

Data-model Hybrid Driven Topology Identification Framework for Distribution Networks

Dongliang Xu, *Student Member, IEEE*, Zaijun Wu, *Member, IEEE*, Junjun Xu, *Member, IEEE*,
and Qinran Hu, *Member, IEEE*

Abstract—Extensive penetration of distribution energy resources (DERs) brings increasing uncertainties to distribution networks. Accurate topology identification is a critical basis to guarantee robust distribution network operation. Many algorithms that estimate distribution network topology have already been employed. Unfortunately, most are based on data-driven alone method and are hard to deal with ever-changing distribution network physical structures. Under these backgrounds, this paper proposes a data-model hybrid driven topology identification scheme for distribution networks. First, a data-driven method based on a deep belief network (DBN) and random forest (RF) algorithm is used to realize the distribution network topology rough identification. Then, the rough identification results in the previous step are used to make a model of distribution network topology. The model transforms the topology identification problem into a mixed integer programming problem to correct the rough topology further. Performance of the proposed method is verified in an IEEE 33-bus test system and modified 292-bus system.

Index Terms—Data-model hybrid driven, DBN-RF, mixed-integer programming, topology identification.

I. INTRODUCTION

THOPOLOGY identification is an essential prerequisite for partition control and state estimation of distribution networks [1]. Lack of accurate identification of distribution networks has long been cited as the main reason for operation failure and recovery difficulty of distribution networks. In daily operation, switching actions and other procedures may change the topology of the distribution network [2]. Extensive penetration of DERs in distribution networks may change current topology and bring uncertainties to the distribution network [3]. It is a tremendous challenge for distribution network topology identification.

Unlike a power transmission system topology, which is rarely reconfigured and can be identified by generalized state estimation [4], monitoring equipment of a distribution network is scarce. Although some line parameters can be obtained through a state estimation algorithm [5], it is challenging to

replace real-time measurement due to the precision and real-time performance of state estimation technology. Therefore, traditional transmission system identification technology cannot be applied to distribution networks.

During early research on distribution network topology identification, connection changes caused by switch actions are used to erect an incidence matrix or adjacency matrix [6], [7]. Network analysis and state estimation algorithm are employed to judge correctness of the network structure. However, this method is challenging to deal with the situation that multiple structural errors and bad data occur at the same time.

With rapid development of distribution automation systems, measurement becomes more accurate and diverse. With popularization of measurement equipment such as SCADA, PMU, and SM, the p , v , q , and θ of buses become clear to operators. Measurement parameters with time information can even be easily obtained [8], which brings more opportunities for topology identification of distribution networks. Topology identification methods based on measurement information have been significantly developed [9]–[11]. In [9] and [10], based on measurement obtained from a smart meter, a graphic model is presented to describe the probabilistic relationship between different voltage measurements. With power flow analysis, identification algorithm based on mutual information is proposed to deal with tree and partially meshed networks [9]. In [10], the Chow-Liu algorithm is employed, expanding the method from single-phase representation to multi-phase balanced systems with wrong bus phase labels. In [11], a novel single-phase-to-ground (SPTG) faulted line identification method is proposed; line errors in the topology are well-identified using measurements from PMU. Literature [12] offers a way to jointly estimate a polyphase distribution network's admittance parameters and topology from available telemetry data. This method makes most of the limited voltage and current phasor measurements from inexpensive distribution PMUs identify the admittance matrix of a polyphase power distribution network.

Development of measurement technology also promotes application of the data-driven method in modern distribution network identification [13]–[16]. Literature [13] applies an approach based on mixed-integer quadratic programming (MIQP), and a weighted least square method (WLS) is used to effectively identify the topological state from data in the measurement residual. In [14], a robust principal component analysis coupled deep belief network (PCA-DBN) surrogate model is proposed for distribution network topology identifica-

Manuscript received August 22, 2021; revised December 1, 2021; accepted February 8, 2022. Date of online publication January 25, 2023; date of current version March 28, 2024.

D. L. Xu, Z. J. Wu, J. J. Xu and Q. R. Hu are with School of Electrical Engineering, Southeast University, Nanjing 210096, China.

J. J. Xu is also with College of automation, Nanjing University of Posts and Telecommunications, Nanjing 210096, China.

DOI: 10.17775/CSEEJPES.2021.06260

tion, which maintains robustness to measurement noise. Power line communication (PLC) network overlying the distribution network is also helpful for topology reconstruction [15]. Analysis of time-series measurements is also employed for topology estimation [16]; time-series data taken from a dynamic system show specific patterns regarding state transitions compared against a library of signatures derived from topology simulations to estimate topology.

Unfortunately, most research focus on model-free and data-driven identification. A single data-driven method is often divorced from a real distribution network model, which is challenging to deal with the ever-changing distribution network physical structure. Traditional model-based topology identification schemes [17], [18] provide a lot of valuable references. Literature [17] addresses assigning every substation the set of buses that it is feeding, given field data, which provides a bus clustering method and transforms the problem into a nonconvex optimization problem. In [18], network topology changes (including line topology changes, node topology changes, and switch actions) are related to resultant perturbations in the voltage amplitude covariance matrix, which can monitor topology with a relatively minor investment.

To deal with increasingly complex modern distribution networks, an effective combination of data-driven and model-driven methods, and exploring a new topology identification strategy hybrid driven by data and model have received extensive attention from scholars [19], [20]. Literature [19] provides a numerical method for topology and line parameter identification. First, network information is preliminarily identified by the process of data regression. On this basis, the Newton-Raphson iteration model is used to realize fine identification of network information. In [20], a generalized state estimation method employs the real measurement and pseudo measurement data to realize topology identification of the network. At the same time, the network splitting problem is solved by modeling the switching elements separately. The above data-model hybrid driven method is more suitable for real changing network topology.

With rapid development of renewable energy power generation, topology identification technology needs to consider characteristics of renewable energy in distribution networks [21], [22]. A hybrid framework is proposed in [21] to deal with these uncertainties from renewable energy systematically and theoretically. Based on the metric space design, jointly temporal-spatially analyzing the network topology. Literature [22] proves that in practice, it is reasonable and feasible to list all possible models to deal with uncertainty caused by renewable energy in a distribution network.

A data-model hybrid driven method proposed in this paper differs from existing ones in that it has the following innovations and contributions for distribution network topology identification:

- 1) A two-step topology identification framework driven by data and model is established to improve accuracy of topology identification while ensuring speed of identification.
- 2) A DBN-RF rough identification algorithm is designed to estimate the topology by learning the relationship between

measurement data and topology connection.

3) The topology identification problem is transformed into a mixed-integer programming problem based on bounded measurement error and circuit constraints to model the distribution network.

Case studies in the IEEE 33-bus test system and modified 292-bus system verify performance of the proposed method and demonstrate applicability of the proposed method in high penetration of DERs and different load types.

The rest of this paper is organized as follows: Section II introduces the motivation, framework, and the exact steps of the proposed method. Section III gives a case study of this proposed method based on an IEEE 33-bus system and modified 292-bus system. Finally, Section IV summarizes the proposed method and concludes.

II. PROPOSED TOPOLOGY IDENTIFICATION

A. Motivation and Framework

Distribution networks are large and complex. Due to operation of sectionalizing switches, capacitor banks, and tie-switches, the structure of the distribution network often changes. Topology analysis of the distribution network is based on the connectivity of electrical components. Topology connection analysis is carried out regarding the whole distribution network as a topology diagram of line and point combination according to the power supply nodes, switch nodes, etc. It is the basis of many distribution network analyses, such as state estimation, power flow calculation, fault location, isolation and power supply recovery, and network reconstruction.

In practice, a distribution network is different from a loop network system of the transmission network. Nodes and switches are the main components of a radial distribution network topology. Topology of the distribution network can be described by the 0/1 state of node connection and switch action. Therefore, topology identification of the distribution network is the binary state problem of identifying node connection and branch switch switching from network parameters obtained from measuring equipment. In this case, topology identification can be regarded as regressing the topology state from network parameters.

The relationship between network parameters and network topology is complex; deep belief (DBN) and random forest (RF) can solve nonlinear regression problems. This paper proposes a data-model hybrid-driven method that combines topology identification based on DBN-RF with a topology identification method based on model optimization to avoid overtraining deep learning and prolonging identifying time.

The framework of the proposed method is shown in Fig. 1. First of all, regressing the topology information of the distribution network quickly and roughly by DBN-RF. Then, optimize the rough identification results by mixed integer programming for further identification. Through the two-step identification framework, distribution network topology can be identified with high accuracy and speed.

B. Data-driven Rough Identification

In terms of radial distribution networks, it is assumed the load is adjusted by similar factors, and such essential

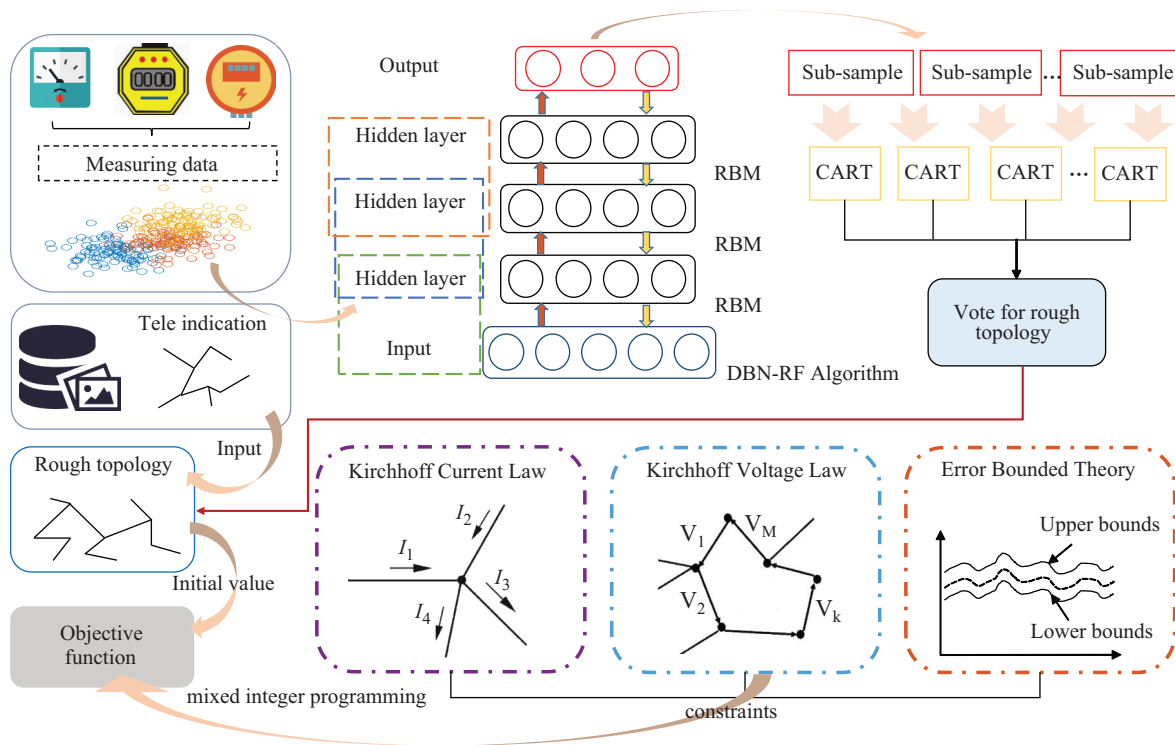


Fig. 1. Framework of data-model hybrid driven method for distribution network topology identification.

events rarely occur as capacitor bank switching or topology change. In this case, voltage is determined by three main statistical characteristics: voltage drop sequence, voltage amplitude sequence, and voltage measurement change point [14]. Therefore, learning from voltage measurement can obtain the topology of a distribution network.

Unlike the intuitive response of voltage data to change of topology, other measurement data such as current data as the input of data-driven part can also reflect real-time topology information to a certain extent. However, it is still inseparable from assistance of voltage information. In this paper, data-driven rough identification is not based on the purpose of directly obtaining accurate system topology. It only needs to provide a high-quality solution for the model driving part. Therefore, without considering other measurement information, only exploring the relationship between voltage data and topology information can realize the function of data-driven rough identification, significantly reducing the burden of data processing and avoiding unnecessary errors.

The proposed DBN-RF rough identification model preliminarily regresses topology information of the distribution network by collecting voltage data provided by measuring equipment. The model includes voltage association feature mining based on DBN and topology information regression based on RF. Processed measurement data are provided to the DBN model to extract the association feature between voltage information and topology connection state. On this basis, the extracted association features are input into the RF model to quickly predict network topology information.

1) Data Generation for DBN-RF Algorithm

Different network topologies can be generated by opening

and closing tie-switches and sectionalizing switches in the distribution network to generate DBN and RF training data. At the same time, high penetration of DERs and load ZIP type are also considered to fit actual operation of the distribution network.

Different operation scenarios are generated based on the orthogonal array sampling (OAS) method [25]. It aims to make sample distribution more uniform and improve topology identifiers' generalization ability, then collects measurements from AMI, SCADA, etc., and compute the power flow of scenarios.

OAS is an improvement of Latin hypercube sampling. It is not layered on all one-dimensional projections of the sample which can make sample distribution more uniform than simple random sampling (SRS).

In practice, it is not easy to ensure accuracy and completeness of the measurement data provided by measuring equipment. It is necessary to preprocess measurement data to ensure reliability of the association feature obtained by DBN. In this paper, the K-nearest neighbor method is employed to deal with the missing problem of measurement data. K samples nearest to the sample with missing data are determined according to Euclidean distance. The weighted average of these K values is used to estimate the missing measurement data [23].

2) Association Feature Mining Based on DBN

After being preprocessed, providing these measurement data to the DBN model for association feature mining, the deep belief network stacked by multiple Restricted Boltzmann Machines (RBM) was proposed by Hinton in 2006 [24]. The RBM is a probabilistic graph model based on energy and can be described by stochastic probabilistic neural networks. Each

RBM consists of a two-layer network consisting of visible layer $v = \{v_1, v_2, v_3, \dots, v_I\}$ and hidden layer $h = \{h_1, h_2, h_3, \dots, h_J\}$, where I and J are the number of neural units in the visible layer and hidden layer, respectively. The corresponding visible layer unit bias is $a = \{a_1, a_2, a_3, \dots, a_I\}$, and the hidden layer unit bias is $b = \{b_1, b_2, b_3, \dots, b_J\}$. Neural units in the same layer are independent of each other, and neural units in different layers are fully connected. The visible layer neural unit v_i is associated with the hidden layer neural unit h_j by weight w_{ji} .

Therefore, given v and h , the energy function $E(v, h)$ of RBM can be defined as (1).

$$E(v, h) = - \sum_{j=1}^J \sum_{i=1}^I w_{ji} v_i h_j - \sum_{i=1}^I v_i a_i - \sum_{j=1}^J h_j b_j \quad (1)$$

According to (1), the joint probability distribution between the hidden and visible layers' units can be shown as (2).

$$p(v, h) = \frac{e^{-E(v, h)}}{\sum_v \sum_h e^{-E(v, h)}} \quad (2)$$

Since neurons in the same layer of RBM are independent, when the state of a certain layer of neurons is given, the form of another layer of neurons can be derived according to (2). The hidden and visible layers' activation probability can be shown as (3) and (4).

$$p(v_i = 1 | h) = \sigma \left(a_i + \sum_{j=1}^J h_j w_{ji} \right) \quad (3)$$

$$p(h_i = 1 | v) = \sigma \left(b_i + \sum_{j=1}^I v_j w_{ji} \right) \quad (4)$$

where σ is the sigmoid function. Sigmoid function σ is defined as, $\sigma = 1/(1 + e^{-x})$ the topology identification as the 0/1 binary state problem.

RBM learning aims to make distribution represented by the RBM network fit the input data as much as possible and maximize the $p(v)$.

Then, the a, b and w are updated following rules (5).

$$\begin{cases} [a_j]_n = \eta[a_j]_{n-1} + \gamma[E_d(v_i) - E_m(v_i)] \\ [b_i]_n = \eta[b_i]_{n-1} + \gamma[E_d(h_j) - E_m(h_j)] \\ [w_{ij}]_n = \eta[w_{ij}]_{n-1} + \gamma[E_d(v_i h_j) - E_m(v_i h_j)] \end{cases} \quad (5)$$

where the learning rate η and γ are in $[0, 1]$, n is the number of iterations, $E_d(\cdot)$ and $E_m(\cdot)$ are the expectations for data distribution and model distribution, respectively.

The DBN network consists of the stacked RBM at the bottom and the classifier at the top. The stacked RBM at the bottom completes the learning and feature extraction of complex data, and the classifier at the top realizes fine-tuning of the structural parameters of the whole model and classification of data. The training of DBN includes two stages: unsupervised layer by layer pre-training and supervised fine-tuning.

The initial data set generated from measurement data by unsupervised learning is provided to the DBN model. Then, the whole model is fine-tuned. In this stage, based on the initial

DBN model, weights and biases between RBMs are retained, and network errors are calculated with labels in the data set as the standard. A Backpropagation (BP) algorithm is used to get the errors of each layer, and the gradient descent method is used to adjust the weights and biases of each layer. The structure of the DBN model is shown in Fig. 2.

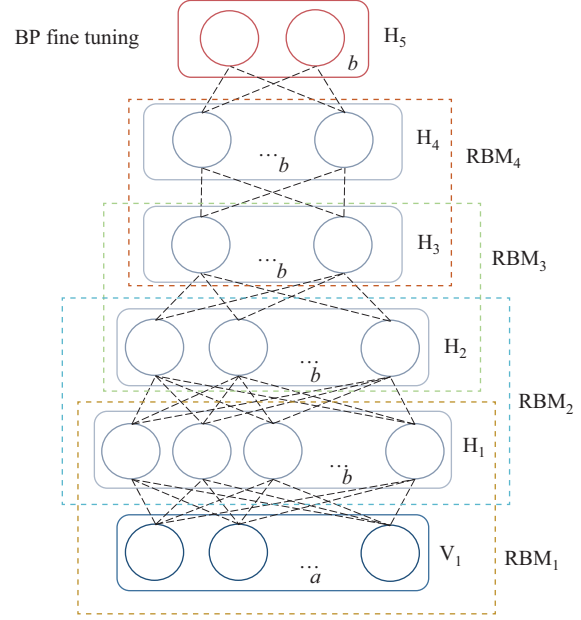


Fig. 2. Structure of DBN model.

In summary, procedure of the proposed DBN-based association feature mining algorithm is as Algorithm 1.

Algorithm 1: Association Feature Mining Algorithm

Input: Measurement data, Maximum number of layers, Number of nodes per layer, Maximum number of iterations

- 1 **Set** layer $i = 1$
- 2 Unsupervised pre-training.
- 3 **if** each layer of DBN has been traversed **then**
- 4 BP based fine tuning
- 5 **else**
- 6 Back to Step 2, and layer $i = i + 1$
- 7 **end**

Output: Association feature between measurement data and topology.

3) Rough Identification Based on RF

RF algorithm is the combination of the Bagging algorithm and Random Subspace algorithm. It can be seen as a combination of multiple decision trees. In a random forest, each decision tree is randomly generated, and different decision trees are independent. Numerous random trees in the forest vote together to determine classification results of the input samples, and a sub-decision tree in the random forest algorithm is as follows:

$$\{f(x, \delta_k), k = 1, 2, 3, \dots, n\} \quad (6)$$

where x is the input vector, δ_k is a set of independent k samples, and $f(x, \delta_k)$ is the basic classifier of the sub-decision tree. The decision tree of the random forest algorithm is based on the CART algorithm, where the Gini coefficient is employed to select features: the smaller the Gini coefficient is, the lower the impurity is, the better the features are.

To improve the features, the calculation process of Gini minimum impurity criterion is as (7):

$$\text{Gini}(s) = 1 - \sum_{i=1}^c [p(i|s)]^2 \quad i = 1, \dots, c \quad (7)$$

where $p(i|s)$ is the probability of class i on node s .

If the sample can be divided into m branches, the Gini coefficient of the node is:

$$\text{Gini}(S) = \sum_{j=1}^m \frac{n_j}{n} \text{Gini}(j) \quad (8)$$

Generating a decision tree includes two parts: a random selection of features and node splitting.

Random selection of features aims to generate m feature quantities randomly for comparison of node splitting, and m can be chosen as (9) usual.

$$m = \log_2 M + 1 \quad (9)$$

where M is the number of input variables.

Node splitting is based on the CART algorithm and Gini minimum impurity criterion. Suppose the size of a random forest is N . In that case, N decision trees should be generated to form a random forest through the above method. Final classification results are decided by the N trees voting, respectively. The principle of the RF algorithm is shown in Fig. 3.

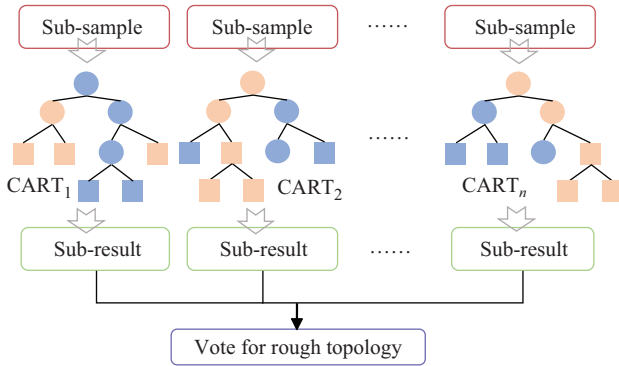


Fig. 3. The principle of RF algorithm.

Based on the theory above, the RF rough identification can be carried out following these steps:

Step 1: Inputting the sub-sample of network topology information represented by feature vectors that are mined from DBN. Bootstrap resampling is employed to sample and form multiple sub-sample sets, which guarantees completeness of samples and avoids overfitting.

Step 2: For each sub-sample set, CART algorithm is used to obtain a wholly split independent sub-decision tree.

Step 3: Training decision tree and select m features from all features randomly to obtain a candidate feature information subset of all the non-leaf branch nodes.

Step 4: Each feature information in the candidate feature information subset is used to cut nodes, and feature information with the best cutting ability is selected from the feature information subset according to Gini minimum impurity criterion.

Step 5: Repeating Step 3 and Step 4 until each node of this sub decision tree is cut according to Gini minimum impurity criterion.

Step 6: Repeating Step 1 to Step 5 until the number of sub-decision trees can meet the RF algorithm's demand and vote, respectively, to obtain rough topology identification results.

C. Model-Based Further Identification

A further topology identification model is established according to electric circuits and error bounded theory principles to optimize topology information after rough identification and minimize errors in topology identification.

During distribution network operation, nodes' current injection follows Kirchhoff Current Law (KCL). The sum of current flowing into the node is equal to the sum of current flowing out of the node: $\sum_{k=1}^n i_{ik} = 0$, where i_{ik} is current flowing through the k th branch connected to this node i . According to the Kirchhoff Voltage Law (KVL), current flows between nodes also depend on the nodes' voltage: $\sum_{k=1}^m v_{ik} = 0$, where m is the number of elements of closed-loop and v_{ik} is the voltage at both ends of the component.

The principles of electric circuits above can be used as the constraint of the optimal topology identification result. The coefficient of connection ξ is employed to adapt to the topology connection problem. If the branch between two nodes is connected, then ξ is one; otherwise, ξ is zero. Constraints KCL and KVL are modified:

$$i_i = \sum_{j=1, j \neq i}^n \xi_{ij} i_{ij} \quad (10)$$

$$i_{ij} R_{ij} = \xi_{ij} (v_i - v_j) \quad (11)$$

where R_{ij} is the resistance between node i and j .

Based on error bounded theory, the measurement equation follows (12):

$$V - e^+ < f(x) < V - e^- \quad (12)$$

where x is the state variable, $f(x)$ is the measurement equation, V is the measurement vector, e^+ and e^- are upper and lower bounds of measurement error, respectively.

To minimize errors in topology identification, the objective function is:

$$\min \sum_{\forall i, j \in \Gamma} |\xi_{ij} - \hat{\xi}_{ij}| \quad (13)$$

where ξ_{ij} is the connection condition between node i and j . $\hat{\xi}_{ij}$ is the connection condition obtained from tele indication, which cannot completely cover all network topology information, and Γ is the set of all nodes in the network.

In summary, model-based further identification is an optimization problem, which is:

$$\begin{aligned}
\min \quad & \sum_{\forall i,j \in \Gamma} |\xi_{ij} - \hat{\xi}_{ij}| \\
\text{s.t.} \quad & V - e^+ < \sum_{\forall i,j \in \Gamma} \xi_{ij} f_{ij}(x) < V - e^- \\
& i_{ij} R_{ij} = \xi_{ij} (v_i - v_j) \\
& i_i = \sum_{j=1, j \neq i}^n \xi_{ij} i_{ij}
\end{aligned} \quad (14)$$

In model (14), ξ_{ij} is the variable related to the v_i and i_i , so the decision variables of the optimized model (14) are $U_{ij}(v_i, v_j)$, and $i_{ij}(i_i, i_j)$.

Unfortunately, several nonlinear constraints in the optimization problem (14) may complicate the solution. The method in [26] is employed to solve this problem.

To overcome the nonlinearity in model (14), transforming (10) to (15):

$$i_i = \sum_{\forall j \in \Upsilon} i_{ij} \quad (15)$$

where Υ is the set of all nodes connected to node i .

Then, $\xi_{ij}(v_i - v_j)$ in (11) and $f(x)$ in (12) are replaced with U_{ij} and χ_{ij} , respectively. If ξ_{ij} is one, then U_{ij} is equal to $(v_i - v_j)$ and χ_{ij} is equal to $f_{ij}(x)$, otherwise U_{ij} and χ_{ij} are zero.

The relationship between U_{ij} , χ_{ij} , ξ_{ij} , $(v_i - v_j)$ and $f_{ij}(x)$ can be achieved in the following constraints:

$$-M_1(1 - \xi_{ij}) \leq U_{ij} - (v_i - v_j) \leq M_1(1 - \xi_{ij}) \quad \forall i, j \in \Gamma \quad (16)$$

$$-M_1 \xi_{ij} \leq U_{ij} \leq M_1 \xi_{ij} \quad \forall i, j \in \Gamma \quad (17)$$

$$-M_2(1 - \xi_{ij}) \leq \chi_{ij} - f_{ij}(x) \leq M_2(1 - \xi_{ij}) \quad \forall i, j \in \Gamma \quad (18)$$

$$-M_2 \xi_{ij} \leq \chi_{ij} \leq M_2 \xi_{ij} \quad \forall i, j \in \Gamma \quad (19)$$

where M_1 and M_2 are the column vectors in the same dimension composed of a positive number large enough.

According to the linear constraints above, optimization problem (14) is transformed into (20) without nonlinear constraints:

$$\begin{aligned}
\min \quad & \sum_{\forall i,j \in \Gamma} |\xi_{ij} - \hat{\xi}_{ij}| \\
\text{s.t.} \quad & i_i = \sum_{\forall j \in \Upsilon} i_{ij} \\
& i_{ij} R_{ij} = U_{ij} \\
& -M_1(1 - \xi_{ij}) \leq U_{ij} - (v_i - v_j) \leq M_1(1 - \xi_{ij}), \quad \forall i, j \in \Gamma \\
& -M_1 \xi_{ij} \leq U_{ij} \leq M_1 \xi_{ij}, \quad \forall i, j \in \Gamma \\
& V - e^+ < \sum_{\forall i,j \in \Gamma} \chi_{ij} < V - e^- \\
& -M_2(1 - \xi_{ij}) \leq \chi_{ij} - f_{ij}(x) \leq M_2(1 - \xi_{ij}), \quad \forall i, j \in \Gamma \\
& -M_2 \xi_{ij} \leq \chi_{ij} \leq M_2 \xi_{ij}, \quad \forall i, j \in \Gamma
\end{aligned} \quad (20)$$

Model (20) is a mixed-integer programming problem; in solving, a rough identification result obtained from the DBN-RF method is the initial value of ξ_{ij} , which significantly

shortens solving time for further optimization distribution network topology identification.

III. CASE STUDY

In this section, effectiveness of the proposed data-model hybrid driven method for topology identification is verified in an IEEE 33-bus system and modified 292-bus system.

Section III-B to Section III-D tests the proposed method in an IEEE 33-bus system with different ZIP types of loads and different penetration of DERs. Sensitivity of the proposed method is also analyzed. Then in the following case, Section III-E, we test the proposed method in a modified 292-bus system, a comparatively large network to show the method's scalability.

All the algorithms are programmed on a Core i5-9400F 2.90Ghz CPU and 32G RAM laptop.

A. Parameters Settings and Input Generation for Data-driven Identification

The normally open or normally closed status of switches in the IEEE-33 bus system is shown in Table I. The feeder structure is shown in Fig. 4, and more information on this test system can be found in [27].

TABLE I
NORMALLY OPEN AND CLOSED SWITCHES IN THE IEEE 33-BUS SYSTEM

Normally open switches	Normally closed switches
33, 34, 35, 36, 37	4, 6, 7, 9, 10, 11, 12, 14, 15, 16, 17, 18, 26, 28, 30, 32

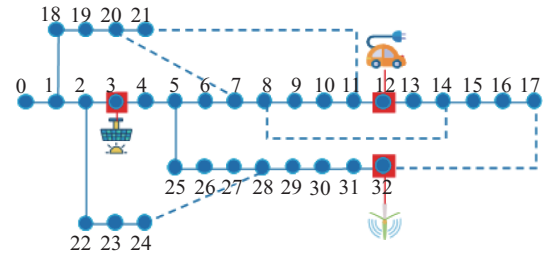


Fig. 4. IEEE 33-bus test system.

During data-driven identification, parameter settings for DBN and RF models have to be considered. Some comparisons are made between different hidden layers to find the best number of hidden layers in the DBN network.

Accuracy of topology identification is given in percentage, which is:

$$\text{accuracy} = 100\% \times \frac{N_S}{N_I} \quad (21)$$

where N_S is the number of total successful topology identifications, N_I is the number of total topology identifications.

As shown in Table II, when the number of hidden layers is smaller than 4, identification accuracy increases with increase of layers; however, setting more hidden layers will not increase identification accuracy, just wasting more time on overtraining. So the number of hidden layers is set as 4. Other parameters for DBN and RF algorithm are shown in Table III.

TABLE II
COMPARISON BETWEEN DIFFERENT NUMBER OF HIDDEN LAYERS

Number of hidden layers	Association feature mining time (ms)	Accuracy of topology identification (%)
2	9.8	89.94
3	10.1	96.53
4	12.4	99.87
5	14.6	99.85
6	16.7	99.89
7	24.4	99.88

TABLE III
PARAMETERS SETTINGS FOR DBN AND RF ALGORITHM

Parameters	Value
Number of neurons in the hidden layer	400-200-100-50
Iteration of pre-training	150
Iteration of BP fine-tuning	100
Initialization of a and b	0
Initialization of w	Random number obeyed normal distribution
Batch-size	128
Learning rate	0.1
Number of trees in random forest	300

In practice, the way to determine the number of scenarios generation and data generation needs specific analysis of a specific distribution network:

a. When the network to be identified is a newly built distribution network or a distribution network with few known operation data, it is suggested that selecting more than ten scenarios with a small number but including basic operation types, such as radial, without-island, without-loop, *et al.* 1000–5000 groups of data are generated as the data input of the data-driven rough identification part of the proposed method to ensure applicable identification accuracy.

b. When the distribution network to be identified is a distribution network with complete known information, it is suggested that selecting 100–400 operation scenarios to ensure most operation states of the distribution network can be included, and generating 10000–40000 groups of data as data input of the data-driven rough identification part of the proposed method to ensure balance between identification accuracy and time consumption

In the case study of this paper, input data of the data-driven identification is generated from operation scenarios. Total 250 different operation scenarios are generated based on the OAS, which includes radial, without-island, without-loop, *et al.* Based on these scenarios, 25000 groups of data are generated as input to the data-driven off-line learning part.

B. Performance of the Proposed Method

Monte Carlo simulation is employed for this test to show the performance of the proposed identification method. One hundred trials are made to deliver performance of the proposed method and comparisons between existing identification methods.

Meanwhile, to show the importance of combining the data-driven rough identification and model-driven further identification proposed in this paper, comparison of only model-based method results are also added. (The model-driven part of

the proposed method is solved based on commercial software Gurobi.)

In Table IV, time consumption of the proposed method is the sum of time consumption of rough identification and further identification. It can be found the average time consumed of the proposed method is 20.46 ms, which is far lower the traditional regression method, RF method, and DBN method. Moreover, from the table, we can find that time consumption of the proposed method is concentrated in many trials, which also proves the stability of the proposed method is high.

TABLE IV
TIME CONSUMPTION OF DIFFERENT METHODS

Method	Maximum time consumed (ms)	Minimum time consumed (ms)	Average time consumed (ms)
Rough identification	13.46	12.01	12.68
Further identification	9.04	6.51	7.80
Proposed method	22.01	18.63	20.46
Only model-based method (Gurobi)	163.77	109.12	141.06
Only model-based method (CPLEX)	283.32	188.78	240.11
Only model-based method (Rglpk)	2871.12	1629.79	2398.03
DBN	31.03	27.95	29.49
RF	30.52	26.12	28.41
Regression	46.61	41.43	44.05

Figure 5 shows comparison of accuracy between different methods, accuracy of the proposed method can stably exceed 99% in many trials, which is enough to meet the needs of topology observation in practical operation. Although accuracy of the traditional regression method is also high, accuracy changes significantly in many trials, and stability of the regression method performance is poor.

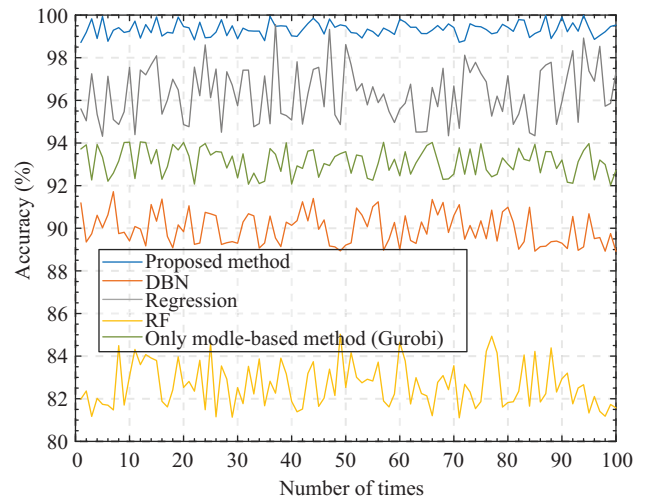


Fig. 5. Comparison of accuracy.

Accuracy and efficiency of the only model-based method are shown in Table IV and Fig. 5. It can be seen the model-driven part of this paper, as a supplement to the data-driven part, can quickly obtain high-precision solutions based on high-quality initial solutions. In the MILP problem (20), decision variables are $U_{ij}(v_i, v_j)$, and $i_{ij}(i_i, i_j)$. In the IEEE 33-bus test system, the number of v_i and i_i is only 33. However, when the

connection condition is entirely unknown (this situation is rare in practice), the scale of U_{ij} and i_{ij} will be pretty massive. U_{ij} and i_{ij} contain $\sum_{n=1}^{33-1} n = 528$ possibilities, respectively. Corresponding to this, colossal scale constraints also bring significant challenges to solution of commercial software.

Comparison between different commercial software-based only model-based methods is also shown in Table IV. Performance of the Gurobi-based method is the best of them. Accuracy of the only model-based method is lower than the proposed method, and its time consumption is almost seven times that of the best performance of the proposed method.

In the above methods, except for traditional regression, it takes 10–20 minutes to train the data, but this process takes time to match a system measurement update. It can be completed off-line, which has no impact on application of on-line topology identification.

C. Penetration of DERs

Extensive penetration of DERs will lead to terrible uncertainty of the distribution network, making voltage amplitude vary more. Accuracy of traditional topology identification is greatly affected.

To verify effectiveness of the proposed topology identification method with penetration of DERs, tests in this part are accessed in the IEEE 33-bus. Six scenes are designed, whose penetration rate of DERs are 10%, 20%, 30%, 40%, 50%, 60%, respectively.

As shown in Fig. 6, high penetration of DERs greatly influences all kinds of topology identification methods. With increasing penetration of DERs, accuracy of topology identification has decreased significantly. However, when penetration rate is less than 30%, the proposed method benefits from both data learning and physical constraints of the system, and identification accuracy is high. With increase of DERs penetration, the relationship between voltage and topology changes, bringing significant challenges to topology identification. However, even when the penetration rate reaches 60%, accuracy of the proposed method can still scratch 57%.

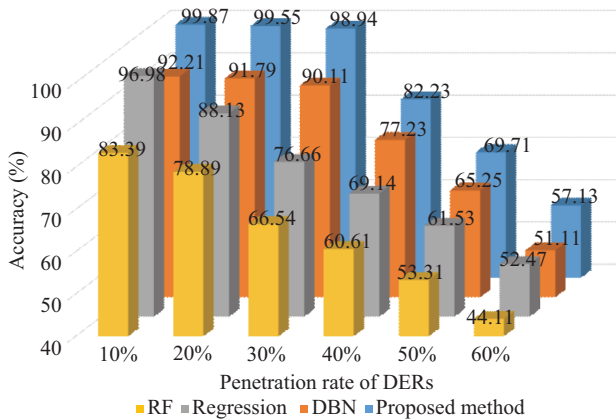


Fig. 6. Impacts of different penetration rate of DERs.

In engineering applications, accuracy is unacceptable when it falls below 90%. The DSO could find it very difficult to recover real topology from an inaccurate one with too

many errors. Through repeated tests, the author sets a reliable penetration threshold of 36% to ensure accuracy of topology identification. (When DERs penetration is 36%, identification accuracy of the proposed method is 93.97%). The proposed method can obtain very reliable topology identification accuracy when penetration is lower than 36%. 36% is a DERs penetration rate that cannot be achieved in most distribution networks at present. Therefore, the method proposed in this paper still has considerable application value in the face of distribution networks with high penetration of DERs in recent years.

On the scenes above, DERs are located in nodes 3, 12, 32. Another five scenes with 20% DERs penetration rate and different DERs locations are designed to consider the impacts of DERs location on topology identification. In Fig. 7, case 1 to case 5 are the five scenes with different DERs locations. Identification accuracy of the RF method varies significantly under different areas of DERs. Different locations of DERs have certain influence on the regression method and DBN method but have little effect on identification accuracy of the proposed method.

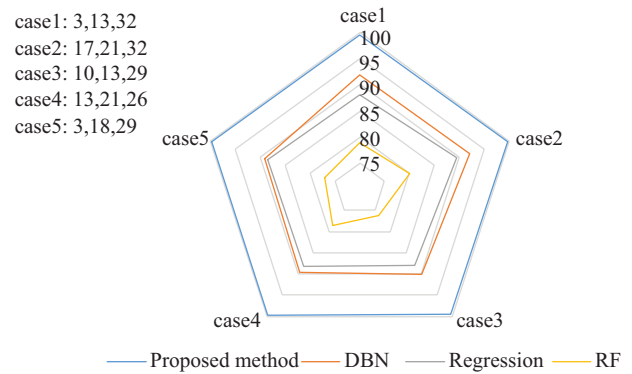


Fig. 7. Impacts of different locations of DERs.

Therefore, compared with other methods, the proposed method has excellent advantages in accuracy and adaptability to penetration of DERs.

D. Sensitivity Analysis

1) Sensitivity to Data Missing

Due to unstable factors such as communication, data missing is inevitable. As is introduced in Part A of Section III, the K-nearest neighbor method is employed to deal with data missing. In this paper, missing data rate is defined as (22):

$$\text{missingrate} = 100\% \times \frac{N_{\text{missing}}}{N_{\text{total}}} \quad (22)$$

where N_{missing} is the number of missing samples, and N_{total} is the number of total samples.

Some values in the measurement set are replaced with null values to test sensitivity to data missing of the proposed method, and the missing data rate is 0%, 5%, 10%, 15%, 20%, 25%.

As is shown in Fig. 8, in the case of data missing, the traditional regression method is greatly affected because output results are very dependent on data information. RF and DBN

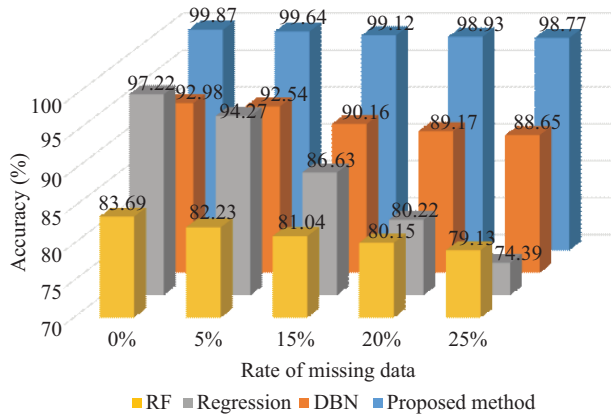


Fig. 8. Sensitivity to data missing.

and the proposed method are much less affected by data missing, especially the proposed method because the K-nearest neighbor method is used to supplement missing data. Even if the proportion of missing data reaches 25%, accuracy of topology identification is still very high.

2) Sensitivity to Load Type

In practice, loads in the distribution network are varied, including residential loads, industrial loads, agricultural loads, etc. ZIP model is usually used to show the type of load models:

$$S = S_0 \left[z \left(\frac{U}{U_0} \right)^2 + i \left(\frac{U}{U_0} \right) + p \right] \quad (23)$$

where S and S_0 are actual load power and static load power, respectively, U and U_0 are actual node voltage and static node voltage, respectively, z is the rate of constant impedance loads, i is the rate of constant current loads, p is the rate of constant power loads.

As is shown in Fig. 9, different $z : i : p$ load types are set to verify the proposed method's sensitivity to load type. Results show the accuracy of topology identification of RF and DBN methods fluctuates wildly with different load types. In contrast, the traditional regression method and the proposed method are less affected by load types. The proposed method has high accuracy in topology identification with various load types.

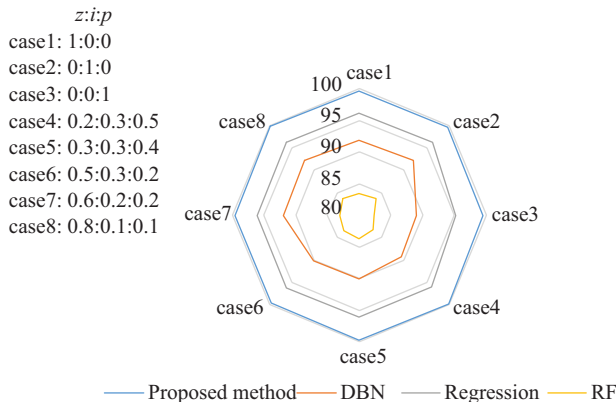


Fig. 9. Sensitivity to loads type.

3) Sensitivity to Measuring Noise

In practice, measurement data obtained from the measuring equipment are noisy, and a smart meter has typically three accuracy levels of 0.5, 1.0, and 2.0. The allowable error of class 0.5 meter is within $\pm 0.5\%$, that, of course, 1.0 meter is within $\pm 1\%$, and that, of course, 2.0 meter is within $\pm 2\%$. Therefore, it is necessary to verify performance of the proposed algorithm under different noise levels. Fig. 10 shows the average error rate and standard deviation of error rate under different noise levels. With increase of the error level, average error rate increases significantly. Accuracy of topology identification is lower than of perfect measurement. However, location of the feeder bus is usually known. Error topology branches can be removed after topology identification, which can improve accuracy of topology identification. From Fig. 10, it can also be found that standard deviations of error rate are slight, and performance of the proposed method under different noise levels is reliable.

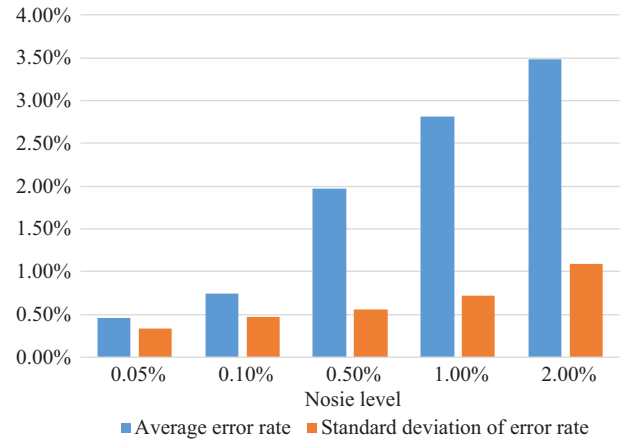


Fig. 10. Sensitivity to measuring noise.

4) Sensitivity to Insufficient Data

When a newly built distribution network needs to be identified, small samples and insufficient data will bring many problems, especially with the method based on a big data drive. Because the final identification result output of the proposed method comes from the model-driven identification part, it should have good performance in the case of insufficient data in theory.

To verify sensitivity to insufficient data of the proposed method, in addition to 25000 groups of inputs for 250 scenarios consistent with other examples, case studies of 15000 groups of inputs for 150 scenarios, 5000 groups of inputs for 50 scenarios, and 1000 groups of inputs for ten scenarios are supplemented.

As is shown in Fig. 11, the proposed method can still maintain high accuracy of topology identification for the input of small samples. In contrast, several other big data support methods have high requirements for the number of samples. When input data is insufficient, accuracy will decline catastrophically, resulting in identification failure. Time consumption of the four methods under a different number of data groups is shown in Table V. Reduction of input samples will significantly shorten identification time of traditional

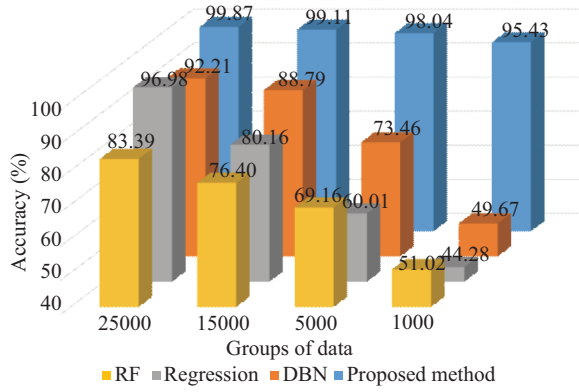


Fig. 11. Sensitivity to insufficient data.

TABLE V
TIME CONSUMPTION UNDER DIFFERENT NUMBER OF DATA GROUPS

Method\Groups	25000 (ms)	15000 (ms)	5000 (ms)	1000 (ms)
RF	28.41	26.03	20.15	14.66
Regression	44.05	39.13	30.74	21.07
DBN	29.49	26.04	19.06	14.63
Proposed method	20.46	22.17	36.41	43.19

data-driven methods, but this is at the cost of identification accuracy. When data is insufficient, off-line training time and rough identification time of the proposed method will be significantly reduced.

In contrast, overall on-line identification time will become more because insufficient off-line training samples reduce initial value accuracy obtained by the rough identification part. It increases the difficulty of solving the model-driven identification, which takes more time than the general situation, but identification results are still very reliable.

5) Sensitivity to Accuracy of Rough Identification

The method proposed in this paper is to closely connect data-driven and model-driven to obtain fast and accurate topology identification. The data-driven part aims to provide a high-quality initial solution for the model-driven part. As a supplement of the data-driven part, model-driven can quickly solve optimization problems based on high-quality initial solutions and quickly obtain accurate topology information of distribution network. So accuracy of Model-Based Further Identification is closely related to accuracy of Data-driven Rough Identification.

To quantify the relationship between accuracy of Model-Based Further Identification and accuracy of Data-driven Rough Identification, rough topology information with different accuracies is used as the solution of the data-driven part and supplied to the model-driven part. As is shown in Table VI, the higher the accuracy of the initial solution provided by

TABLE VI
PERFORMANCE OF MODEL-DRIVEN PART UNDER DIFFERENT ACCURACY OF ROUGH IDENTIFICATION

Description	90%	80% (%)	70% (%)	60% (%)	50% (%)
Accuracy	99.91%	99.84%	97.63%	95.47%	93.09%
Average time consumption of model-driven part	7.31 ms	7.92 ms	26.08 ms	37.96 ms	50.14 ms

rough identification, the higher the accuracy of the topology identification results output by the model driving part, and the less time consumption. When the initial solution accuracy reaches more than 70%, the model-driven part can provide topology identification results with acceptable accuracy while ensuring efficiency.

E. Scalability in Large Distribution Network

Real-sized distribution systems in practice have the order of 200 buses or even much more. To thoroughly verify the scalability of the proposed method on large-scale systems, a modified 292-bus test system that is part of the Elmira distribution system in New York [28] is employed. The topology of the modified 292-bus system is shown in Fig. 12.

All loads in this 292-bus system are modified as ZIP loads with $0.2:0.3:0.5 z : i : p$. 20% DERs penetration rate, which is twelve 0.45 MW PVs are placed in the modified 292-bus system, and the location of DERs is shown in Table VII.

To obtain reliable test results, the parameters of the data-driven part are reset, and new parameters are shown in Table VIII.

After 100 trials, time consumption of the proposed method and other comparative methods are shown in Table IX, and accuracy of that is shown in Fig. 13.

According to Table IX and Fig. 13, it can be found that under the condition of ensuring sufficient input data, even in a large-scale distribution network, identification accuracy of the methods involved in the comparison has not decreased

TABLE VII
LOCATION OF THE TWELVE PVs

#	Location	#	Location
DER 1	Bus-16	DER 7	Bus-83
DER 2	Bus-24	DER 8	Bus-128
DER 3	Bus-40	DER 9	Bus-187
DER 4	Bus-56	DER 10	Bus-194
DER 5	Bus-61	DER 11	Bus-235
DER 6	Bus-79	DER 12	Bus-287

TABLE VIII
PARAMETERS SETTINGS FOR DBN AND RF ALGORITHM

Parameters	Value
Number of neurons in the hidden layer	550-250-200-50-50
Iteration of pre-training	150
Iteration of BP fine-tuning	150
Initialization of a and b	0
Initialization of w	Random number obeyed normal distribution
Batch-size	512
Learning rate	0.15
Number of trees in random forest	400

TABLE IX
TIME CONSUMING OF DIFFERENT METHODS

Method	Maximum time consuming (ms)	Minimum time consuming (ms)	Average time consuming (ms)
Rough identification	44.08	36.23	40.02
Further identification	22.12	19.30	20.92
Proposed method	63.89	54.17	60.94
DBN	77.03	71.12	74.76
RF	74.09	69.85	71.18
Regression	179.94	113.27	129.76

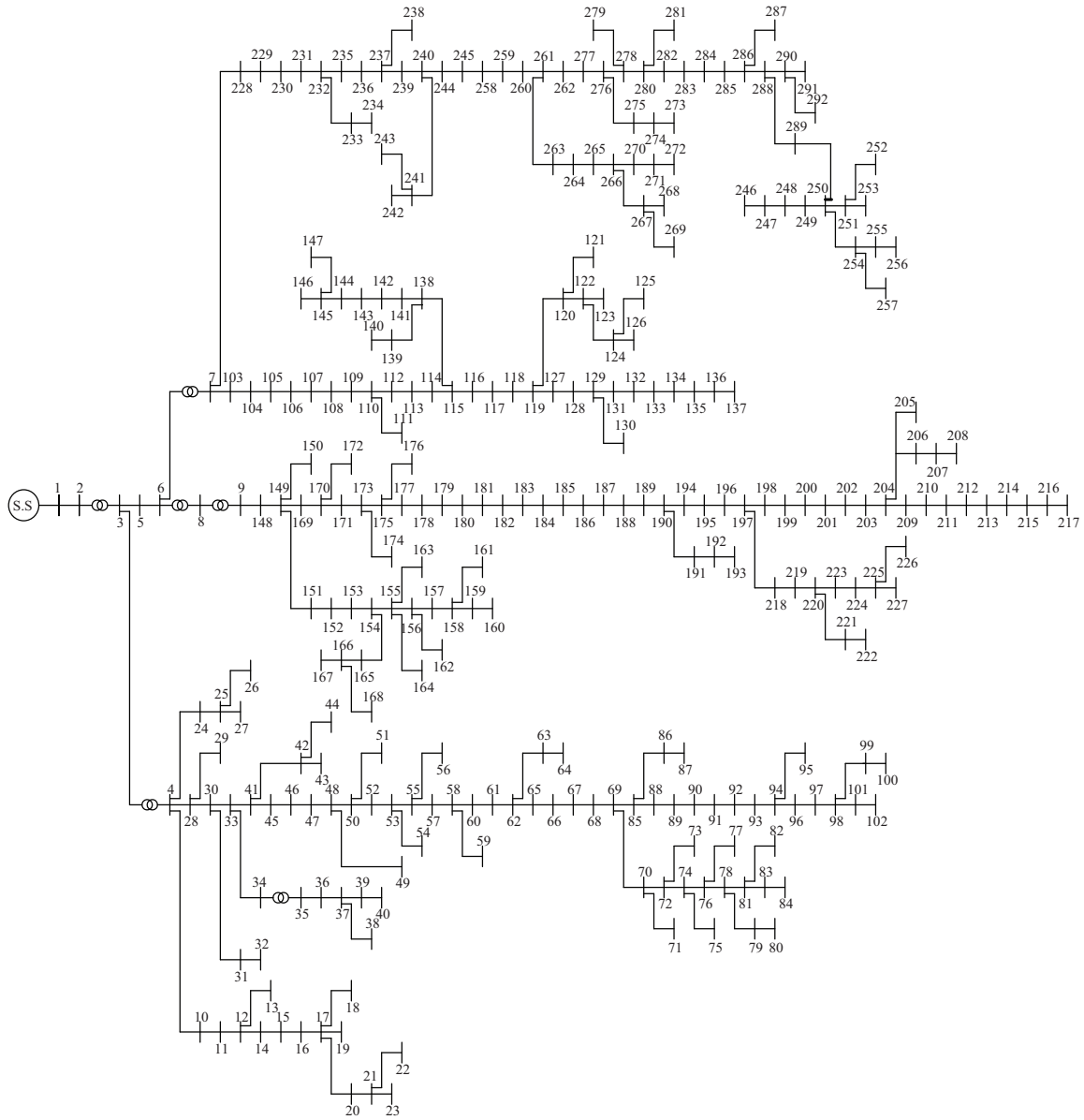


Fig. 12. The topology of a modified 292-bus system.

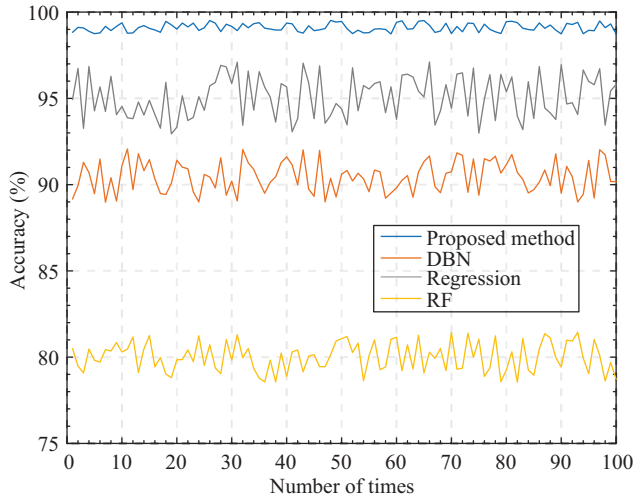


Fig. 13. Comparison of accuracy.

significantly. In particular, the method proposed in this paper can achieve an average identification accuracy of 99.12%, which is a reliable identification accuracy. Facing the large-scale system, time consumption of each method increases significantly due to increase of the amount of data. Compared with other methods, efficiency of the method proposed in this paper has significant advantages.

The above analysis shows the great scalability of the proposed method in large-scale systems.

IV. CONCLUSION

In this paper, a data-model hybrid driven topology identification method is proposed for a distribution network. A DBN network is employed to mine the association features of topology from measurement data, and the RF method is used to quickly and roughly identify network topology. Then, the rough identification result is taken as the initial value

for further identification. Based on the principle of minimum topology error, considering the bounded measurement error and physical constraints of the circuit, the topology identification problem is transformed into a mixed integer programming problem to estimate the distribution network topology accurately. The proposed method is tested on an IEEE standard system. Test results show the proposed method has excellent advantages in accuracy and speed compared with other methods. Our approach can well adapt to a distribution network with high penetration of DERs. Besides, the excellent sensitivity of the proposed method to data missing, load type, measuring noise has been verified, and results are reliable and accurate in various system situations.

Whether the proposed method applies to a three-phase unbalanced distribution network or larger distribution network, how to realize dynamic observation of distribution network topology with PMU and other measuring equipment, and how to improve topology identification accuracy under high penetration DERs are worthy of further study. We will show these studies in our following works.

REFERENCES

- [1] Y. L. Liu, P. Wang, "Partial Correlation Analysis Based Identification of Distribution Network Topology," *CSEE Journal of Power and Energy Systems*, vol. 9, no. 4, pp. 1493–1504, Jul. 2023.
- [2] H. X. Wang, B. Wang, P. Luo, and F. Q. Ma, "State evaluation based on feature identification of measurement data: for resilient power system," *CSEE Journal of Power and Energy Systems*, vol. 8, no. 4, pp. 983–992, Jul. 2022.
- [3] 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, Dec. 2017.
- [4] R. L. Lugtu, D. F. Hackett, K. C. Liu, and D. D. Might, "Power system state estimation: detection of topological errors," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-99, no. 6, pp. 2406–2412, Nov. 1980.
- [5] X. R. Zhang, C. Lu, Y. Wang, Q. T. Ruan, H. B. Ye, and W. H. Wang, "Identifiability Analysis of Load Model Parameters by Estimating Confidential Intervals," *CSEE Journal of Power and Energy Systems*, vol. 9, no. 5, pp. 1666–1675, Sep. 2023.
- [6] F. Aboytes and B. J. Cory, "Identification of measurement parameter and configuration errors in static state estimation," in *Proc. of the PICA Conf.*, 1975.
- [7] 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, Jul. 2014.
- [8] B. Appasani, A. V. Jha, S. K. Mishra, and A. N. Ghazali, "Communication infrastructure for situational awareness enhancement in WAMS with optimal PMU placement," *Protection and Control of Modern Power Systems*, vol. 6, no. 1, pp. 9, Mar. 2021.
- [9] Y. Weng, Y. Z. Liao, and R. Rajagopal, "Distributed energy resources topology identification via graphical modeling," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2682–2694, Jul. 2017.
- [10] Y. Z. Liao, Y. Weng, G. Y. Liu, Z. Y. Zhao, C. W. Tan, and R. Rajagopal, "Unbalanced multi-phase distribution grid topology estimation and bus phase identification," *IET Smart Grid*, vol. 2, no. 4, pp. 557–570, Dec. 2019.
- [11] J. J. Xu, Z. J. Wu, X. H. Yu, and C. Z. Zhu, "Robust faulted line identification in power distribution networks via hybrid state estimator," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 9, pp. 5365–5377, Sep. 2019.
- [12] O. Ardakanian, V. W. S. Wong, R. Dobbe, S. H. Low, A. Von Meier, C. J. Tomlin, and Y. Yuan, "On identification of distribution grids," *IEEE Transactions on Control of Network Systems*, vol. 6, no. 3, pp. 950–960, Sep. 2019.
- [13] Z. Tian, W. C. Wu, and B. M. Zhang, "A mixed integer quadratic programming model for topology identification in distribution network," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 823–824, Jan. 2016.
- [14] L. Zhao, Y. B. Liu, J. B. Zhao, Y. C. Zhang, L. X. Xu, Y. Xiang, and J. Y. Liu, "Robust PCA-deep belief network surrogate model for distribution system topology identification with DERs," *International Journal of Electrical Power & Energy Systems*, vol. 125, pp. 106441, Feb. 2021.
- [15] T. Erseghe, S. Tomasin, and A. Vigato, "Topology estimation for smart micro grids via powerline communications," *IEEE Transactions on Signal Processing*, vol. 61, no. 13, pp. 3368–3377, Jul. 2013.
- [16] G. Cavraro, R. Arghandeh, G. Barchi, A. Von Meier, "Distribution network topology detection with time-series measurements," in *Proceedings of 2015 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2015, pp. 1–5.
- [17] G. Cavraro and A. Bernstein, "Bus clustering for distribution grid topology identification," *IEEE Transactions on Smart Grid*, vol. 11, no. 5, pp. 4080–4089, Sep. 2020.
- [18] N. Zhou, L. G. Luo, G. H. Sheng, and X. C. Jiang, "Power distribution network dynamic topology awareness and localization based on subspace perturbation model," *IEEE Transactions on Power Systems*, vol. 35, no. 2, pp. 1479–1488, Mar. 2020.
- [19] J. W. Zhang, Y. Wang, Y. Weng, and N. Zhang, "Topology identification and line parameter estimation for non-PMU distribution network: A numerical method," *IEEE Transactions on Smart Grid*, vol. 11, no. 5, pp. 4440–4453, Sep. 2020.
- [20] G. N. Korres and N. M. Manousakis, "A state estimation algorithm for monitoring topology changes in distribution systems," in *Proceedings of 2012 IEEE Power and Energy Society General Meeting*, 2012, pp. 1–8.
- [21] X. He, R. C. Qiu, Q. Ai, and T. Y. Zhu, "A hybrid framework for topology identification of distribution grid with renewables integration," *IEEE Transactions on Power Systems*, vol. 36, no. 2, pp. 1493–1503, Mar. 2021.
- [22] 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, Jul. 2018.
- [23] X. D. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z. H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg, "Top 10 algorithms in data mining," *Knowledge and Information Systems*, vol. 14, no. 1, pp. 1–37, Jan. 2008.
- [24] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.
- [25] W. Jarosz, A. Enayet, A. Kensler, C. Kilpatrick, and P. Christensen, "Orthogonal array sampling for Monte Carlo rendering," *Computer Graphics Forum*, vol. 38, no. 4, pp. 135–147, Jul. 2019.
- [26] M. Farajollahi, A. Shahsavari, and H. Mohsenian-Rad, "Topology identification in distribution systems using line current sensors: an MILP approach," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1159–1170, Mar. 2020.
- [27] M. E. Baran and F. F. Wu, "Network reconfiguration in distribution systems for loss reduction and load balancing," *IEEE Power Engineering Review*, vol. 9, no. 4, pp. 101–102, Apr. 1989.
- [28] S. X. Wang, L. Han, and L. Wu, "Uncertainty tracing of distributed generations via complex affine arithmetic based unbalanced three-phase power flow," *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 3053–3062, Nov. 2015.



Dongliang Xu received the B.S degree in Electrical Engineering from Nanjing Normal University 2019. He is currently working toward the Ph.D. degree in Electrical Engineering with the School of Electrical Engineering, Southeast University, Nanjing, China. His research interests include state estimation and distribution network topology identification.



Zaijun Wu received the B.Eng. degree in Power System and Its Automation from the Hefei University of Technology, Hefei, China, in 1996, and the Ph.D. degree in Electrical Engineering from Southeast University, Nanjing, China, in 2004. He was a Visiting Scholar with the Ohio State University, Columbus, OH, USA, from 2012 to 2013. He is currently a Professor of Electrical Engineering with the School of Electrical Engineering, Southeast University. His research interests include microgrid, active distribution network, and power quality.



Qinran Hu received the B.S. degree from Chien-Shiung Wu Honors College, Southeast University, Nanjing, China, in 2010, and M.S. and Ph.D. degrees from the University of Tennessee, Knoxville, TN, USA, in 2013 and 2015, respectively, all in Electrical Engineering. He was a Postdoc Fellow with Harvard University, Cambridge, MA, USA, from 2015 to 2018. In October 2018, he joined the School of Electrical Engineering, Southeast University, Nanjing, China. His research interests include power system optimization, demand aggregation, and elec-

tricity market.



Junjun Xu received the B.S. degree in Power System and Its Automation from the Nanjing Institute of Technology, Nanjing, China, in 2012, and the M.S. degree in Agricultural Electrification and Automation from the School of Electrical and Information Engineering, Jiangsu University, Zhenjiang, China, in 2015, and the Ph.D. degree in Electrical Engineering from Southeast University, Nanjing, China, in 2019. He is currently an Associate Professor of Nanjing University of Posts and Telecommunications. His research interests include distribution network

state estimation, self-healing control, and uncertainty modeling approaches.