EAI Endorsed Transactions
on Energy Web

Research Article EAI.EU

Construction and Application Analysis of an Intelligent Distribution Network Identification System Based on Deep Neural Networks

Yu Ma[1,*]

[1]School of Information and Automation, Qilu University of Technology (Shandong Academy of Sciences), Jinan, 250353, China

Abstract

INTRODUCTION: At present, the communication between measuring data and network topology in the distribution system cannot be accurately established. Therefore, deep neural networks were utilized to learn the mapping relationship between the measurement data and network topology, achieving topology structure discrimination under different working conditions.
OBJECTIVES: This study aims to establish a machine learning-based Intelligent Distribution Network (IDN) online topology recognition model to address the limited measurement equipment in distribution networks and improve the accuracy and efficiency of network topology recognition.
METHODS First, light GBM was used for feature selection to reduce computational complexity and improve learning efficiency. Then, a DNN model was constructed for topological identification and enhances the model scalability through incremental and transfer learning mechanisms. In addition, the Cross-Validation Grid Search Algorithm (GSA) was used to optimize the hyperparameters to ensure that the model can achieve the optimal performance on different data sets. Finally, a new intelligent distribution network identification model (Intelligent Distribution Electricity Network Identification System, IDENIS) was constructed.
RESULTS: The study was experimentally verified on the distribution system of IEEE 33 and PG&E 69. The experimental results showed that the accuracy of the DNN-based model reached 0.9817 on the test set, while the accuracy after feature selection only decreased by 1.3%, and the features decreased by 81.8%. In the PG&E 69 node system, the features were reduced by 85.5%, while the identification accuracy was decreased by only 0.51%. These results demonstrated that the proposed method maintained high identification accuracy while reducing the computational resource consumption.
CONCLUSION: Its efficient computing speed fully meets the real-time requirements in practical applications. This paper provides new ideas and methods for achieving intelligent distribution network topology recognition of high proportion distributed power sources.

Keywords: Deep neural network; Online topology identification; Topological labels; Intelligent distribution network; Light GBM algorithm.

Received on 26 03 2024, accepted on 29 08 2024, published on 21 11 2024

doi: 10.4108/ew.5547

*Corresponding author. Email: mayu202401@126.com

1. Introduction

Electricity is currently the most widely used secondary energy source. However, its production still relies mainly on fossil fuel combustion. The development of smart grids has become a necessary path for the energy and power industry to improve the reliability and economy of the power grid to achieve large-scale access, consumption, and transmission of clean energy such as wind and electricity [1-2]. Deep Neural Network (DNN) is an effective machine learning method that can extract complex patterns from massive data and has strong adaptability and generalization ability [3-4]. DNN can excavate the nonlinear correlations between various operating states and influencing factors in distribution networks, providing important support for real-time identification of distribution networks [5-6]. The network topology of the distribution network may undergo dynamic changes due to factors such as equipment failures, maintenance work, and changes in energy demand. Network topology refers to the connections between the nodes and links in the network. Feature selection can not only reduce the computational complexity by reducing irrelevant or redundant features, but also improve the predictive performance and generalization ability of the model. Joint training is the simultaneous optimization of parameters for multiple related tasks or multiple models during training. The molecular features are used for automatic extraction and analysis. At present, it is impossible to accurately establish the correlation between measured data and network topology in most distribution network systems. In this study, DNN is used to learn the mapping relationship between measured data and network topology, realizing the topological structure identification under different working conditions. This experiment aims to provide new technical means for the operation and management of Intelligent Distribution Network (IDN) to further improve the operation efficiency and security of the power system.

The research mainly includes four parts. Firstly, an overview of the research background is provided, and a summary of the research in the relevant field is provided. Then, the establishment of an IDN identification system for DNN is elaborated. Secondly, experimental verification is conducted on the adaptive ability and other performance of the IDN identification system for DNN. Finally, a summary and outlook are provided for the entire study.

2. Related works

The distribution network topology identification system is an important tool for modern power operation and development. As the recognition framework of IDN develops and advances, more and more scholars are beginning to apply machine learning algorithms to various systems. DNN is also widely used in various fields.

Yin L et al. proposed a three-state energy consumer model to promote flexible scheduling of renewable energy. Then, an economic intelligent power generating controlling structure was designed to make the traditional multi-time scale structure replaced. The economic intelligent power generation control allowed consumers of three states of energy to enter and exit freely, improving economic dispatch's efficiency. In addition, a scalable adaptive dynamic programming means was put forward. This adaptive dynamic programming could match the entry and exit characteristics of three state energy consumers and had fine algorithmic accuracy and computational speed, while making network redundancy in adaptive dynamic programming reduced [7]. Lee Rd et al. conducted a survey on delivering systems, using neural enhancing technology for achieving quick response time et al. This paper introduced some composition of the content transferring systems, emphasized difficulties, using neural enhancing models as countermeasures [8]. Yang A et al. developed a multitasking DNN for automatic extraction of molecular features and integrated tree like Long Short-Term Memory (LSTM) with multiple feedforward neural networks for correlation analysis of multiple features. Molecular features were encoded, calculated, and extracted in the molecular tree diagram without the need for manual user operation or preliminary molecular descriptor calculation. The proposed multi-task DNN was trained using both joint training and alternative training methods, which could capture the relevant information and commonalities between multiple target characteristics [9]. Alizadeh Bidgoli M et al. proposed a system for the energy management of microgrids. First, each microgrid used historical data to predict the load demand of users. Then, the cooperating game means was utilized to achieve microgrid's scheduling and energy trading. A predicting model called deep learning Artificial Neural Network (ANN) was established using ANN and rough neural water cycle algorithm to predict uncertain parameters [10]. Kumari P et al. proposed a new hybrid deep learning model, LSTM-CNN, for hourly global level radiation (Global Horizontal Irradiance, GHI) prediction. Spatiotemporal features were modeled by integrating LSTM and Convolutional Neural Network (CNN). First, the proposed hybrid model used LSTM to extract temporal features, and then used CNN to extract spatial features from the correlation matrices of multiple variables at the target and its adjacent locations [11].

Xu Z et al. proposed a recognition method using transferring terminal unit measurement results to ensure that the topology of the managing system was consistent with the actual topology of the distributing network. It expressed the topology recognition problem as a minimizing problem. Adding variables to unavailable phase angles in minimization problems could lead to nonlinearity and non-convexity. Asynchronous sampling time could lead to sampling errors when surveying power injection fluctuations [12]. Distribution network monitoring may improve service levels by reporting the cause of fault events and informing the nature of remedial measures. Jiang X et al. proposed a new structural similarity measure applied to relevant power quality waveforms, which inferred the cause of faults from substation current data using a very small amount of historical fault data. This method improved classification accuracy compared with similar technologies [13]. Zhang W et al. proposed a fast error locating and isolating method for distributing networks. Three major techniques were introduced. They included a distribution network phase divider with permanent magnet operation structure, a phase selector based on transient current characteristics, and a temporary and permanent fault identification method based on power frequency voltage characteristics. These proposed strategies could avoid secondary impacts on upstream switches and power outages in upstream sections during permanent faults and shorten fault handling time [14]. Chen Y et al. proposed a fault localization method based on equivalent admittance distortion rate. The equivalent admittance distortion rate was obtained by measuring the equivalent admittance before and after the fault. A phase having the highest equivalent admittance distorting rate was treated as the faulty one. A feeder having the highest distorting rate was treated as the faulty one. A segment having the highest distorting rate was treated as the faulty one [15]. Dua G S et al. proposed a new method for detecting configurations by obtaining and processing real-time measurement values of optimized microwave measurement units. The placement of microwave measurement units was formulated based on many working configurations of this distributing network. Distribution network configuration recognition was solved by obtaining measurement data from nodes [16].

In summary, the identification of IDN occupies a core position in the operation of the power system. Traditional identification methods are mainly based on fixed mathematical models. Faced with the dynamic changes in the network topology of the distribution network, it is difficult for traditional identification methods to cope with the complexity and dynamism of the distribution network. This leads to the inability to accurately capture and establish the communication between measurement data and network topology in the distribution system. As deep learning technology develops, DNN has indicated significant advantages in handling nonlinear problems. Faced with topology discrimination problems under different operating conditions, DNN is used to explore the mapping relationship between measurement data and network topology. This is crucial for further research on IDN identification systems based on DNN.

3. DNN-based online topology identification system for distribution networks and grid searching algorithm for cross-validating

A machine learning-based IDN network topology recognition model is established based on these measurement and operational characteristics of the distributing network. The specific implementation steps are discussed in detail. Firstly, the data used in this experiment are analyzed. Second, an online identification method for network topology based on Light Gradient Boosting Machine-Gradient Boosting Decision Tree (Light GBM) is studied considering the limited measurement equipment in the distribution network.

3.1 DNN-based online topology identification system for distribution networks

Light GBM is an efficient and lightweight Gradient Boosting Decision Tree (GBDT). These two new data processing algorithms proposed in this study, GOSS and EFB, can significantly reduce computational complexity and improve learning efficiency while ensuring model accuracy. They are suitable for distributed power systems with massive and high feature dimensions, such as large distribution networks [17-18]. Light GBM, a histogram-based decision tree algorithm, is used to discretize continuous features into discrete histogram features, thus reducing the storage space and computational complexity of the data. This discretization approach can reduce the complexity of feature processing and enable better handling of high-dimensional sparse data. A sampling method called Gradient-based One-Side Sampling (GOSS) and a feature bundling method of Exclusive Feature Bundling (EFB) are used to make the model efficient and vertically parallelized computing when training. This speeds up the training of the model, which is especially effective when dealing with large-scale datasets. Conventional gradient lift tree algorithms are grown by layer. Light GBM uses a growth strategy called leaf-wise. The leaf-wise growth strategy selects the current optimal leaf node to split each time, which can find the direction where the loss function decreases fastest faster, thus accelerating the training of the model. On this basis, GOSS is proposed, which utilizes sampled data to model training without all samples. The second is based on mutually exclusive feature binding, which constrains multiple attributes to find the optimal solution, thus avoiding searching for each attribute separately. This method can greatly improve the model's learning effectiveness and make computing difficulty reduced while maintaining its accuracy. This study adopts machine learning methods to study the topology recognition of distribution networks from two aspects: offline learning and online application. Figure 1 is a topology identification framework based on machine learning.

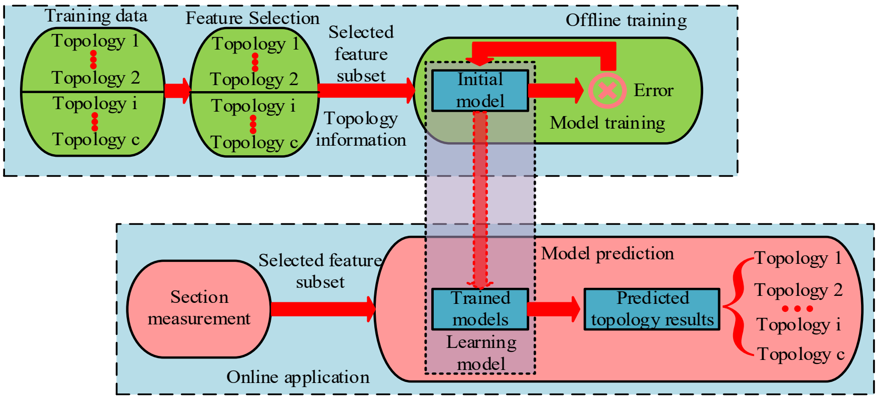


Figure 1. A topology identification framework based on machine learning

In Figure 1, the number of topologies for a given distribution network to function properly is determined. Various measuring devices are installed at certain key points. Each sampling point contains a set of time-domain measurement data for the operational variables, represented by colored solid lines to refer to the current network connectivity. In offline learning, a large amount of time-domain segmented measurement data with different topological structures are collected and used as training samples. On this basis, the original observation features are first selected for feature selection. The most efficient feature subset is selected. The corresponding topological structure is used as the output. On this basis, iterative optimization is carried out on the constructed model. In online applications, the cross-sectional measurement features under unknown topology are input into the trained model to obtain the corresponding network structure. The initialization of the gradient boosting decision tree algorithm is represented by formula (1).

$$f_0(x) = \arg\min_c \sum_{i=1}^{N} L(y_i, c) \quad (1)$$

In formula (1), $L(y_i, c)$ is the loss function. For $i = 1.2.....N$, the decision tree is represented by formula (2).

$$r_{mi} = -[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}]_{f(x)=f_{m-1}(x)} \quad (2)$$

In formula (2), for $j = 1.2.....J$, the decision tree is represented by formula (3).

$$c_{mj} = \arg\min \sum_{x_i \in R_{mj}} L(y_i, f_{m-1}(x_i) + c) \quad (3)$$

In formula (3), the updated network structure is represented by formula (4).

$$f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J} c_{mj} I(x \in R_{mj}) \quad (4)$$

The obtained regression tree is represented by formula (5).

$$\hat{f}(x) = f_M(x) = \sum_{m=1}^{M} \sum_{j=1}^{J} c_{mj} I(x \in R) \quad (5)$$

The output equation of ANN neurons is represented by formula (6).

$$y_j = f(\sum_{i=1}^{n} w_{ij} x_i + b_j) \quad (6)$$

Light GBM is used for feature selection. DNN is used for distribution network topology identification. Figure 2 shows the DNN structure.

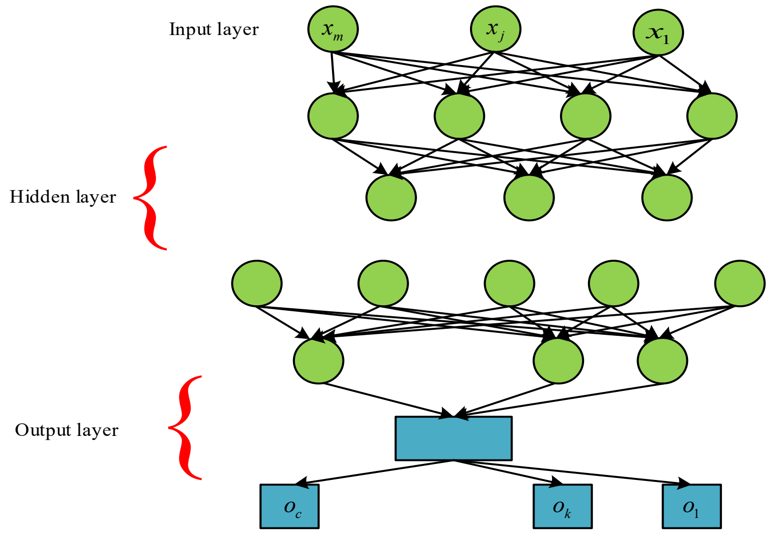


Figure 2. Schematic diagram of DNN structure

In Figure 2, the input feature values in the network are taken as the input parameters of the network. The number of network nodes is taken as the type of network structure. The ReLU function is chosen as the excitation function of the neural network to effectively solve the gradient vanishing and improve the computing efficiency of the algorithm. The output layer is activated using the standardized indicator function Softmax. For the input-output relationship in neural network learning, the i th hiding layer neuron's input is represented by formula (7).

$$net_i = \sum_{j=1}^{M} \omega_{ij} x_j + b_i \tag{7}$$

The i th hiding layer neuron's output is represented by formula (8).

$$o_i = \phi(net_i) = \phi(\omega_{ij} x_j + b_i) \tag{8}$$

The k th hiding layer neuron's input is represented by formula (9).

$$net_k = \sum_{i=1}^{q} \omega_{ki} y_i + a_k = \sum_{i=1}^{q} \omega_{ki} \phi(\sum_{j=1}^{M} \omega_k x_j + b_i) + a_k \tag{9}$$

The k th hiding layer neuron's output is represented by formula (10).

$$o_k = \psi(net_k) = \psi[\sum_{i=1}^{q} \omega_{ki} \phi(\sum_{j=1}^{M} \omega_k x_j + b_i) + a_k] \tag{10}$$

The adjusting direction of weight is represented by formula (11) as the negative gradient of the objective function.

$$\Delta\omega_{ij} = -\eta \frac{\partial L(E,O)}{\partial \omega_{ij}} \tag{11}$$

The updating rule is represented by formula (12).

$$\omega_{ij} \longleftarrow -\omega_{ij} + \Delta\omega_{ij} \tag{12}$$

On this basis, the forward transmission signal and the backward transmission error are iterated repeatedly until the network parameters and loss function tend to a series of stable constants.

3.2 Cross-validation grid search algorithm

In machine learning, hyperparameter optimization is the process of setting an optimal set of hyperparameters. The hyperparameter optimization is used for controlling the algorithmic learning and adjusting other parameters [19-20]. The best hyperparameter tuple obtained can make the prior loss function minimized in known data, thereby obtaining the most accurate prediction results. When establishing DNN, the quantity of layers and neurons in the network is first determined. Other hyperparameters are then set to make the algorithmic learning controlled. The hyperparameters of the network can be obtained using a Grid Search Algorithm (GSA) based on cross-validating. Each parameter's value is determined, and a GSA parameter is used. Figure 3 is a schematic diagram of GSA.

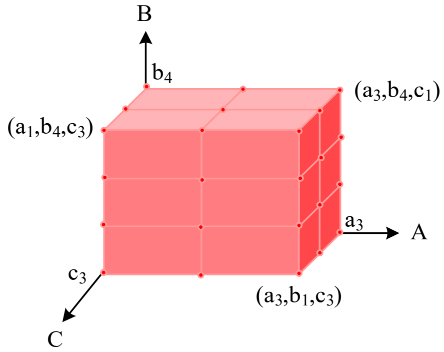


Figure 3. Schematic diagram of parameter grid searching

In Figure 3, each parameter's potential discrete values are arranged and combined to construct a parameter grid. There are three possible values for parameter A, four possible values for parameter B, and three possible values for parameter C. Therefore, there are a total of 36 parameter combinations for the parameter grid. For model performance evaluation, a 5-fold cross-validation method is used, and Figure 4 shows its principle.

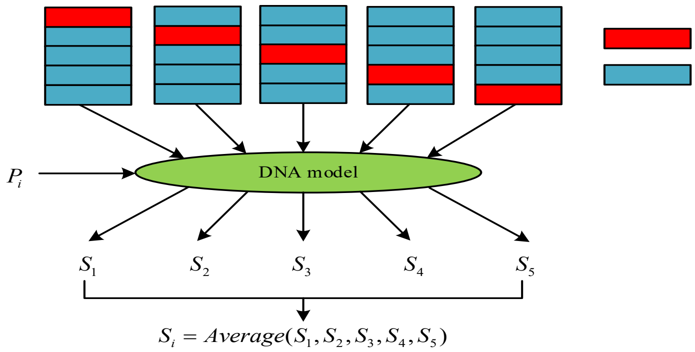


Figure 4. 5-fold cross-validation principle

In Figure 4, the dataset is randomly divided into 5 equal parts. One of the data is selected as the validation set without repetition each time. The remaining data are used as the training set, resulting in five pairs of datasets. The performance score of the learning model is obtained by using the allocated dataset for training and validation. The performance score of the parameter combination is the average score of the five sets. A new method for identifying unknown topological structures is proposed. This method includes two aspects: first, utilizing the variance between samples for unknown topological identification, and second, incremental learning. Figure 5 shows a positional topology processing mechanism based on minimum variance.

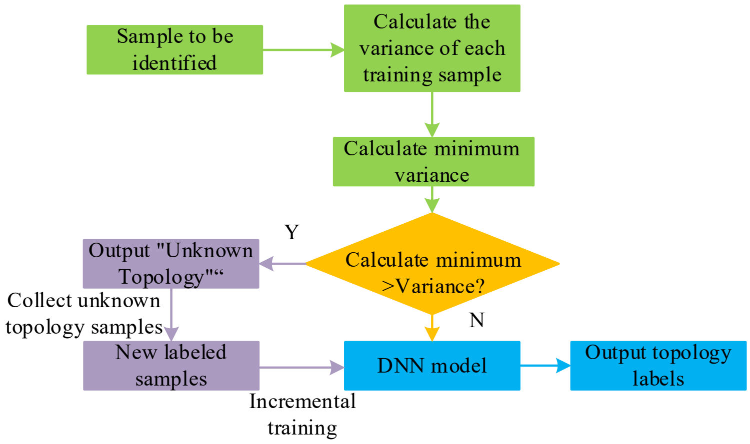


Figure 5. Flowchart of unknown topology processing mechanism based on minimum variance

In Figure 5, the identification samples and the variance of each training set are first calculated, followed by calculating the minimum variance and comparing them. When the sample variance > the minimum variance, the sample corresponds to a certain topology in the training set and is identified with DNN. When the sample variance <= the minimum variance, the sample corresponds to an unknown topology. Measurement samples are collected from unknown topology structures and annotated. Based on this, existing deep learning networks are learned to update the network topology structure. It is possible to promptly discover previously overlooked or newly generated network operating topologies by adopting the processing mechanism. The existing DNN is continuously revised and improved through an incremental learning mechanism, dynamically updating the topology knowledge base. This applies to IDNs with flexible and ever-changing topology and constantly emerging new operating scenarios. The minimum-maximum normalization method is selected for data standardization, and its transformation function is represented by formula (13).

$$\widehat{f_i} = \frac{f_i - f_i^{\min}}{f_i^{\max} - f_i^{\min}} \qquad (13)$$

In formula (13), $f_i^{\max}$ and $f_i^{\min}$ refer to feature points' maximum and minimum values, respectively. f_i is the feature's initial value. $\widehat{f_i}$ is the standard characteristic value. The definition of DNN is represented by formula (14).

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \qquad (14)$$

The cross entropy in DNN is represented by formula (15).

$$H(p,q) = -\sum_x p(x)\log q(x) \qquad (15)$$

In formula (15), p is the probability distribution. q is the actual output. The research introduces the incremental learning based on the semi-supervised paradigm to increase the scalability of the IDN topology identification system based on DNN. This mechanism allows the model to learn and adapt to new network topologies in real-time at runtime by utilizing partially labeled data in the form of pseudo supervision. In addition, transfer learning is used to pre-train models on similar but smaller networks and reuse these pre-trained knowledge to initialize model training on larger and more complex networks. The knowledge learned from an environment is used to help with the learning tasks in a new environment. This improves the training efficiency and convergence speed of the model on the new network. Meanwhile, the adaptability of the model is enhanced to the identification accuracy of the new network topology, thus further enhancing the scalability and practicability of the system. Besides, the model is integrated into the existing power system management infrastructure. Then, it is necessary to develop the development data universal interface to ensure that the model can receive and process the measurement data from the existing system. Finally, an IDENIS model that combines Light GBM, DNN, and GSA is constructed. Light GBM is responsible for feature selection. The improved DNN is responsible for the topological recognition model. GSA is used to optimize the hyperparameters of DNN.

The performance of the model may be affected under conditions of imbalanced data. If certain topological structures are not adequately represented in the training set, the model may have difficulty accurately identifying these minority categories, leading to a decrease in classification performance. To address these issues, research is conducted to increase the cost of minority class classification errors and to achieve a balance of data during training by discarding some majority class samples. The parameter class_weight='balanced' is used to automatically adjust the weight to penalize misclassification of minority classes.

4. Feature selection results and comparative analysis based on Light GBM

The correctness and superiority of the IDENIS algorithm were demonstrated through numerical examples and comparisons to verify the feature selection method's effectiveness based on Light GBM. Taking the IEEE33 and PG&E69 node distribution networks as research objects, this study focused on typical distributed energy access scenarios and conducted research and experiments on two topologies: radiative and weak loop networks.

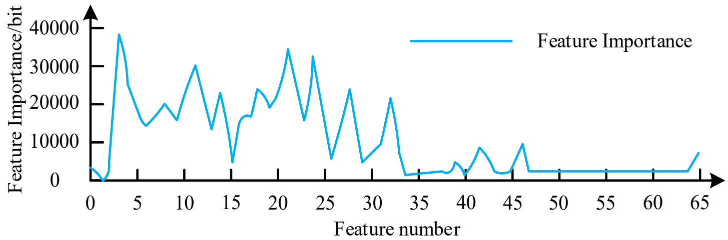
4.1 Feature selection results and comparative analysis based on Light GBM in IEEE33 node distribution system

In the running system, 28 representative topology structures to be identified were selected and divided into 0-27. 0-19 mean a radial structure, and 20-27 refer to a grid-like structure. Under each topology, 1500 sampling points were simulated, with a total of 42000 samples. The features included the time-domain cross-sectional measurements of voltage amplitude and active power of system nodes, totaling 66 dimensions. The training set accounted for 80%, and the testing set was 20%. Table 1 presents typical topology line conditions and label diagrams in the IEEE33 node system.

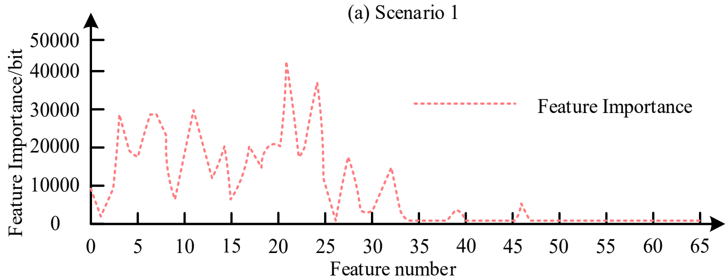
Table 1. IEEE33 node system typical topology

Line disconnection		Topology labels	Line disconnection		Topology labels	Line disconnection		Topology labels
Open branch	Connecting branches	/	Open branch	Connecting branches	/	Open branch	Connecting branches	/
/	/	0	4-5	8-21	10	/	25-29	20
32-33	18-33	1	9-10 14-15	9-15 12-22	11	/	12-22	21
11-12	12-12	2	3-23	18-33	12	/	8-21	22
14-15	9-15	3	31-32	25-29	13	/	18-33	23
7-8	8-21	4	6-26 8-9	25-29 12-22	14	/	9-15	24
28-29	25-29	5	7-8 11-12	8-21 12-22	15	20-21	8-21 12-22	25
17-18	18-33	6	3-4	12-22	16	13-14	9-15 18-33	26
24-25	25-29	7	3-4	25-29	17	11-12	12-22 18-33	27
10-11 12-13	12-22 18-33	8	5-6	25-29	18	/ /	/ /	/ /
2-19	12-22	9	5-6 29-30	12-22 18-33	19	/ /	/ /	/ /

In Table 1, it is possible to understand the characteristics and relationships of different line breaks and topology labels to better understand and analyze the lines and topology structure in the network. A 5-fold cross-validating GSA was used to obtain the optimal parameters of Light GBM. Two scenarios were defined. Scenario 1: The decision trees are 1000, with a maximum depth of 11 for each tree and a learning rate of 0.1. Scenario 2: The decision trees are 1000, with a maximum depth of 9 for each tree and a learning rate of 0.1. Figure 6 presents the feature importance calculated by the Light GBM-based feature selection method in the IEEE33 node system.



(a) Scenario 1



(b) Scenario 2

Figure 6. IEEE33 node system characteristic importance

Figure 6 shows the importance analysis of the characteristics of the IEEE33 node system. 0 to 32 are voltage amplitude characteristics. 33 to 65 are active characteristics. The importance of each function varied greatly. Features with low importance had little impact on structural recognition and could be excluded in structural recognition. Overall, the importance of voltage amplitude characteristics was much greater than active power, indicating that topology identification using node voltage amplitude was more effective. This conclusion was consistent with the physical mechanisms of voltage and power distribution at various nodes in the power grid. However, no matter how the topology of the distribution system changed, the power of a node was determined by the load it was connected to. The power of a node was not affected by the wiring mode, making it difficult to truly reflect the topology characteristics of the network. The study compared it with some other common machine learning algorithms to verify the advantages of DNN. Each algorithm was optimized using a GSA that had undergone cross-validation in the corresponding scenario. Starting from all features, these features were reduced in order of importance to obtain the corresponding subset of features to observe the relationship between the quantity of selected features and the accuracy. Then, the learned model was trained separately. Figure 7 shows the communication between the quantity of features and the accuracy of the test set.

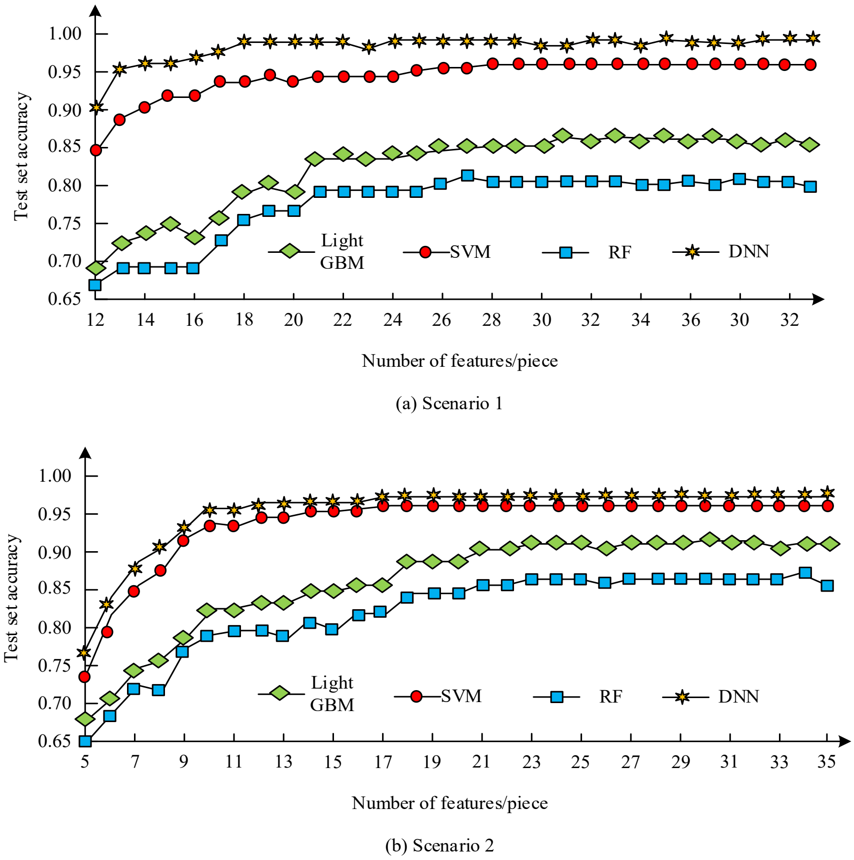


Figure 7. The relationship between the accuracy of four learning algorithm test sets and the number of features

In Figure 7, as the features decreased, the accuracy of test cases showed a decreasing trend. These experiments confirmed that DNN-based models had the largest training samples and the best topology recognition ability. All four training methods could meet the needs of online applications. In Figure 7 (a), the accuracy of the test set reached 0.9817. After the features exceeded 16, the accuracy of the test samples remained stable. However, when the features were less than 16, the accuracy of the test samples dropped sharply. In Figure 7 (b), compared with the original feature set, the extracted feature subset required 81.8% less features, while the recognition accuracy only decreased by 1.3%, reaching 0.9724. Compared with Support Vector Machine (SVM), this algorithm improved accuracy by 1.6% when the number of features was large. When the number of features was small, it decreased by about 3.6%.

4.2 Feature selection results and comparative analysis based on Light GBM in PG&E 69 node distribution system

Under supervised learning, the gradient descent method was used to minimize the model error and optimize the parameters of the neural network. In each iteration of the loop, the loss function of each loop changes in a gradient manner, thereby controlling the learning rate in Figure 8.

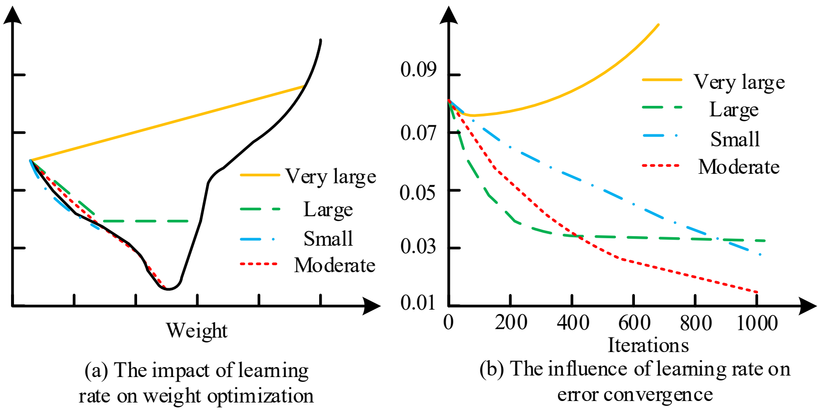


Figure 8. The impact of different learning rates on gradient descent

In Figure 8, the starting point of the training was set at the vertex to the left of the error curve. Overfitting might occur if the learning rate was below 0.03. Then, the network parameter updated slowly, and the training cost was slower. The lowest point and the optimal parameters of the curve were well found with the appropriate learning rate. When it was higher than 0.05, the system parameters oscillated around the optimization solution, but they could not reach the optimal solution. At this point, a large number of parameter updates were required. As iterations increased, the loss function also increased, making it difficult to find the optimal parameters. Therefore, the learning rate should be set within an appropriate range. In practical applications, continuous hyperparameters had infinite possible values, while discrete hyperparameters reduced the search space, making the search process more efficient. Using a 5-fold cross-validating GSA, Light GBM's best parameters were obtained and two scenarios were defined. Scenario 1: The number of decision trees is 1500, with a maximum depth of 20 for each tree and a learning rate of 0.05. Scenario 2: The number of decision trees is 2000, with a maximum depth of 18 for each tree and a learning rate of 0.1. Figure 9 presents the feature importance calculated by the Light GBM-based feature selection method in the PG&E 69 node system.

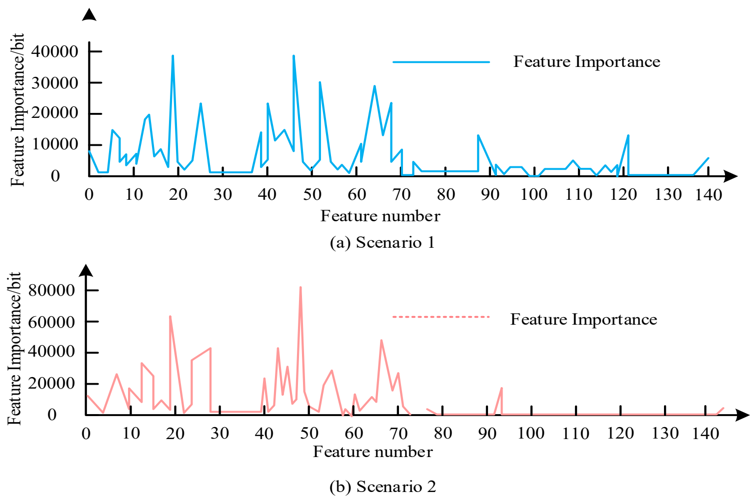


Figure 9. Feature importance of PG&E 69 node system

In Figure 9, numbers 0-68 refer to voltage amplitude characteristics, and 69-137 refer to active power characteristics. The importance of some active power features is 0, which has no impact on topology identification. The reason is that these nodes do not contain loads or power generation equipment. The active power injection is always 0. The feature selection process was used to study the sorting of features based on their importance after evaluating the importance of features, gradually reducing the number of features to observe their impact on the model performance. Through this approach, researchers could determine which features are redundant (i.e., the information they provide can be replaced by other features) or have less information (i.e., their contribution to model prediction is minimal). Figure 10 shows the performance of four learning algorithms after training.

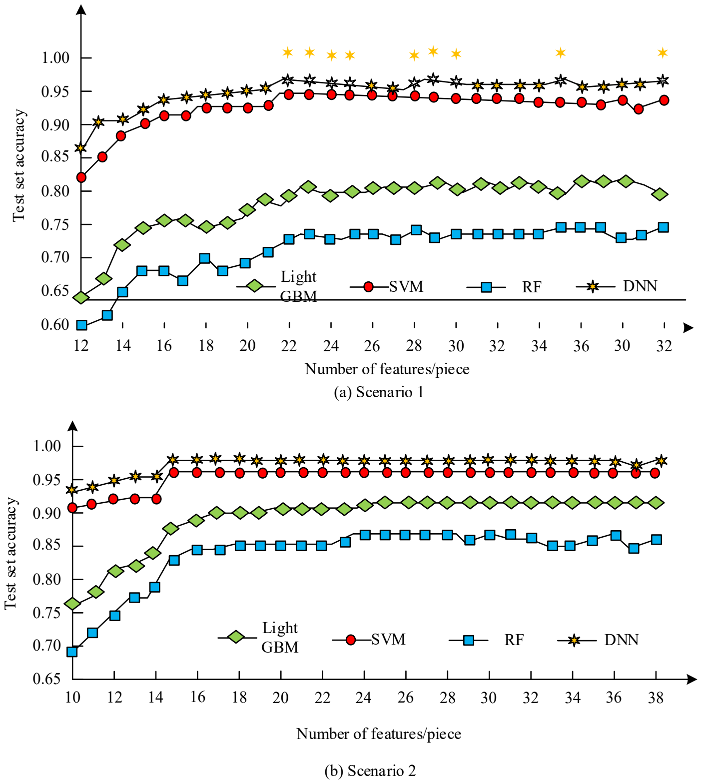


Figure 10. The relationship between the accuracy of four learning algorithm test sets and the number of features

In Figure 10 (a), the required features were reduced by 85.5% compared to the original feature set, while the identification accuracy could reach 0.9738, only a decrease of 0.51%. The Light GBM algorithm determined the importance of the feature by training the model and evaluating the contribution of each feature in the model. This allowed Light GBM to identify which features influenced the model predictions most and which features might be less important. Then, the feature selection process was reduced. When using the original feature set for training, the test set's accuracy was 0.9788. Considering the trade-off between the number of features and classification accuracy, the top 20 important features were selected to form a feature subset. Compared to SVM, DNN had an accuracy increase of about 1.5% when there were more features, and about 2.1% when there were fewer features. In Figure 10 (b), its accuracy in the test set was 0.9795. The top 15 important features were selected to form a feature subset. The required number of features was reduced by 89.1% compared to the original feature set. The identification accuracy could reach 0.9717, only a decrease of 0.8%. Compared to SVM, DNN had an accuracy increase of about 1.1% when there were more features, and about 2.4% when there were fewer features. Four learning models were trained using the selected feature subset. Table 2 shows the accuracy and calculation time of the test set. The calculation equation for its accuracy is shown in formula (16).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{16}$$

In formula (16), TP is the number of samples correctly predicted as positive. TN represents the number of samples correctly predicted as negative. FP is the false positive case FN indicates the number of samples falsely predicted as negative. Accuracy directly reflects the classification performance of the model, which is a key indicator for measuring the accuracy of model prediction. The calculation time reflects the operational efficiency of the model.

Table 2. Performance of four learning algorithms

Learning algorithm	Scenario 1		Scenario 2	
	Test set accuracy	Calculation time/s	Test set accuracy	Calculation time/s
RF	0.7439	7.06×10^{-4}	0.8223	8.07×10^{-4}
Light GBM	0.8022	1.07×10^{-5}	0.8697	4.97×10^{-4}
SVM	0.9557	4.22×10^{-4}	0.9611	$2.73\times10\text{-}^{4}$
DNN	0.9738	2.93×10^{-6}	0.9717	9.26×10^{-6}

In Table 2, applying Light GBM reduced the required number of features by 89.1% compared to the original feature set. The identification accuracy could reach 0.9717, only a decrease of 0.8%. Compared to SVM, DNN had an accuracy increase of about 1.1% when there were more features, and about 2.4% when there were fewer features. Multiple experimental groups were designed to verify the strong robustness of the IDENIS-based online topology recognition system for distribution networks against noise, measurement missing, and other situations. Firstly, the IDENIS model (Group 1) without introducing any noise or measurement missing information was applied. Subsequently, the uncertainty in actual operation was simulated by gradually increasing Gaussian noise (Group 2 to Group 4), random measurement missing (Group 3 to Group 4), and extreme salt and pepper noise and measurement missing (Group 5). In addition, traditional machine learning algorithms such as SVM (control group A) and Random Forest (RF, control group B) were introduced and evaluated to compare the performance of the IDENIS model under the same noise and measurement missing conditions. The evaluation indicators include accuracy, precision, recall, F1 score, and calculation time to comprehensively evaluate the performance of different models under different conditions. The experimental results are shown in Table 3.

Table 3 Experimental results of model robustness testing

Group	Noise type	Measurement missing rate	Accuracy	Recall	F1 score	Computation time (s)
Group 1	Nothing	0%	0.9817	0.9790	0.9824	0.05
Group 2	Gaussian noise	5%	0.9780	0.9750	0.9780	0.06
Group 3	Nothing	10%	0.9700	0.9675	0.9700	0.07
Group 4	Gaussian noise	10%	0.9650	0.9630	0.9650	0.08
Group 5	Pretzel noise	20%	0.9500	0.9470	0.9495	0.10
Comparison group A	Pretzel noise	20%	0.9400	0.9250	0.9300	0.09
Comparison group B	Pretzel noise	20%	0.9200	0.9100	0.9125	0.11

According to Table 3, the IDENIS model exhibited strong robustness in dealing with uncertainty factors such as noise and measurement missing. In the absence of noise and missing measurements (Group 1), the IDENIS model achieved high accuracy, recall, and F1 score. With the increase of noise and measurement missing, although the model performance had declined, it still maintained a relatively stable level. Especially under extreme conditions (Group 5), the IDENIS model can still maintain higher accuracy, recall, and F1 score compared to traditional SVM (Group A) and RF (Group B), demonstrating its superior robustness. In addition, the IDENIS model also performed well in terms of computation time, maintaining relatively efficient computation speed even under complex conditions, further proving its effectiveness in online topology recognition tasks in distribution networks.

5. Conclusion

The IDN network's topology structure will continue to change to ensure the safe and economical operation of the power system. In response to the lack of measurement equipment in distribution networks, an IDN topology recognition framework suitable for large-scale distribution networks was constructed. This was transformed into a multi-class classification problem under a machine learning framework to solve the topology characteristics of radial and weak loop networks in distribution networks. A new neural network-based IDN network topology recognition algorithm was proposed in this study. The proposed algorithm was experimentally validated using the IEEE33 node system. Its adaptive ability was studied under different noise levels and missing measurement data conditions. These experiments confirmed that the proposed method could identify the real-time topology of the distributing network only by measuring the instantaneous cross-section of local nodes. Its computational speed was fast and could meet the real-time requirements. Compared to SVM, DNN had an accuracy increase of about 1.1% when there were more features, and about 2.4% when there were fewer features. However, this method still has some shortcomings, such as the high cost of manually annotating the data used in research. On this basis, future work will delve into how to achieve automatic topology labeling of measurement data, thereby improving the adaptability and promotion ability of topology recognition. Future methods can automatically annotate measured data to reduce the cost of manual annotation and further optimize the feature set.

References

[1] Moise I M, Meltzer V, Pincu E. A study on pellets as an alternative source of energy for fossil fuel using adiabatic combustion calorimetry. Revue Roumaine de Chimie, 2020, 65(2):211-215.

[2] Hua W, Chen Y, Qadrdan M, Jang J, Sun H, Wu J. Applications of blockchain and artificial intelligence technologies for enabling prosumers in smart grids: A review. Renewable & sustainable energy reviews, 2022, 161(6):112308-112320.

[3] Fan D, Ren Y, Feng Q, Liu Y, Wang Z, Lin J. Restoration of smart grids: Current status, challenges, and opportunities. Renewable and Sustainable Energy Reviews, 2021, 143(5):110909-110925.

[4] Wang K, Mao W, Song H, Evinemi E I. A multi-data training method for a deep neural network to improve the separation effect of simultaneous-source data. Geophysical Prospecting, 2023, 71(1):63-84.

[5] Chen Y H, Lin W T, Liu C W.Image recognition of interference fringes in polishing by convolutional neural network with data augmentation by deep convolutional generative adversarial network. Optical Engineering, 2022, 61(4):102-114.

[6] Lee S H, Yu W F, Yang C S. ILBPSDNet: Based on improved local binary pattern shallow deep convolutional neural network for character recognition. IET image processing, 2022,16(3):669-680.

[7] Yin L, Luo S, Ma C. Expandable depth and width adaptive dynamic programming for economic smart generation control of smart grids. Energy, 2021, 232(3):120964-120936.

[8] Lee R, Venieris S I, Lane N D. Deep Neural Network-based Enhancement for Image and Video Streaming Systems: A Survey and Future Directions. ACM computing surveys, 2022, 54(8):169-198.

[9] Yang A, Su Y, Wang Z, Jin S, Ren J, Zhang X, Shen W, Clark J H. A multi-task deep learning neural network for predicting flammability-related properties from molecular structures. Green Chemistry, 2021, 23(12):4451-4465.

[10] Alizadeh Bidgoli M, Ahmadian A. Multi-stage optimal scheduling of multi-microgrids using deep-learning artificial neural network and cooperative game approach. Energy, 2022, 239(6):122036-122050.

[11] Kumari P, Toshniwal D. Long short term memory–convolutional neural network based deep hybrid approach for solar irradiance forecasting. Applied Energy, 2021, 295(8):117061-117081.

[12] Xu Z, Jiang W, Xu J, Wang D, Wang Y, Ou Z. Distribution Network Topology Identification Using Asynchronous Transformer Monitoring Data. IEEE Transactions on Industry Applications, 2023,59(1):323-331.

[13] Jiang X, Stephen B, Mcarthure S. Automated Distribution Network Fault Cause Identification With Advanced Similarity Metrics. IEEE Transactions on Power Delivery, 2021, 36(2):785-793.

[14] Zhang W, Chang Z, Zhang C, Song G, Tan W. A quick fault location and isolation method for distribution network based on adaptive reclosing. IET Generation, Transmission & Distribution, 2022, 16(4):715-723.

[15] Chen Y, Yin J, Li Z, Wei R. Location for single-phase grounding fault in distribution network based on equivalent admittance distortion rate. IET Generation, Transmission & Distribution, 2021, 15(11):1716-1729.

[16] Dua G S, Tyagi B, Kumar V. A Novel Approach for Configuration Identification of Distribution Network Utilizing μPMU Data. IEEE Transactions on Industry Applications, 2021, 57(1):857-868.

[17] Tiwari D, Bhati B S, Nagpal B, Sankhwar S, AI-Turjman F. An enhanced intelligent model: To protect marine IoT sensor environment using ensemble machine learning approach. Ocean engineering, 2021, 242(11):110180-110190.

[18] Kunzweiler F, Biltzinger B, Greiner J, Burgess J M.Automatic detection of long-duration transients in Fermi-GBM data. Astronomy and astrophysics, 2022,665(3):22-31.

[19] Du X, Xu H, Zhu F. Understanding the Effect of Hyperparameter Optimization on Machine Learning Models for Structure Design Problems. Computer-Aided Design, 2021, 135(5):103013-103028.

[20] Mokayed H, Quan T Z, Alkhaled L, Sivakumar V. Real-time human detection and counting system using deep learning computer vision techniques. Artificial Intelligence and Applications. 2023, 1(4): 221-229.