Distribution Network Electrical Topology Identification Algorithm Based on Deep Learning

Ming Qu¹, Tao Wang¹, Fangshuo Li¹, Lina Liu¹, Qilin Li¹, Hailian Long¹, Yuhang Zhang², Jie Xu²*

¹ State Grid Sichuan Electric Power Corporation Metering Center, Chengdu, Sichuan, 610045, China
² School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, 611731, China
*Corresponding author’s e-mail: xuj@uestc.edu.cn

Abstract. With the rapid development of smart grids, power data can be collected online and increase in a large amount. Data problems such as missing data, abnormal data have become more and more prominent. Electrical topology is the basis for the implementation of the distribution network. The further development of the smart grid makes new requirements for the accuracy of the electrical topology identification algorithm. Deep learning relies on a large number of data inputs for feature extraction, and shows great tolerance for data missing and fluctuations. This paper proposes a two-channel 1DCNN (One-Dimensional Convolutional Neural Network) model for the electrical topology identification. In this model, voltage and current data are respectively input in each channel for feature extraction with a two-layer stacked CNN. And BN (Batch Normalization) layer and ReLU (Rectified Linear Unit) are added behind each CNN to help convergence. The experiment results show that the proposed model has good accuracy in electrical topology identification.

1. Introduction

With the rapid development of the Ubiquitous Power IoT, collection devices such as smart meters have achieved widespread access. The growth of data types and amount has brought huge pressure to communications networks, data storage and data processing. At the same time, with the development of the "last mile" in the distribution network, more and more high-performance devices are connected in the low-voltage side, providing hardware conditions for carrying out more extensive research. On this basis, researches emerge based on edge computing for distribution Internet of Things. Edge computing decentralizes part of the functions to the edge nodes from the central server. Data processing can be realized on the edge computing equipment, which reduces the pressure of data transmission and data processing, and improves the work efficiency of the entire distribution system [1]. As the foundation of the high-level applications in the distribution network, the identification of the electrical topology is particularly important. How to realize the topology identification more accurately and efficiently at the edge side is the key problem. As one of the main representatives of artificial intelligence, deep learning has superior prospects under the framework of edge computing.

Deep learning aims to make the machine self-learn the inherent laws of data, and automatically extract data features through neural network. The learning effect depends on a large number of input samples, which is in line with the trend of the expanding scale of power data. In addition, there are often
a lot of wrong data and bad data in the collected data in application scenarios. Deep learning has a strong
tolerance for the uncertainty of data, and it can adaptively identify and correct bad data. It has shown
great advantages in the practical application of distribution network topology identification [2].

The structure of this paper is as follows: section 2 introduces the related work of topology
identification and deep learning; section 3 shows the datasets and deep learning network model used in
this paper; section 4 compares and analyses the multiple test experiments conducted in our research;
section 5 is the conclusion of this paper.

2. Related work
In terms of the electrical topology identification of the distribution network, the traditional method that
relies on manual inspection is inefficient and cannot guarantee the topology update timely. As a result,
the real topology is inconsistent with the files in the central server, and advanced applications such as
station area line loss rate analysis and fault location cannot be effectively carried out. To solve these
problems, a series of researches on topology identification methods have emerged.

Zhang et al. [3] presented a topology analysis method with μPMU and SCADA measurement data.
Chen et al. [4] designed a phase identification system (PIS) for phase measurement based on a fuzzy
microprocessor-based controller. Both of the two methods can achieve topology identification with high
accuracy, but they all additionally depend on specific devices at the user-side. Zhang et al. [5] proposed
a method based on the customer voltage profile similarity by calculating Pearson correlation coefficient.
Chen et al. [6] used a two-stage method to identify the connection status between the transformers and
the feeders. Based on the data association analysis, Yang et al. [7] solved the problems of large amount
of calculation, inaccuracy of calculation results, and inability to verify based on the existing big data
mining methods. Lu et al. [8] proposed a method for topology check based on fuzzy C-Means clustering
algorithm.

All the methods above have got good effect on topology identification, but the accuracy is based on
complex pre-processing of data. Under the rapid growth of grid data, the efficiency of these methods
has been affected. As an advanced application form of the smart grid, the Ubiquitous Power Internet of
Things make new requirements for the data processing performance and computing capabilities. Deep
learning stands out due to its strong tolerance for data quality. Bagheri et al. [9] pointed out that deep
learning is very suitable for processing big data in smart grids, and explained the application potential
of deep learning. Chen et al. [10] regarded the network topology as a graph, and built a deep neural
network with load as input. Luo et al. [11] proposed a deep learning-based DC distribution network fault
location method. Gu Haitong et al. [12] construct an identification model of the topological relationship
between households in the station area based on carrier data and CNN-LSTM. All the methods above
are the achievements of deep learning obtained in various applications in the distribution network,
verifying the application value of deep learning in the context of big data in the smart grid.

Considering the data characteristics and collection conditions of power data, this paper proposes an
algorithm of electrical topology identification of distribution network based on deep learning with multi-
source heterogeneous data. With reference to the one-dimensional convolutional neural network
(1DCNN), we propose a two-channel 1DCNN model to extract features respectively from the voltage
and current data of the low-voltage side users, and then use the convolutional layer again for deep feature
fusion. And we verify the accuracy of the model through multiple sets of experiments.

3. Two-channel 1DCNN model
CNN is a type of feedforward neural network with convolution calculations and deep structure. It is one
of the representative algorithms of deep learning [13]. Since ImageNet in 2012, deep learning has got
more and more attention. CNN has become a conventional method in the field of computer vision and
has created great value. For the two-dimensional data of computer vision research, we often use the two-
dimensional CNN (2DCNN) model. For one-dimensional data, we need to make dimensional
conversion to use 2DCNN, which is complicated and ineffective. In 2015, Kiranyaz et al. [14] proposed
an adaptive model of 1DCNN when analysing patient ECG data for the first time. Since then, 1DCNN has been widely used in one-dimensional data processing.

1DCNN is often used in sequence models and natural language processing. It is suitable for analysing and processing time series data and periodic data. 1DCNN can mine the relationship between multi-source heterogeneous data and extract the deep features of power data. How to choose model and effectively extract features from multi-source heterogeneous data is the key of topology identification based on deep learning.

Figure 1. The two-channel 1DCNN model

The topology identification model structure of the two-channel 1DCNN is shown in figure 1. With sufficient data, deep learning shows great tolerance to missing data and abnormal data. Therefore, we directly input the real data sampled from distribution network into our model, and extracts features from the voltage and current data through two channels respectively. At the same time, to enhance the robustness of the two-channel model and prevent over-fitting, we add Additive Gaussian Noise before the first convolutional layer, which also improves model's adaption of data fluctuations in practice.

Figure 2. Feature extraction module

There are two modules in our model: feature extraction module and feature fusion module. Figure 2 specifically shows the structure of the feature extraction module. Considering the scale of distribution network data and efficiency of algorithm operation, we use two-layer stacked 1DCNN+BN+ReLU. Among them, we add a batch normalization (BN) layer before the activation function layer of each convolution block, which can speed up the training and convergence process, prevent overfitting of training [15], and improve the performance of subsequent activation layers. The activation function layer uses a Rectified Linear Unit (ReLU) function, which can effectively avoid the problem of gradient disappearance and make sure that the model has a stable convergence rate.

In the feature fusion module, it is difficult to highly fuse the features of two channels only through data splicing or addition and multiplication. Therefore, after the two-channel feature maps are stitched
together, we add 1DCNN+BN+ReLU convolutional layer again to achieve feature fusion, and finally completes the classification task through SoftMax function.

4. Experiments
This section firstly introduces the datasets and model used in the experiments of this paper, then verifies the effectiveness and stability of the model through comparative experiments. The experiments have two parts: compare the accuracy of the single-channel model and the two-channel model to show the effect of current data; compare the accuracy of deep learning and other algorithms based on raw data (data without processing such as missing value filling and outlier elimination).

4.1. Datasets
This paper uses power data in a district of Sichuan Province to make the distribution network topology identification dataset, which included 4483 users under 22 transformers (22 station areas). The user’s voltage and current data are sampled at the same time. The sampling interval is 15 minutes and the duration is 1 day. Each user’s data forms a sample, the station area ID and phase of the user is the label of sample according to the file record of central server. The data dimension is 2×96=192.

4.2. Implementation details
Our experiment uses Ubuntu16.04.6 system, and CPU is Intel(R)Core (TM)i7-7700K CPU@4.20GHz, and GPU is NvidiaGeforce1080Ti with 11GB video memory. All the experiments are completed based on an open source framework of deep learning named Keras. Keras is a highly modular neural network library developed by Python, which can complete the rapid development of deep learning on platforms such as Tensorflow and Theano.

The single-channel 1DCNN model used in this article is shown in figure 3. The number of convolution kernels in each layer is 32, 32, 64, and the size of the first two kernels is 3, the last one is 1.

Figure 3. 1DCNN model

The parameters of each layer in the two-channel 1DCNN model are shown in table 1. For the parameter optimizer, we use Adam to minimize the objective function, the initial learning rate is set to 0.001, the exponential decay rate of the first-order moment estimation is 0.9, and the exponential decay rate of the second-order moment estimation is 0.999. The training batch size used in our experiment is 10 and the epoch is 200. All models are tested on the datasets mentioned in Section 4.1.

| Module              | Layer       | Parameters                             |
|---------------------|-------------|----------------------------------------|
| Feature extraction  | Gaussian noise | $\mu = 0, \sigma = 0.5$                |
|                     | CNN1        | Core size = 3, Core number = 32, $l_2$ regularizarion = 0.02 |
|                     | CNN2        | Core size = 3, Core number = 32, $l_2$ regularizarion = 0.02 |
| Feature fusion      | CNN         | Core size = 1, Core number = 64, $l_2$ regularizarion = 0.02 |
|                     | Dropout     | Dropout = 0.2                          |

4.3. Results and Comparisons
Based on the multi-source heterogeneous data collected from the distribution network, we select voltage and current as the two-channel input, and voltage as the single-channel input, to analyse the performance of single-channel and two-channel model and the influence of two types of data.
Table 2. The comparison of the identification accuracy of single-channel and two-channel 1DCNN

| Model          | Accuracy | Recall  | AUC  |
|----------------|----------|---------|------|
| Single-channel | 99.76%   | 99.71%  | 100% |
| Two-channel    | 99.82%   | 99.78%  | 100% |

In the experimental results of table 2, we can see that two models have high accuracy both, and the two-channel model has higher accuracy and recall. It shows that current data can help improve identification effect, and verify the stability and generalization ability of our model. It is more conducive to the application of identification algorithms in real scenarios.

Traditional electrical topology identification algorithms mostly rely on additional data collection and cannot realize remote identification. At the same time, as the amount of data increases, problems such as missing data and abnormal data become more prominent. Topology identification algorithms based on voltage correlation need to fill the missing point and reduce the dimensionality of data to achieve high accuracy, while deep learning shows great tolerance to fluctuations and lack of data. In this paper, under the original data condition, a five-fold cross-validation comparison experiment was respectively performed based on KNN, decision trees, random forests and two-channel 1DCNN. Here 96 sampling points of each user were used as a sample input. The experimental results are shown in table 3.

Table 3. The comparison of the identification accuracy based on four models.

| Model              | Accuracy |
|--------------------|----------|
| KNN                | 57.97%   |
| Decision trees     | 93.40%   |
| Random forests     | 95.84%   |
| Two-channel 1DCNN  | 99.82%   |

From the experimental results, we can see that without data filling and data filtering, the accuracy of traditional algorithms has declined, and it is difficult to meet the requirements of practical applications. But the two-channel 1DCNN model can maintain high accuracy. It can be seen that deep learning has a strong tolerance for numerical fluctuations and missing samples, and is more suitable for data collection scenarios on the low-voltage side of the distribution network.

Figure 4 shows the loss value curve of training based on our model. It shows that after about 100 epochs’ training, the loss value of both train set and test set has stabilized, which means the model has converged and the training is complete.

5. Conclusions
The rapid development of the Ubiquitous Power Internet of Things has provided sufficient data and hardware equipment for various research on the smart grid, and also makes new requirements for the efficiency and accuracy of the identification of electrical topology. Based on the multi-source heterogeneous data of the distribution network, considering the data characteristics and collection conditions of power data, this paper proposes a distribution network electrical topology identification
algorithm based on deep learning. With two-channel 1DCNN model, we verify the accuracy of the model through a wealth of comparative experiments, and analyse the advantages of the model in the practical application of the distribution network.

Acknowledgments
This work was supported by State grid Sichuan Electric Power Company Science and Technique Research Program (52199719001M).

References
[1] Chen, J., Jiang, W., Jiao, H., Guo, Y., Nie, G. and Chen, W., (2018). Application Prospect of Edge Computing in Smart Distribution. In: 2018 China International Conference on Electricity Distribution (CICED), Tianjin, pp. 1370-1375.
[2] Xie, X., Zhou, J. and Zhang, Y. (2020). Application and challenge of deep learning in Ubiquitous Power Internet of Things. Electric Power Automation Equipment, v.40; No.312(04), 83-93.
[3] Zhang, X., Li, Y., Yang, C., Wang, S., Xie, W. and Ling, P., (2018). Topology Analysis of Distribution Network Based on \( \mu \)PMU and SCADA. In: 2018 International Conference on Power System Technology (POWERCON), Guangzhou, pp. 3427-3433.
[4] Chen, C. S., Ku, T. T., and Lin, C. H., (2011). Design of phase identification system to support three-phase loading balance of distribution feeders. In: Industrial and Commercial Power Systems Technical Conference (I&CP). Baltimore, USA. 1–8.
[5] Zhang, M., Luan, W., Guo, S., and Wang, P., (2018). Topology Identification Method of Distribution Network Based on Smart Meter Measurements. In: 2018 China International Conference on Electricity Distribution (CICED), Tianjin, pp. 372-376.
[6] Chen, Y., Chen, J., Jiao, H., Guo, Y., Jiang, W. and Tang, H., (2019). Two-Stage Topology Identification Method for Distribution Network Via Clustering Correction. In: 2019 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia), Chengdu, China, pp. 280-284.
[7] Yang, Z., Shen, Y., Yang, F., Lei, Y., Su, L. and Yan, F., (2020). Topology identification method of low voltage distribution network based on data association analysis. In: 2020 5th Asia Conference on Power and Electrical Engineering (ACPEE), Chengdu, China, pp. 2226-2230.
[8] Lu, C., Zhao, L. and Li, Y., (2020). Topology Checking Method for Low Voltage Distribution Network Based on Fuzzy C-Means Clustering Algorithm. In: 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, pp. 1077-1080.
[9] Bagheri, A., Bollen, M. H. J. and Gu, I. Y. H., (2017). Big data from smart grids. In: 2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), Torino, pp. 1-5.
[10] Chen, X., Tan, H., Yu, B., Li, C. and Chen, X., (2018). Multi-Objective Distribution Network Reconfiguration Based on Deep Learning Algorithm. In: 2018 IEEE International Conference on High Voltage Engineering and Application (ICHVE), ATHENS, Greece, pp. 1-4.
[11] Luo, G., Tan, Y., Yao, C., Liu, Y. and He, J. (2019). Deep learning-based fault location of DC distribution networks. The Journal of Engineering, vol. 2019, no. 16, pp. 3301-3305.
[12] Gu, H., Zhang, Y., Lu, X., Cui, Z. and Du, J. (2020). Research on Identification Method of Topological Relationship Based on Deep Learning. China CIO News, (3):150-151,153.
[13] Goodfellow, I., Bengio, Y., Courville, A. (2016) Deep learning. Cambridge: MIT press.
[14] Kiranyaz, S., Ince, T., Hamila, R. and Gabbouj, M., (2015). Convolutional Neural Networks for patient-specific ECG classification. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, pp. 2608-2611.
[15] Ioffe, S. and Szegedy, C., (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift. https://arxiv.org/abs/1502.03167.