

Research Article

Deep-Learning-Assisted Topology Identification and Sensor Placement for Active Distribution Network

Juhua Hong ¹, Linyao Zhang ¹, Yufei Yan ², Zeqi Wang ² and Pengzhe Ren ²

¹State Grid Fujian Institute of Economics and Research, Fuzhou 350109, China

²College of Electrical Engineering, Sichuan University, Chengdu 610065, China

Correspondence should be addressed to Pengzhe Ren; rpz@stu.scu.edu.cn

Received 8 September 2021; Revised 26 September 2021; Accepted 29 September 2021; Published 19 October 2021

Academic Editor: Qiuye Sun

Copyright © 2021 Juhua Hong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In response to the demand for identification of distribution network topology with a high percentage of renewable energy penetration, a distribution network topology analysis method based on decision trees and deep learning methods is proposed. First, the decision tree model is constructed to analyze the importance of each node's characteristics to the observability of the distribution network topology. Next, we arrange the node feature importance from large to small and select the node measurement data with high importance as the training sample set. Then, the principal component analysis (PCA)-deep belief network (DBN) model is used to analyze the changes in the observability of the distribution network topology, and the nodes are selected as the optimal location for the measurement device when the distribution network is completely observable. Finally, the IEEE-33 bus system with a high proportion of renewable energy is used to verify that the method proposed has a good effect in the identification of the distribution network topology.

1. Introduction

With the large-scale grid connection of renewable energy and the wide application of power electronic devices, the operation of the power system becomes more complex and changeable. Meanwhile, as the uncertainty of the system increases, security issues become increasingly prominent [1]. The topology identification for distribution network is the basis for other advanced application analysis such as state estimation, power flow analysis, and demand response of the distribution network. With the gradual increase in the proportion of renewable energy in the power grid [2], the application of technologies such as distributed power generation [3], controllable loads [4], electric vehicles [5], energy storage access [6], and demand response [7] has caused the power flow of the distribution network to become bidirectional, which increases the difficulty of topology identifying for distribution network. With the large-scale integration of renewable energy into the distribution network, the uncertainty of grid operation has increased, leading to more frequent topology reconstruction. For

example, in distribution network with a large-scale distributed photovoltaic power, the network topology may change every 8 hours [8], which increases the difficulty of obtaining real-time and accurate network topology information. Furthermore, the integration of renewable energy has significantly increased the complexity of the dynamic behavior of the power system, which brings challenges to accurate topology identification of the distribution network.

Topology identification methods of distribution network mainly include matrix method, graph theory method and data-driven method. The core idea of the matrix method [9, 10] is to use the node-branch incidence matrix or adjacency matrix to represent the branch commissioning state and switching state of the power grid, and then expand the distribution network topology analysis and identification. The advantage of the matrix method is that the algorithm is intuitive and the principle is simple. However, it has a large amount of calculation and slow calculation speed, which cannot meet the requirements of real-time monitoring of the topology status. Hence, it is usually only suitable for small-scale power grids. The core idea of graph theory [11, 12] is to

transform the power grid into a graph model. When the information of the distribution network admittance matrix is not fully known, this method cannot detect the topology changes of the actual distribution network. Therefore, it is difficult to use in an actual system. The data-driven method is based on historical measurement data or data collected by smart meters and uses statistical methods to analyze the mapping relationship between the measured data and the topology. Lioa et al. [13] used the group lasso algorithm to calculate the regression coefficients between nodes and then judged the connectivity between nodes to achieve the effect of topology identification. Ahmad et al. [14] stored various network topologies in the model library and used recursive Bayesian method to analyze the state estimation results of each model and obtained the real-time topology. In Mestav et al.'s study [15], the node injected power is used as the input data to train the deep neural network and then to evaluate the state of the distribution network. Oliveira et al. [16] proposed a probabilistic graph model that can mine the topology information of the distribution network with distributed generation from a large amount of voltage data. In Duan and Stewart's study [17], switch action recognition is defined as a multilabel classification problem. Researchers use distributed energy data with dynamic characteristics, such as time series data such as three-phase frequency and voltage amplitude, to train the convolutional neural network to recognize the system's multiphase and multiswitch actions.

As the available energy sources are connected to the distribution network, the uncertainty and randomness of the system increase. Therefore, it is necessary to configure corresponding power distribution automation equipment, such as advanced metering infrastructure (AMI) and phasor measurement unit (PMU). However, due to the high cost of distribution automation measurement devices, in the actual distribution network, measurement devices are not installed at all nodes. Therefore, the configuration principle of the measurement device of the distribution network is to achieve the power grid topology identification with the smallest number of measurement devices. To optimize the distribution network topology measurement device placement, Singh et al. [18] adopted the theory of order optimization and the method of accurate probability. The model takes the minimum relative error value of the system voltage and phase angle as the objective function to calculate the optimal configuration scheme of the measuring device. Zhao et al. [19] proposed a PMU placement optimization method for distribution network state estimation. This method uses binary integer linear programming model to optimize the configuration of measurement devices. Mabaning et al. [20] aimed to minimize the number of measurement devices installed and used greedy algorithm to solve the problem of optimal configuration of measurement devices for distribution network. Existing research has achieved certain results for the identification of distribution network topology and the optimization of distribution network topology measurement points, but it does not consider the impact of the high proportion of renewable energy penetration on the configuration plan. Although Zhao et al. [21] considered the distribution network topology identification of renewable energy, the

performance of the proposed method is significantly reduced when the renewable energy penetration rate exceeds 40%.

A topology analysis method is proposed for distribution network with high proportion of renewable energy based on decision tree and deep learning. First, a method for selecting measurement feature attributes of distribution network based on a decision tree is proposed. It can screen the measurement feature and analyze the decision tree. Compared with the data-driven method based on all node measurement data, this method can effectively quantify and analyze the importance of nodes, resulting in the reduction of the measurement points and the measurement data. Then, a distribution network topology analysis method based on the principal component analysis (PCA) and deep belief network (DBN) coupling model is proposed. The PCA is used to select superior robust features and remove redundant measured data. The measured data of active distribution network nodes have nonlinear and nonsequential characteristics. Compared with neural networks such as recurrent neural network (RNN) and long short term memory (LSTM), DBN used is a better one for processing the nonsequential sample data. In the DBN model, the measured data of the distribution network nodes are used as input, and the branch connection status is used as the output. DBN is used to map the voltage amplitude of the distribution network nodes and the connection status of the switches, so as to realize the identification of the distribution network topology.

The main contribution of this paper can be summarized as follows:

- (1) This paper studies the distribution network topology analysis method with a high proportion of renewable energy access and analyzes the performance of the proposed method under different renewable energy penetration rates. The results show that when the renewable energy penetration rate reaches 50%, the topology identification has good accuracy.
- (2) Propose a method for selecting characteristic attributes of distribution network measurement based on decision tree. By calculating the importance of node features and quantifying the importance of each node feature to topology analysis, the number of measurement devices in the distribution network can be effectively reduced while ensuring that the topology can be identified.

The following paper can be organized as follows: the basic topology theory is presented in Section 2; the details of the data-driven topology identification method are presented in Section 3; Section 4 shows the process of the algorithm implementation. Several cases are carried out in Section 5, and Section 6 concludes this paper.

2. Topology Theory for Distribution Network with Renewable Energy Connection

2.1. Topology Model and Its Representation Method. The structure of the distribution network is generally radial or ring shaped. Similar to the general distribution network, the structure of the distribution network with a high proportion

of renewable energy access is also a closed-loop design and open-loop operation. In addition, the distribution network has many equipment components and poor objectivity, which makes it inconvenient and inconvenient to directly perform topological analysis, state estimation, and other tasks. To perform topological analysis quickly and efficiently, it is necessary to simplify the topological structure of the distribution network with a high proportion of renewable energy access. Generally, the physical topology of the distribution network is simplified to a topological graph model composed of points and lines. Specifically, substations, distribution transformers, renewable energy power generation equipment, and loads are abstracted as nodes, while branch feeders, disconnect switches, and tie switches are abstracted as lines. The topological graph model is defined as $G = \{V, \Psi\}$, where $V = \{1, 2, \dots, n\}$ represents the set of nodes and $\Psi = \{x_{ij}, i, j \in V\}$ represents the set of edges between nodes.

The topological graph model of distribution network needs to be expressed mathematically to facilitate calculation and analysis. The connection relationship of the topological graph can be expressed algebraically through a matrix. For a distribution network diagram model $G = \{V, \Psi\}$ with a relatively fixed structure, the order of its adjacency matrix A_{adj} is constant, and the properties of its matrix elements are as follows:

$$a_{ij} = \begin{cases} 1, & x_{ij} \in \varepsilon, \\ 0, & \text{others.} \end{cases} \quad (1)$$

Equation (1) represents that when the nodes and are connected, the value of the corresponding element is 1, otherwise it is 0. The adjacency matrix has the following properties:

- (1) The state variable, whose matrix element value is 0/1, represents the state of the segment switch and the contact switch
- (2) For undirected graphs, the matrix is a symmetric square matrix symmetrical about the main diagonal
- (3) The diagonal elements of the matrix are all 0

2.2. Topology Identification Problem Description and Its Solution. The core of the distribution network topology identification problem is to accurately and in real-time judge the operational status of the bus and distribution transformers in the distribution network, as well as the connection status of the section switch and the tie switch. From the perspective of matrix algebra, it is to determine the value of each position element in the matrix. As shown in Figure 1, the topological structure of the two different operation modes of the same structure of the distribution network is also different. Their adjacency matrix is expressed as $A_{\text{adj}1}$ and $A_{\text{adj}2}$, respectively. Topology identification is to accurately obtain whether each element of the adjacency matrix is 0 or 1 in the current operation. When each element of the adjacency matrix is determined, it can be mapped to obtain the real-time operating topology of the distribution network.

However, due to the lack of real-time remote signaling or inaccurate remote signaling information in the distribution network, the mis-operation and maintenance of some switches will cause the distribution network topology change information to not be updated in time, thereby affecting the real-time topology identification results. The access of renewable generation has exacerbated this problem. The volatility and uncertainty of the energy output of renewable energy make the operation mode of the active distribution network flexible and changeable, leading to more frequent topology changes. However, the budget for building real-time measurement devices in the distribution network is limited, and the measurement devices are often not enough to cover the entire distribution network. This increases the difficulty of real-time topology identification of the distribution network.

Therefore, the machine learning method is chosen to replace the traditional topology identification model. By mining the relationship between the distribution network operation data and the distribution network topology, an accurate and efficient online identification method of the distribution network topology is proposed. Based on the aforementioned analysis, the problem of distribution network topology identification can be attributed to the problem of using the machine learning model to map the measurement data of the distribution network and the connection status of the tie switch and the section switch under different operation modes of the distribution network.

2.3. Optimal Deployment of Measurement Devices. The observability of the distribution network refers to the estimation and determination of the distribution network topology and operating status through sufficient measurement data. From the perspective of the distribution network algebraic model, if the adjacency matrix and the Jacobian matrix of the distribution network can be uniquely determined, and both the adjacency matrix and the Jacobian matrix are full rank matrices, the distribution network is completely observable. From the perspective of the distribution network graph model, for a distribution network $G = \{V, \Psi\}$, where V is the set of all nodes and Ψ is the set of edges between nodes, the distribution network graph model obtained by measuring the network is $\widehat{G} = \{\widehat{V}, \widehat{\Psi}\}$. If the two sets of node information in the measured graph model belong to the real model, namely $\widehat{V} \in V$ and $\widehat{\Psi} \in \Psi$, the graph model of the measured network can cover the graph model of the actual network. Thus, the distribution network system can be treated as observable.

Based on the observability requirements of the distribution network topology [20], the configuration principle of the distribution network measurement device is shown in Figure 2.

According to the configuration principle of the distribution network measuring device, to solve the problem of the optimal location of the measuring device and the minimum number of configurations is to solve the problem of the observability of the distribution network. The objective function of the mathematical model to solve this problem is as follows:

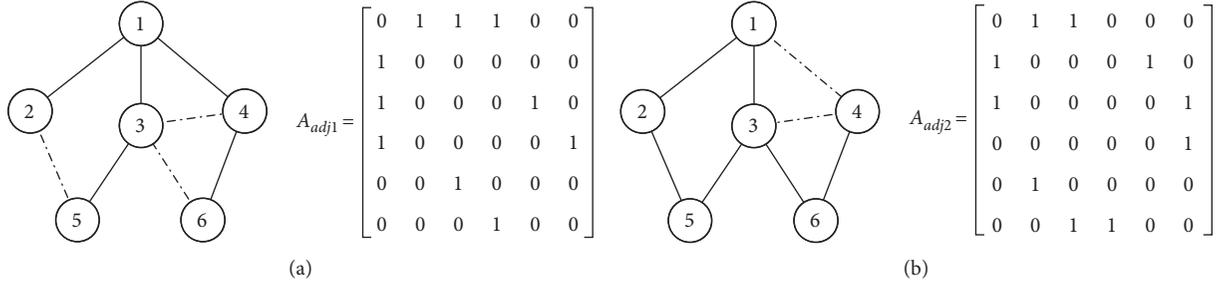


FIGURE 1: Topological structure of distribution network under different operation modes.

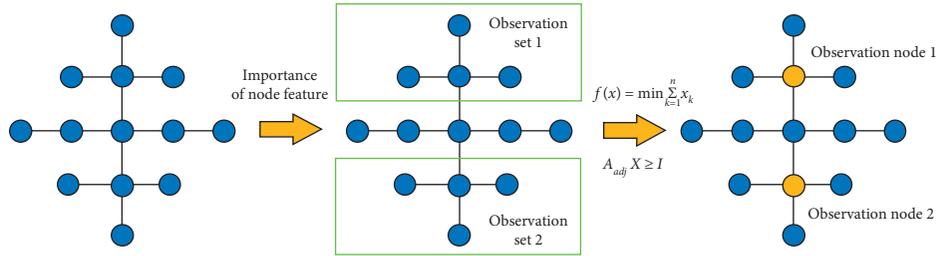


FIGURE 2: The schematic diagram of optimal deployment for measuring device.

$$f(x) = \min \sum_{k=1}^n x_k, \quad (2)$$

where x_k represents the configuration attribute of the node. If the node is equipped with a measuring device, then $x_k = 1$, otherwise $x_k = 0$. The specific formula is as follows:

$$x_k = \begin{cases} 1, & \text{equip,} \\ 0, & \text{not equip,} \end{cases} \quad (3)$$

$$A_{adj} X \geq I, \quad (4)$$

where $X = [x_1, x_2, \dots, x_n]^T$ is the attribute matrix of the distribution network node, which is an n -dimensional column vector, indicating whether the node is equipped with a measuring device. The matrix A_{adj} is the adjacency matrix of the distribution network, and its value is shown in equation (1). The matrix I is an n -dimensional column vector with all 1 elements.

3. Topology Identification Method Based on Machine Learning

3.1. Feature Attribute Selection. Decision tree algorithm is a commonly used supervised machine learning algorithm used to solve regression problems and classification problems. Decision tree learning algorithm includes two parts: feature selection and decision tree generation. The important features of the sample set classification can be obtained through feature selection. Indexes such as information gain, information gain ratio and Gini coefficient are usually used for feature selection. X and Y are respectively input and output variable in the training set $T = \{(x_1, y_1), (x_2, y_2), \dots,$

$(x_n, y_n)\}$. Suppose the training set samples have K categories, the number of samples in the K th category is C_K . To build a decision tree, the sample set needs to be divided first. Usually select the j th feature x_j of the node training set T and its value s as the segmentation variable and partition node of the sample, and divide the sample set into two subdata sets T_{s1} and T_{s2} . The division principle is as follows:

$$\begin{cases} T_{s1}(j, s) = \{x | x_j \leq s\}, \\ T_{s2}(j, s) = \{x | x_j > s\}. \end{cases} \quad (5)$$

For the sample set T , the information gain of the sample feature A is $g(T, A)$, shown as follows:

$$g(T, A) = H(T) - H(T|A), \quad (6)$$

$$H(T) = - \sum_{k=1}^K \frac{|C_K|}{|T|} \log_2 \frac{|C_K|}{|T|}, \quad (7)$$

$$H(T|A) = \sum_{i=1}^n \frac{|T_i|}{|T|} H(T_i), \quad (8)$$

where $H(T)$ is the information entropy of the data set, $H(T_i)$ is the information entropy of the data set T_i , $H(T|A)$ is the information entropy of the data set T to the feature A , n is the number of values of the feature A , and T_i is the i th sample set of the feature A in the data set T .

For the sample set T , the information gain ratio of the sample feature A is $g_R(T, A)$, shown as

$$g_R(T, A) = \frac{g(T, A)}{H(T)}. \quad (9)$$

The Gini coefficient expression of the sample T is

$$\text{Gini}(T) = 1 - \sum_{k=1}^K \left(\frac{C_k}{T} \right)^2. \quad (10)$$

Selecting features with the goal of minimizing the Gini coefficient. The size of the Gini coefficient represents the impurity of the model. The smaller the Gini coefficient, the better the features selected by the model. The decision tree model starts from the root node, through training the data set obtained by minimizing the Gini coefficient, recursively build a binary decision tree. The establishment process of the decision tree model is as follows:

- (1) Initialize training sample T and Gini coefficient threshold g_{ini} .
- (2) Calculate the feature value and loss function partition nodes for sample nodes. The loss function is the mean square error of the two subdata sets T_{s1} and T_{s2} after the division. Select the feature segmentation variable x_j and the partition node s corresponding to the minimum loss function. The calculation formula is as follows:

$$\min \left[\min_{c_1} \sum_{x_i \in T_{s1}(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in T_{s2}(j,s)} (y_i - c_2)^2 \right], \quad (11)$$

where c_1 and c_2 represents the average output value for subdata sets T_{s1} and T_{s2} , respectively.

$$c_l = \frac{1}{n} \sum_{x_i \in T_{s2}(j,s)} y_i, \quad l = 1, 2. \quad (12)$$

Through Step (2), the input training data can be divided into two subset T_{s1} and T_{s2} .

- (3) Calculate the Gini coefficients of two divided nodes. If the Gini coefficients are smaller than the threshold g_{ini} , stop node recursive partition; otherwise, return to Step (2).
- (4) Finally, the input sample set is divided into K subspaces, each subspace contains part of the sample data and the average value of the subspace output value c_K . The mathematic model of decision tree can be written as

$$f(x) = \sum_{k=1}^K C_k I(x \in T_{sK}), \quad (13)$$

where $I(x \in T_{sK})$ represents model indicator function.

According to the proposed decision tree model, the sample feature importance index is calculated. First, for a sample set that contains multiple features, the decision tree can calculate the division standard value for each feature in the sample. Then, the division standard value is used as an index for calculating the importance of each feature. The feature importance can be used to determine the contribution of each feature in the sample to the target variable.

The feature importance calculation method is shown as follows:

$$I_F = \frac{N_t}{N} \times \left(H - \frac{N_{tR}}{N_t} \times H_{\text{right}} - \frac{N_{tL}}{N_t} \times H_{\text{left}} \right), \quad (14)$$

where N and N_t represent the total number of samples and the number of samples of the current node respectively and N_{tR} and N_{tL} represent the number of samples of the left subtree of the current node and the left subtree of the current node, respectively. H represents the impurity of the current node, and the impurity is the sum of squared errors of the node samples. The calculation of impurity is shown as follows:

$$H_{T_s} = \sum_{x_i \in T_s(j,s)} (y_i - c_1)^2. \quad (15)$$

H_{right} and H_{left} represent the impurity of the right subtree of the current node and the left subtree of the current node respectively. According to equation (15), it can be seen that the lower the impurity of the node, the higher the feature importance of the node.

3.2. Topological Observability Analysis Method Based on Deep Learning Model. For large-scale distribution network systems, the samples are characterized by large numbers and high dimensions. Therefore, directly using them for training will significantly increase the complexity and time of training. In addition, data noise will also bring bias to the trained model. In fact, the neural network only needs some key characteristics of the sample data to get good results. The proposed PCA-DBN coupling model uses principal component analysis to select superior robust features without sacrificing the accuracy of the algorithm and to remove redundant information about the voltage amplitude. Then, the selected robust features are supplied to the deep belief network for training and learning, which improves the learning efficiency of DBN, reduces the influence of noise on the training model, and enhances the robustness and compatibility of the model.

The bottom layer of the BN model uses a multilayer restricted Boltzmann machine (RBM) structure. The greedy algorithm is used to train and learn the sample data layer by layer. The parameters obtained by training the first layer of RBM are used as the input of the second layer of RBM, and the parameters of each layer are obtained by analogy. This process belongs to unsupervised learning. The top layer of the model uses back propagation (BP) neural network to fit and optimize the prediction results. The abstract features learned from the bottom model are used as the input of the top-level BP neural network. The prediction result is output through the fitting of BP neural network. At the same time, the BP algorithm needs to be used to finetune and optimize the obtained model parameters. This process is supervised learning. The joint probability of the visible layer and the hidden layer in RBM can be expressed as

$$E(v, h | \theta) = - \sum_{ij} w_{ij} v_i h_j - \sum_i a_i v_i - \sum_j b_j h_j, \quad (16)$$

where v_i is the state of the visible layer node, h_j is the state of the hidden layer node; and a_i and b_i respectively represent the bias values corresponding to the visible layer node and the hidden layer node. w_{ij} represents the connection weight between the visible layer and the hidden layer. According to the aforementioned formula, the joint probability density of the visible layer and the hidden layer can be obtained as

$$P(v, h | \theta) = \frac{1}{Z(\theta)} \exp(-E(v, h | \theta)). \quad (17)$$

In the process of unsupervised learning, the purpose of training RBM is to obtain model parameters θ , which can be obtained through the loglikelihood function:

$$L(\theta) = \sum_{n=1}^N \ln P(v^n, h), \quad (18)$$

$$\theta^* = \arg \max L(\theta) = \arg \max \sum_{n=1}^p \log P(v^n | \theta). \quad (19)$$

According to a first-order Markov process, the following equations can be expressed as

$$P(\Theta) = P(\theta_1, \theta_2, \dots, \theta_n) = P(\theta_1)P(\theta_1 | \theta_2) \cdots P(\theta_n | \theta_{n-1}), \quad (20)$$

$$P(V | \Theta) = P((v_1, v_2, \dots, v_n) | (\theta_1, \theta_2, \dots, \theta_n)) = P(v_1 | \theta_1)P(v_2 | \theta_2) \cdots P(v_n | \theta_n), \quad (21)$$

where p is the number of the main components.

4. Algorithm Implementation

A distribution network topology analysis method driven by decision trees and deep learning models is proposed. The model and its application framework are shown in Figure 3.

The specific steps of the proposed topology identification method:

- (1) Clarify the sample data T and sample Gini coefficient threshold g_{ini} of the distribution network operation status and initialize the model parameters of the decision tree model and the PCA-DBN coupling model.
- (2) Build a decision tree model based on equations (10) to (13) to select the measurement characteristics of the distribution network.
- (3) Based on the results of the supervised learning of the decision tree model, the importance of each feature of the sample set can be calculated by equations (14) and (15), and the sum of the importance of each feature of the node is regarded as the importance of the node.
- (4) Arrange the importance of nodes in descending order, and select the measurement data of the top n nodes of node importance to form a training sample set.
- (5) Input the training sample set into the PCA-DBN coupling topology identification model to analyze the observability of the distribution network topology. If the distribution network topology is completely observable, the first n nodes selected are the best measuring device placement locations; otherwise, return to Step (4) and select the measurement data of the first $n+1$ nodes with the largest node importance to form the training sample set.

- (6) Obtain the optimal location of the distribution network measuring device, and assist the distribution network to identify the real-time topology.

5. Results and Discussion

Based on the modified IEEE-33 bus system, as shown in Figure 4, several case studies are tested to prove the effectiveness of the proposed method. There are 12 normally open switch and 5 normally closed switch used for topology reconfiguration. The total load is 3.715 MW and the types of loads are set as ZIP models. The proportions for the constant power load, constant impedance load, and constant current load are 50%, 30%, and 20%, respectively. Three photovoltaic power plants with rated capacity of 0.5 MW are placed at nodes 17, 21, and 32.

The feasible topology is generated by Monte Carlo method, and the structure constraints of distribution network are considered [22]. The load fluctuations, ranged from 0.8 to 1.2, are generated by Latin hypercube sampling. The node voltage measurement under different combinations of topologies and load are obtained by power flow calculation. Therefore, the initial sample contains the node voltage amplitudes and phase angles from all nodes and the statuses of switches. About 80% of samples are used for training, 10% of samples are used for validation, and the rest are used for test.

5.1. Analysis of Topology Identification Results. According to the proposed decision tree optimization model, the importance of the distribution network node feature is analyzed, as shown in Figure 5. The abscissa is the node importance. It can be learned that the feature importance of different nodes is quite different. For example, the voltage amplitude and phase angle importance of the 0th node are both 0. This is because node 0 is a balanced node, and its voltage amplitude and phase angle are unchanged; in

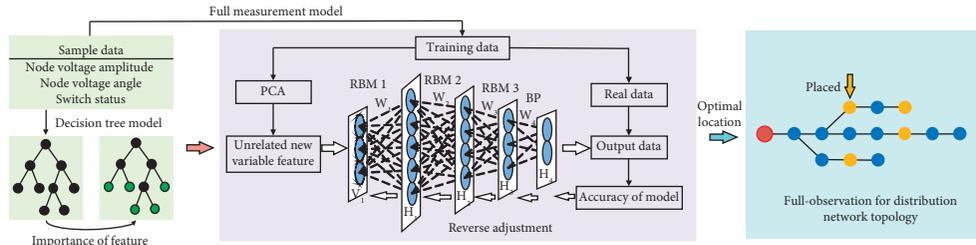


FIGURE 3: Flow chart of optimization model for observability analysis of distribution network topology.

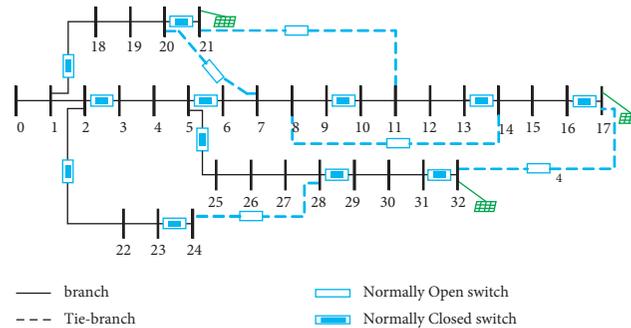


FIGURE 4: Modified IEEE-33 bus system.

addition, the voltage amplitude feature and voltage phase angle feature of each node are not the same in importance. For example, the voltage amplitude importance of node 7 is less than the phase angle importance, while node 18 is the opposite.

The proposed method is used to calculate the accuracy results of the observability of the distribution network topology with different number of measuring devices. The measurement data corresponding to the number of points of the measuring device are selected as the training sample set. Taking the produced training sample set as input, expand the training of the PCA-DBN coupling topology identification model, and calculate its topology identification accuracy. It should be noted that the training parameters and structure parameters of DBN are set by manual tuning, and the optimal training parameters are shown in Table 1. To further verify the calculation effect of the proposed method, the accuracy of topology identification is compared with other machine learning methods, such as random forest (RF), multioutput regression (MOR), and DBN. The comparison results of the accuracy of the topology identification of the four methods are shown in Figure 6.

Considering the number of measurement devices in the distribution network and the observability of the distribution network topology, the top 11 nodes with the highest node importance are selected as the best measurement device locations in the IEEE-33 bus system. The selected node set is {7, 21, 18, 17, 24, 4, 3, 6, 9, 11, 32}.

5.2. Analysis of Topology Identification Accuracy versus Renewable Energy Penetration Rates. The performance of the algorithm under different photovoltaic penetration rates is analyzed. The total capacity of the photovoltaic power plants

is generated by 10%~50% of the total load in IEEE-33 bus system and allocated to nodes 17, 21, and 32 evenly.

Figure 7 shows the topology identification accuracy of four methods versus renewable energy penetration rates. To simulate the practical engineering environment [23], the voltage measurement error of 1% is added. It can be concluded that as the penetration rate of DERs increases, the proposed method has better performance than other machine learning algorithms. Moreover, when the penetration of renewable energy reaches 50%, the accuracy of topology identification is still as high as 98%. The performance of the DBN algorithm is slightly worse than that of the PCA-DBN algorithm. However, it has obvious advantages compared to MOR and RF. Therefore, it is proved that DBN is more suitable for solving this kind of nonlinear and nonsequential problems.

5.3. Robust Analysis of Topology Identification Accuracy.

The noise in measurements will adversely affect the accuracy of distribution network topology identification [24]. In addition, the differences of ZIP model proportions also have varying degrees of impact on topology identification accuracy, as the proportions of the ZIP model will affect the voltage distribution. Therefore, it is necessary to explore the robustness of the proposed topology identification. According to the meter standard [23], the error of measurements is 0.1%~10%. The performance of different topology identification algorithms under different errors is shown in Table 2.

It can be observed that the proposed PCA-DBN maintains a high accuracy with the error range is 0.1%~5%. The accuracy drops about 4% when the error increases from 0.1% to 10%. Therefore, the proposed algorithm is robust to measurement error.

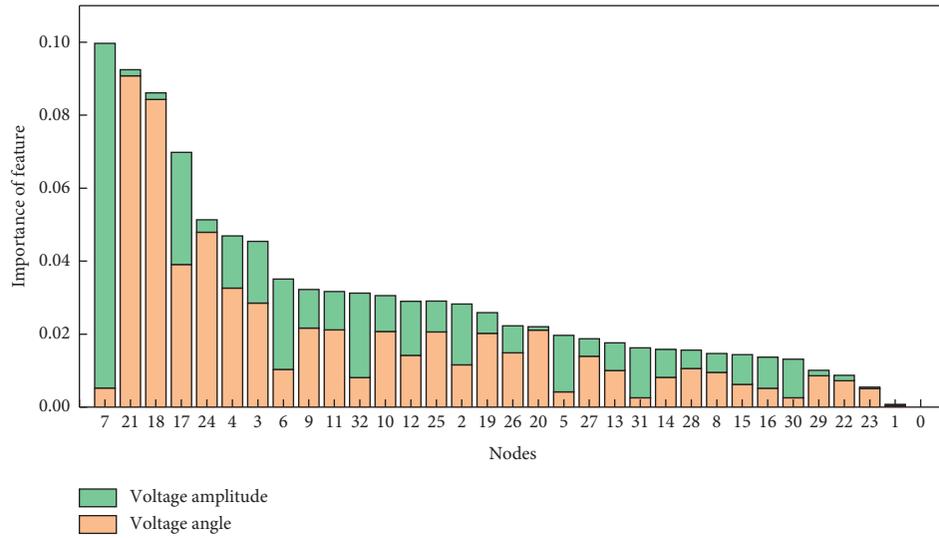


FIGURE 5: Node feature importance of IEEE-33 bus system.

TABLE 1: The training paraments and structure paraments of DBN.

Paraments	Value	Paraments	Value
Pretraining iteration	100	Reverse fine-tuning learning rate	0.01
RBM learning rate	0.0005	Reverse fine-tuning epoch	1500
Activation function	Sigmoid	Structure of hidden layer	300-100-50
Training batch-size	64	Samples of pretraining per batch	150

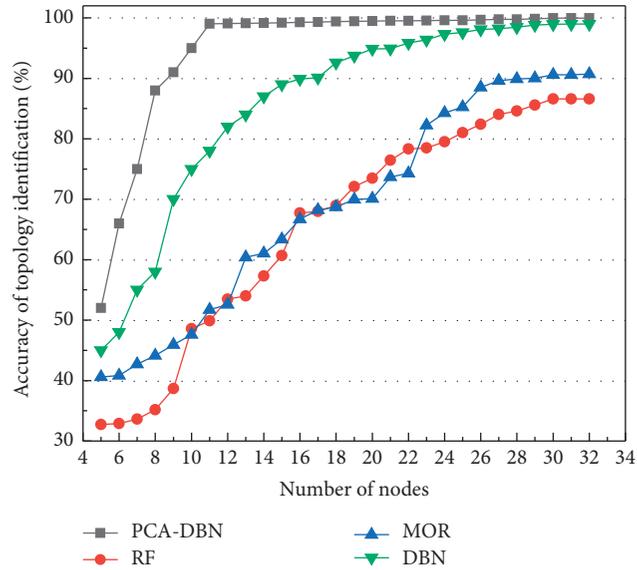


FIGURE 6: Topology identification accuracy results of four different models.

The impact of ZIP model proportions is shown in Table 3, and the measurement error of 1% is added. It can be observed that the drops of topology identification accuracy

of proposed algorithm are limited to within 1%. Therefore, the proposed algorithm is robust to the different proportions of ZIP model.

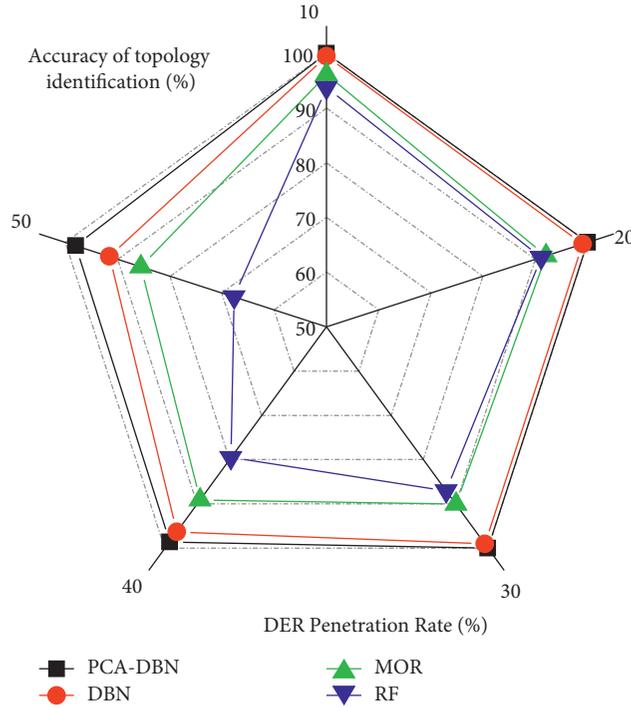


FIGURE 7: The topology identification accuracy versus different renewable energy penetration rates.

TABLE 2: Topology identification accuracy versus different noise.

Algorithm	Measurement error				
	0.1%	0.5%	1%	5%	10%
PCA-DBN	98.77	98.54	98.41	96.68	94.72
DBN	77.94	77.62	77.23	76.12	74.31
MOR	53.96	53.77	53.37	50.17	46.24
RF	50.62	50.51	50.02	46.34	41.57

TABLE 3: Topology identification accuracy versus different ZIP model proportions.

Algorithm	ZIP model proportions				
	1:0:0	0:1:0	0:0:1	0.5:0.3:0.2	0.6:0.2:0.2
PCA-DBN	97.82	98.22	98.29	98.41	98.34
DBN	76.07	76.83	76.54	77.23	77.09
MOR	52.14	53.07	52.64	53.37	53.34
RF	42.31	40.86	40.77	50.02	48.62

6. Conclusion

To solve the problem of distribution network topology analysis under a high renewable energy penetration rate, a distribution network topology analysis method based on decision trees and deep learning is proposed. The PCA-DBN model is used to analyze the observable changes of the distribution network topology under different number of measuring devices. With the goal of complete observability of the distribution network, the optimization of the location of the measurement device is carried out. And several cases are carried out on the IEEE-33 bus system to test the performance of the proposed topology identification algorithm. These results can be summarized as follows:

- (1) In the IEEE-33 bus system, using the method proposed, only the measurement data of 11 important nodes can realize the complete observability of the distribution network topology.
- (2) The results show that the proposed analysis algorithm shows good accuracy at 10% to 50% renewable energy penetration rate. In particular, when the penetration of renewable energy reaches 50%, the topology identification accuracy rate of 98% can still be achieved.

In summary, the proposed method has good adaptability for the distribution network topology identification with high renewable energy penetration rates, which have

significant meaning for the investment and operation of distribution network. In the future, we will further the impact of the renewable energy's locations on the measuring devices' configuration.

Data Availability

The data used to support the findings of this study have not been made available because the algorithm calls programs from different programming languages. The program can be reproduced through the process and ideas provided in this paper.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study is funded by National Natural Science Foundation of China (No. 51977133).

References

- [1] B. Wei, Z. Qiu, and G. Deconinck, "A mean-field voltage control approach for active distribution networks with uncertainties," *IEEE Transactions on Smart Grid*, vol. 12, no. 2, pp. 1455–1466, 2021.
- [2] V. Telukunta, J. Pradhan, A. Agrawal, M. Singh, and S. G. Srivani, "Protection challenges under bulk penetration of renewable energy resources in power systems: a review," *CSEE Journal of Power and Energy Systems*, vol. 3, no. 4, pp. 365–379, 2017.
- [3] Q. Cui, X. Bai, and W. Dong, "Collaborative planning of distributed wind power generation and distribution network with large-scale heat pumps," *CSEE Journal of Power and Energy Systems*, vol. 5, no. 3, pp. 335–347, 2019.
- [4] X. Qiao, Y. Luo, J. Xiao et al., "Optimal scheduling of distribution network incorporating topology reconfiguration, BES and load response: a MILP model," *CSEE Journal of Power and Energy Systems*, 2020.
- [5] W. Sun, F. Neumann, and G. P. Harrison, "Robust scheduling of electric vehicle charging in LV distribution networks under uncertainty," *IEEE Transactions on Industry Applications*, vol. 56, no. 5, pp. 5785–5795, 2020.
- [6] V. V. S. N. Murty and A. Kumar, "Multi-objective energy management in microgrids with hybrid energy sources and battery energy storage systems," *Protection and Control of Modern Power Systems*, vol. 5, pp. 1–20, 2020.
- [7] W. Huang, N. Zhang, C. Kang, M. Li, and M. Huo, "From demand response to integrated demand response: review and prospect of research and application," *Protection and Control of Modern Power Systems*, vol. 4, pp. 1–13, 2019.
- [8] G. Cavraro and R. Arghandeh, "Power distribution network topology detection with time-series signature verification method," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 3500–3509, 2018.
- [9] V. Freitas and A. S. Costa, "Integrated State & topology estimation based on a priori topology information," in *Proceedings of the 2015 IEEE Eindhoven PowerTech*, pp. 1–6, Eindhoven, Netherlands, July 2015.
- [10] S. M. S. Alam, B. Natarajan, and A. Pahwa, "Distribution grid state estimation from compressed measurements," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1631–1642, 2014.
- [11] A. N. Samudrala, M. H. Amini, S. Kar, and R. S. Blum, "Optimal sensor placement for topology identification in smart power grids," in *Proceedings of the 2019 53rd Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–6, Baltimore, MD, USA, March 2019.
- [12] D. Deka, M. Chertkov, and S. Backhaus, "Topology estimation using graphical models in multi-phase power distribution grids," *IEEE Transactions on Power Systems*, vol. 35, no. 3, pp. 1663–1673, 2020.
- [13] Y. Liao, Y. Weng, G. Liu, and R. Rajagopal, "Urban MV and LV distribution grid topology estimation via group lasso," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 12–27, 2019.
- [14] F. Ahmad, A. Rasool, E. E. Ozsoy, A. Sabanović, and M. Elitas, "Successive-over-relaxation based recursive bayesian approach for power system configuration identification," *COMPEL: The International Journal for Computation & Mathematics in Electrical & Electronic Engineering*, vol. 36, no. 4, pp. 1043–1058, 2017.
- [15] K. R. Mestav, J. Luengo-Rozas, and L. Tong, "Bayesian state estimation for unobservable distribution systems via deep learning," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4910–4920, 2019.
- [16] R. Oliveira, R. Bessa, and V. M. Iranda, "Identifying topology in power networks in the absence of breaker status sensor signals," in *Proceedings of the 2018 19th IEEE Mediterranean Electrotechnical Conference (MELECON)*, pp. 160–165, Marrakesh, Morocco, May 2018.
- [17] N. Duan and E. M. Stewart, "Deep learning based power distribution network switch action identification leveraging dynamic features of distributed energy resources," *IET Generation, Transmission & Distribution*, vol. 13, no. 14, pp. 3139–3147, 2019.
- [18] R. Singh, B. C. Pal, R. A. Jabr, and R. B. Vinter, "Meter placement for distribution system state estimation: an ordinal optimization approach," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2328–2335, 2011.
- [19] Z. Zhao, Yu Hao, Li Peng et al., "Optimal placement of PMUs and communication links for distributed state estimation in distribution networks," *Applied Energy*, vol. 256, Article ID 113963, 2019.
- [20] A. A. G. Mabaning, J. R. C. Orillaza, and A. von Meier, "Optimal pmu placement for distribution networks," in *Proceedings of the 2017 IEEE Innovative Smart Grid Technologies - Asia (ISGT-Asia)*, pp. 1–6, Auckland, New Zealand, December 2017.
- [21] L. Zhao, Y. Liu, J. Zhao et al., "Robust PCA-deep belief network surrogate model for distribution system topology identification with DERs," *International Journal of Electrical Power & Energy Systems*, vol. 125, Article ID 106441, 2021.
- [22] M. Lavorato, J. F. Franco, M. J. Rider, and R. Romero, "Imposing radiality constraints in distribution system optimization problems," *IEEE Transactions on Power Systems*, vol. 27, no. 1, pp. 172–180, 2012.
- [23] American National Standard for Electricity Meters, ANSI Standard c12.20-2010, pp. 1–11, 2010.
- [24] J. Yu, Y. Weng, and R. Rajagopal, "A data-driven parameter and topology joint estimation framework for time-varying system in distribution grids," *IEEE Transactions on Power Systems*, vol. 34, no. 3, pp. 1682–1692, 2019.