    | 18.836    |
| 00:05:09.424| 19.516    | 18.632    |

**TABLE 2.** Sampled data of \( I_a \) and \( I_c \) after sorted

| Time Values | \( I_a \) | \( I_c \) |
|-------------|-----------|-----------|
| 00:05:00.921| 19.924    | ...       |
| 00:05:02.174| 19.924    | 19.244    |
| 00:05:03.461| 19.924    | 19.040    |
| 00:05:04.359| 20.060    | 19.040    |
| 00:05:07.057| 20.060    | 18.836    |
| 00:05:08.199| 19.652    | 18.836    |
| 00:05:09.424| 19.516    | 18.632    |

**TABLE 3.** features of the proposed model and its meaning

| Symbol      | Quantity       |
|-------------|----------------|
| \( I_a \)   | A-phase current|
| \( I_c \)   | C-phase current|
| \( P \)     | Active power   |
| \( Q \)     | Reactive power |

\[ \cos \varphi \] Power factor

\( T \in [t_a, t_c] \) represents a time windows of single-phase ground fault, where \( t_a \) is the start time of the ground fault. For a ground fault that has ended, \( t_c \) is the end time of fault; for
the one that has not completed, $t_c$ is the data acquisition time. From the feeder operation data stored in the master station, extract the feature parameters data in the time window $T$, and construct a feature data set $X$, which is described as equation (9):

$$X = \left( \begin{array}{c} x^1 \\ x^2 \\ \vdots \\ x^l \\ \vdots \\ x^{l+u} \end{array} \right)^T = \left( \begin{array}{c} x^1, x^2, \ldots, x^l \\ x^1_1, x^2_1, \ldots, x^l_1 \\ \vdots \\ x^1_n, x^2_n, \ldots, x^l_n \end{array} \right)^T$$

where $n$ is the number of distribution feeders, $l$ is the number of unlabeled samples, and $u$ is the number of labeled samples. There are five features per feeder, and $5n$ features for $n$ feeders.

The feeder operation data are normally of different orders of magnitude. Therefore, some feature parameters cannot participate in the training effectively if $X$ is input into the model directly, which will result in the ‘large numbers to eat decimals’ situation. To avoid this problem and speed up the convergence of the algorithm, the min-max normalization method is used to transform the element $x$ in $X$ linearly, and all the data is mapped into the interval $[0, 1]$, as shown in equation (10).

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

where $x'$ represents the normalized value of the parameter, $x$ represents the original value, and $x_{\max}$ and $x_{\min}$ represent the maximum and minimum values of $x$, respectively.

C. LABEL OF DATA SAMPLE

In the faulty feeder detection feature data set $X$, the output $Y$ corresponding to the $u$ labeled data samples can be described as equation (11):

$$Y = \left( \begin{array}{c} y^1 \\ y^2 \\ \vdots \\ y^u \end{array} \right)^T = \left( \begin{array}{c} y^1_1, y^2_1, \ldots, y^u_1 \\ y^1_2, y^2_2, \ldots, y^u_2 \\ \vdots \\ y^1_n, y^2_n, \ldots, y^u_n \end{array} \right)^T$$

where $Y$ represents the label of the sample data, and its value is related to the feeder where the ground fault is located, as shown in Table 4.

| Grounded feeder | $y = (y_1, y_2, \ldots, y_n)$ |
|-----------------|-------------------------------|
| $l_1$           | $(0, 0, \ldots, 0, 0, 1)$    |
| $l_2$           | $(0, 0, \ldots, 0, 1, 0)$    |
| $\ldots$        | $\ldots$                      |
| $l_{n-1}$       | $(0, 1, 0, \ldots, 0, 0)$    |
| $l_n$           | $(1, 0, \ldots, 0, 0, 0)$    |

D. BUILD THE FAULTY FEEDER DETECTION MODEL

The modeling process of the faulty feeder detection model based on DBN is shown in Fig.4. Combining with the faulty feeder detection problem, the proposed method adds a softmax classifier to the output layer to identify the single-phase grounded feeder corresponding to the data sample, and outputs the classification results.

The specific implementation steps are as follows:

Step 1: From the feeder operation data stored in the master station of the power dispatching system, extracting the input feature data of all feeders in the single-phase ground fault time window $T$ and forming a feature parameter data set $X$, and then normalize it.

Step 2: Divide the data set $X$ into a pre-training set and a fine-tuning set in a ratio of 8:2. Among them, the pre-training set is an unlabeled data set, and the tuning set is a labeled data set.

Step 3: Construct a DBN-based faulty feeder detection model. The number of neurons in the DBN input layer is set to $5n$. Multiple experiments determine the amount of hidden layers and the number of neurons in each layer. The form of the softmax classifier in the output layer is determined according to the number of feeders and the data sample label.

Step 4: Pre-training. Set the parameters of the training process, such as learning rate and training period. Use the data sample in the pre-training set to train the RBMs one by one through the CD algorithm, and obtain the connection relationship $w$ and offset values $b$ and $c$ of each network layer of the identification model.

Step 5: Fine-tuning. After the pre-training, the fine-tuning set is used as the input of the pre-trained model. The label matrix $Y$ corresponding to the fine-tuning set is used as the output of the softmax classifier. Then use the BP algorithm to optimize the whole network parameters. After that, save the trained model structure and parameters $\theta = \{w, b, c\}$. 

![Fig.4 Modeling process of faulty feeder detection](image-url)
Step 6: When a single-phase ground fault occurs, read the structure and parameters $\theta$ obtained in step 5, and input the feature data collected in real time into the model for judgment. The output $y$ is the identification result.

Step 7: After verifying the grounded feeder, the collected feature data in step 6 and its correct label are added to the historical fault data set, and the DBN-based faulty feeder detection model is trained again to improve the generalization ability of the model further.

IV. CASE ANALYSIS
A. SYSTEM AND PARAMETERS
Taking an actual substation of Lipu Power Grid, Guangxi, China as the research object. Its wiring diagram is shown in Fig.5. $l_1$, $l_2$, and $l_3$ are the distribution feeders on the bus, $load_1$, $load_2$, and $load_3$ are the feeder loads.

According to the records of single-phase grounded fault occurred in the system from August 25, 2015, to December 31, 2017, obtaining all feature data during faults from the master station of the power dispatching system. After classification and sorting, 2873 groups of valid data sample are obtained. The sample size distribution of each feeder is shown in Table 5.

| Feeder   | Sample size |
|----------|-------------|
| $l_1$    | 722         |
| $l_2$    | 1021        |
| $l_3$    | 1130        |

After randomly sorting respectively, 2600 groups are selected as the training set, and 273 groups as the test set. They are used to test the DBN-based faulty feeder detection model, which is implemented by TensorFlow, Google’s open source deep learning framework.

B. EFFECT OF THE ACTIVATION FUNCTION
The activation function plays a role of character mapping in deep learning model, and the model has different convergence performance under different activation functions. Commonly used activation functions include ReLU, sigmoid (equation (12)), etc.

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

To select an appropriate activation function, sigmoid and ReLU are used as the activation function of the DBN-based faulty feeder detection model. For comparing the training errors, 2000 samples from the training set are randomly selected for testing.

The results are shown in Fig.6.

![Comparison of activation function performance](image)

As can be seen from Fig.6, with the increase of iteration number, the training error decreases rapidly, and reaches the minimum when the number of learning is about 150. After that, it shows a slight upward trend and eventually tends to be stable. In the whole iterative learning process, the error rate trend of ReLU is smaller than sigmoid, which means it has better convergence characteristics.

Based on the analysis above, this paper selects ReLU as the activation function, and the number of iterative learning was 150.

C. INFLUENCE OF TRAINING SAMPLES AND NETWORK STRUCTURE
There is no mature method for the construction of the DBN structure [31]. Theoretically, the more neurons in the hidden layer, the easier it is to mine the raw data characteristics. However, too many neurons can easily lead to over-fitting, which increases the computational cost.

For evaluating the influence of network structure and training samples on the performance of the proposed model, four structures of DBN-based faulty feeder detection model are set. The number of hidden layers and the number of neurons in each layer are 100 (DBN-1), 100-50 (DBN-2), 200-100-50 (DBN-3) and 200-100-50-10 (DBN-4), respectively. Under
a condition that the sample size of the training set is 100, 300, 500, 800, 1000, 1500, 2000, and 2600, respectively, the identification accuracy of the four structures are compared. During the training, the learning rate of pre-training and fine-tuning are both $10^{-3}$. In order to reduce the influence of random factors and improve the credibility of detection results, each experiment was repeated ten times, and then take the average accuracy as the final result (the same as below).

The results are shown in Fig.7.

![Detection accuracy under different DBN structure and different training sets](image)

**FIGURE 7.** Detection accuracy under different DBN structure and different training sets

Fig.7 shows that the accuracy of DBN-based faulty feeder detection model is related to both the network structure and the number of training samples. Under the same training set, as the number of network layers increases, the accuracy of the four DBN structures increases visibly, especially from DBN-1 to DBN-2. With the amount of training samples increases, the accuracy of DBN-2, DBN-3, and DBN-4 increase rapidly. Especially in the process of the sample size from 100 to 300, the accuracy increases significantly and then tends to grow slowly. When the number of training samples exceeds 1500, the accuracy of DBN-4 exceeds 90% and finally reaches 94.70%, which is significantly higher than the shallow structure, and has better fault characteristic extraction ability.

For further compare the stability of different DBN structures, collating the ten experiments results of each DBN structures above, and the results are shown in Fig.8.

According to Fig. 8, benefited from the deep network structure, when the number of training samples exceeds 1500, and the structure is DBN-4, the faulty feeder detection accuracy of each experiment maintains at a high level, and the surface shape tends to be stable, which means the stability of detection result has been strengthened. In these experiments, when the number of a training sample is 2600, the accuracy of the four DBN structures comes to highest. At this point, the accuracy standard deviation of the four structures is shown in Table 6.

As can be seen from Table 6, for the four structures above, when the accuracy reaches the highest, with the deepening of the network structure, the DBN-based detection model has improved the ability to extract the characteristic of the ground fault data, so that the discrete degree of accuracy is gradually reduced and the stability is strengthened.

In practical applications, it is difficult to collect a large number data sample of single-phase ground fault. According to the experimental results above, by deepening structure of the DBN-based faulty feeder detection model, it is possible to obtain a better detection effect under a small number of ground fault samples.

**D. INFLUENCE OF FEATURE PARAMETER COMBINATION**

In the previous research, limited by the performance of data acquisition device and the running time of the algorithm, the faulty feeder detection in neutral non-effectively grounded systems usually separately considers feature parameters such as feeder current, zero-sequence current or power. Therefore, the fault information cannot be comprehensively utilized.

In order to study the influence of different parameter combinations on the result of faulty feeder detection, two feature datasets are reconstructed: the first one contains $I_a$ and $I_c$, a total of 2 features; the other one contains $I_a$, $I_c$, $P$ and $Q$, a total of 4 features. Two datasets above are used to train DBN-4 separately. Then, the results of DBN-4 in Section IV.C are combined to form a comparison of three feature parameter combinations, to study the influence of different feature parameter combinations on the result of faulty feeder detection. The experimental results are shown in Table 7.

| Feature parameter | Accuracy | Growth value |
|-------------------|----------|--------------|
| $I_a$, $I_c$      | 87.62    | + 0          |
| $I_a$, $I_c$, $P$, $Q$ | 90.56    | + 2.94       |
| $I_a$, $I_c$, $P$, $Q$, $\cos \varphi$ | 94.70    | + 7.08       |

According to Table VII, the accuracy reaches 87.62% when only considering the combination of current feature parameters $I_a$, $I_c$. When most of the distribution feeders are only installed with A and C-phase current transformers, the accuracy is good enough in practice. With the addition of $P$, $Q$, $\cos \varphi$, the accuracy has gradually increased, eventually reaching 94.70%. The results above shows that, with deep structure, DBN-4 can still obtain high detection accuracy under a condition of fewer fault feature parameters.
FIGURE 8. Stability of faulty feeder detection under different DBN structure and different training samples

Besides, the addition of power feature parameters $P$, $Q$ and $\cos \varphi$ increases accuracy by 7.08%, which is relatively small when compared with the accuracy of current feature parameters alone. Therefore, when using the feeder operation data collected by power dispatching system for faulty feeder detection, it is critical to obtain high-quality current feature parameter values.

E. PERFORMANCE COMPARISON OF DBN VS. TRADITIONAL METHOD

In a case that the number of training sample and test sample are 2600 and 273, respectively, DBN-based faulty feeder detection model and traditional machine learning methods such as SVM and BP algorithm are compared. Specific contents are as follows:

1) DBN structure is the DBN-4 shown in Section IV.C.
2) Four kernel functions are considered in SVM algorithm, which is polynomial kernel SVM-P (degree = 4), sigmoid kernel SVM-S, radial basis kernel SVM-R (RBF, C = 1, gamma = 3.5) and linear kernel SVM-L (Linear).
3) BP algorithm considers two kinds of hidden layer structures. After traversal search, the structure is set to: single hidden layer 100 (BP-1) and multiple hidden layers 200-100-50-10 (BP-4).

The results are shown in Fig.9.

As shown in Fig.9, based on the same data set, the DBN-4 has the best performance with an accuracy of 94.70%, which is 4.96% higher than SVM-R and 22.17% higher than single hidden layer BP-1. These show that the deep network structure of DBN can mine and express the single-phase ground fault data more essentially and has better characteristic learning ability. While the generalization ability of traditional shallow machine learning algorithms such as SVM and BP is relatively limited when dealing with ground fault data.

When the number of hidden layers and the number of
neurons are the same, DBN-4 has a higher accuracy than BP-4 neural network by 26.03%. These indicate that the unsupervised greedy algorithm used by DBN can more effectively pre-train the network parameters and provide better initial values for the subsequent training process, thereby improving the faulty feeder detection accuracy of the model while the multi-hidden BP neural network uses the randomization method to initialize the model parameters and trains it directly. As its parameters are not effectively optimized, resulting in low accuracy of faulty feeder detection.

V. APPLICATION EFFECT
A faulty feeder detection system based on the proposed method has been implemented in Lipu Power Grid, Guangxi, China. Since May 2018, four single-phase grounding faults occurred successively in the monitored substation. The detection system has captured all the four ground faults and successfully judged the ground feeder, which significantly improves the reliability of the power supply.

Fig.10 shows the change of bus voltage and feeder current of a single-phase ground fault occurred at 07:47:01.258, June 3, 2018. At time $t_1$, the on-site detection device detected the ground fault. Then the staffs cut off feeders $l_1$, $l_2$ and $l_3$ at times $t_2$, $t_3$, and $t_4$, separately. When $l_3$ is cut off, the bus voltage returns to normal, which means $l_3$ is the faulty feeder. In this period, the duration of single-phase ground fault is about 14 minutes, which seriously affects the reliability of the power supply.

For these single-phase ground faults with more complicated grounding process, after obtaining the feature data within 20s from the master station of the power dispatching system after grounded, the DBN-based faulty feeder detection system can still accurately determine the ground feeder. This effectively improves the efficiency of faulty feeder detection in a distribution network, reduce the outage time and improve the reliability of power supply.

VI. CONCLUSION
This paper proposes a faulty feeder detection uses deep belief network to identify the faulty feeder of single-phase ground fault in neutral non-effectively grounded systems. With an actual power dispatching system, feeder operation data of a 10kV ungrounded system is extracted and experiments are carried out. The experiment results are as follows:
1) Using feeder operation data collected by power dispatching system to train a DBN-based faulty feeder detection model with a deep network structure, good faulty feeder detection performance can be obtained.

2) The accuracy of the proposed model is related to the number of hidden layers and training samples. Within a certain range, the increase in the number of layers and the increase in the number of training samples can improve the identification accuracy.

3) Compared with traditional machine learning algorithms such as SVM and BP, the training method using DBN is more effectively by fault data characteristics and improve detection accuracy.

The proposed method has been put into operation on line in Lipu Power Grid, Guangxi, China. The actual operation effect is well, and the result of faulty feeder detection is reliable, which proves that it has a broad application prospect.

REFERENCES
[1] X. Dong and S. Shi, “Identifying single-phase-to-ground fault feeder in neutral noneffectively grounded distribution system using wavelet transform,” IEEE Trans. Power Del., vol. 23, no. 4, pp. 1829–1837, 2008.
[2] C. S. Mardegan and R. Rifat, “Insights into applications of ieee standards for ground-fault protection in industrial and commercial power systems,” IEEE Trans. Ind. Appl., vol. 51, no. 4, pp. 2854–2861, 2015.
[3] X. Wang, J. Gao, M. Chen, X. Wei, Y. Wei, and Z. Zeng, “Faulty line detection method based on optimized bistable system for distribution network,” IEEE Trans. Ind. Informat., vol. 14, no. 4, pp. 1370–1381, 2018.
[4] Y. Wang, Y. Huang, X. Zeng, G. Wei, J. Zhou, T. Fang, and H. Chen, “Faulty feeder detection of single-phase-earth fault using grey relation degree in resonant grounding system,” IEEE Trans. Power Del., vol. 32, no. 1, pp. 55–61, 2017.
[5] S. Jamali and A. Bahmanyar, “A new fault location method for distribution networks using sparse measurements,” International Journal of Electrical Power & Energy Systems, vol. 81, pp. 459–468, 2016.
[6] L. Cheng and Q. Chen, “A survey on faulty line selection technology for single-phase grounded transmission line in small current neutral grounded system,” Power System Technology, vol. 33, no. 18, pp. 219–224, 2009.
[7] M. Guo, X. Zeng, D. Chen, and N. Yang, “Deep-learning-based earth fault detection using continuous wavelet transform and convolutional neural network in resonant grounding distribution systems,” IEEE Sensors J., vol. 18, no. 3, pp. 1291–1300, 2017.
[8] N. I. Elkalashy, A. M. Elhaffar, T. A. Kawady, N. G. Tarhuni, and M. Lehtonen, “Bayesian selectivity technique for earth fault protection in neutral-voltage networks,” IEEE Trans. Power Del., vol. 25, no. 4, pp. 2234–2245, 2010.
[9] B. Jiang, X. Dong, S. Shi et al., “A method of single phase to ground fault feeder selection based on single phase current traveling wave for distribution lines,” Proceedings of the CSEE, vol. 34, no. 34, pp. 6216–6227, 2014.
[10] Q. Zhang, Y. Zhao, X. Han, and Y. Sun, “An improved fault line selection method based on traveling waves in non-solid earthed network,” in 2017 Chinese Automation Congress (CAC). IEEE, 2017, pp. 4553–4557.
[11] H. Livani and C. Y. Evrenosoglu, “A machine learning and wavelet-based fault location method for hybrid transmission lines,” IEEE Trans. Smart Grid, vol. 5, no. 1, pp. 51–59, 2014.
[12] H. Livani, C. Y. Evrenosoglu, and V. A. Centeno, “A machine learning-based faulty line identification for smart distribution network,” in 2013 North American Power Symposium (NAPS). IEEE, 2013, pp. 1–5.
[13] Y. Lei, F. Jia, J. Lin, S. Xing, and X. Xing, “An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data,” IEEE Trans. Ind. Electron., vol. 63, no. 5, pp. 3137–3147, 2016.
[14] B. Li and H. Shi, “A new integration method of fault line detection in resonant earthed system based on instantaneous real power and instantaneous virtual power,” Transactions of China Electrotechnical Society, vol. 6, 2012.
[15] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” nature, vol. 521, no. 7553, p. 436, 2015.
[16] G. E. Hinton, “A practical guide to training restricted boltzmann machines,” in Neural Networks: Tricks of the trade. Springer, 2012, pp. 599–619.

HONGBO WEI received the B.S. and M.S. degrees in electrical engineering from Guangxi University, Nanning, China, in 2015 and 2018, respectively. He is currently working as an electrical engineer in Power Dispatch and Control Center of Guangxi Power Grid Co., Ltd., China Southern Power Grid. His research interests include power dispatch automation and the application of artificial intelligence in power system.

HUA WEI received the B.S. and M.S. degrees in power engineering from Guangxi University, Nanning, China, in 1981 and 1987, respectively, and the Ph.D. degree in power engineering from Tokyo Institute of Technology, Tokyo, Japan, in 2002. He is currently a Professor with Guangxi University. He is also the Director with the Institute of Power System Optimization, Guangxi University. His research interests include power system operation and planning, particularly in the application of optimization theory and cloud technology to power systems.
ZHONGLIANG LYU received the B.S. degree in agricultural electrification and automation from Nanjing Agricultural University, Nanjing, China, in 2015. He is currently working toward the Ph.D. degree in power engineering with Guangxi University, Nanning, China.

His research interest includes the application of optimization methods and cloud technologies to power systems.

XIAOQING BAI received the B.S. and Ph.D. degrees in power engineering from Guangxi University, Nanning, China, in 1991 and 2010, respectively. She is currently a Professor with the Institute of Power System Optimization, Guangxi University.

Her research interest includes power system optimization based on SDP and robust optimization.

JUNYANG TIAN received the B.S. degree in electrical engineering from Hunan University, Changsha, China, in 2004, and the Ph.D. degree in power engineering from Guangxi University, Nanning, China, in 2013. He is currently working as a technical expert in Power Dispatch and Control Center of Guangxi Power Grid Co., Ltd., China Southern Power Grid.

His research interest includes power system fault diagnosis and optimization.