

PAPER • OPEN ACCESS

## Topology Verification of Low Voltage Distribution Network Based on k-means Clustering Algorithm

To cite this article: Junyi Wang *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **569** 052096

View the [article online](#) for updates and enhancements.

# Topology Verification of Low Voltage Distribution Network Based on k-means Clustering Algorithm

Junyi Wang<sup>1</sup>, Xingquan Ji<sup>1\*</sup>, Kejun Li<sup>2</sup>, Qiaoyu Sun<sup>1</sup>

<sup>1</sup> School of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao, Shandong, 266590, China;

<sup>2</sup> School of Electrical Engineering, Shandong University, Jinan, Shandong, 250061, China

\*Corresponding author's e-mail: eesdust@qq.com

**Abstract.** Aiming at the problem of topology connection error existing in GIS system of low voltage distribution network, a topology verification algorithm using AMI voltage measurement data combined with k-means clustering algorithm is proposed. Firstly, the voltage data of the consumers in the low-voltage substation area is obtained by the AMI measurement system. Then k-means clustering algorithm is used by the similarity to cluster the voltage curves to identify and verify the users connected on the incorrectly transformer stations. An improved method of noise processing using data density set is proposed to solve the problem of initial cluster centre selection in k-means. For the problem of  $k$  value selection, an automatic correction of optimal  $k$  value is proposed. The feasibility and effectiveness of the improved k-means clustering algorithm in topology verification are verified by a practical example.

## 1. Introduction

Power companies usually use the Geographical Information System (GIS) to record the topology connection between electrical equipment and users in the low-voltage distribution network, including the connection between the distribution transformers and the users, or distribution transformers and feeder lines. However, the connection structure changed because of the rebuilding of distribution network, and there is a long term record of wrong topological structure caused by untimely updated information in GIS. The grid dispatcher may not obtain the true topology structure in time, which has a significant impact on the operation, control, protection and maintenance work of distribution. In order to solve the problems, it is essential to develop a reasonable topology verification method to quickly and accurately verify the topology recorded in the GIS and obtain the correct topology connection.

Efforts have been made to verify the topology structure. Many researchers have developed various algorithms to update the GIS record when the asset connection changed. The traditional topology verification method mainly based on transmission network's method such as transfer power method<sup>[1]</sup>, information map method, which is not available any more due to the lack of measurement data in the distribution network. The availability of Advanced Meter Infrastructure (AMI) customer data has recently provided user voltage, branch current data for an alternative approach to the GIS accuracy improvement efforts. According to [2], those customers in incorrectly GIS topology record were identified by calculating the correlation factors of user voltage provided by AMI for 24 hours under



the same transformer, and selected to perform similarity analysis with users of adjacent transformer, and determined the location which those actually belong to. The Local Outlier Factor (LOF) algorithm was used to identify the outliers which are in low density, as the user who were recorded in wrong GIS topology, after calculating the voltage's correlation factors between the customers to be verified and the customers in other transformer stations in [4]. And the clipping k nearest neighbour method is proposed to identify the users with bigger distance measured by Fréchet distance to represent the similarity of the costumers 'voltage curve, as the wrong connection users in GIS in [5]. According to above researches, most of existing topology verification method require manual operation and may not be reflected directly by verification result. A low-voltage distribution network topology verification algorithm based on improved k-means clustering algorithm is proposed in this paper. In this algorithm, a noise processing is applied on initial data set using data density, and an adaptive k value selection algorithm is proposed judging by correlation coefficient of data to solve the problem that k cannot be determined before clustering, existing in k-means clustering. The relation between users under the same line or transformer can be revealed directly by clustering result, which shows great superiority of this algorithm.

## 2. Principle of topology verification

The prominent error in GIS records is incorrect connectivity, which is exemplified by customers connected to wrong poles or wrong transformers, or transformers connected to wrong feeder phases [2]. The topology verification algorithms are mainly based on following three fundamental electrical characters. (1) The users' voltage profiles exist difference on different lines under the same transformer due to the different load usage. (2) The closer the user's electrical distance is, the higher the similarity of the node voltage profiles exist. Therefore, the costumers tend to higher similarity connected on the same transformers or feeders. (3) The voltage magnitude decreases along the line on the same transformer. So the upstream customers have higher voltage magnitude than downstream customers.

## 3. K-means clustering

K-means clustering is a popular method for clustering analysis in data mining, aiming to partition n observations into k clusters in which each observation belongs to their own clustering centre. Given a group of observation  $(X_1, X_2, X_3, \dots, X_n)$ , each of them in the group are m dimension vectors. To find k initial clustering centre  $\mu_k$ , and cluster the observation group into k classes  $C = (C_1, C_2, C_3, \dots, C_k)$  which the data in the same classes have the nearest distance to clustering centre. Then, the clustering centres are continuously updated and the observation is re-clustered by new clustering centres, so as to minimize the within-cluster sum of squares(WCSS), the function is as following:

$$E = \sum_{j=1}^k \sum_{x \in C_j} \|x - \mu_j\|_2^2 \quad (1)$$

Where x are initial clustering centers after clustered as well as  $\mu_j$  are new clustering centers after updated.

In general, traditional clustering algorithms are mainly applied to the clustering of data points rarely to cluster curves. Therefore, k-means clustering shows excellent advantages in topology verification as following: (1) It shows extremely high similarity in the same cluster, and distinct differences between the different clusters. (2) The curve of the topology connection error can be clearly displayed by the clustering result. (3) The algorithm is simple and easy to implement, and can be used to quickly cluster large-scale data curves.

Besides, it also appeared short for the topology verification. (1) The random selection of cluster centres may cause the clustering results to be difficult to meet the expectation, and is easily affected by the data noise. (2) Traditional clustering method based on distance cannot express the similarity between curves. (3) The clustering result of k-means clustering algorithm is greatly affected by k value, which cannot be determined before the topology verification.

#### 4. Topology verification process

In view of the shortcomings of k-means clustering algorithm, an improvement is proposed for initial data set processing, clustering measurement and  $k$  selection. The process of low-voltage distribution network topology verification is described.

##### 4.1 Processing initial data

Influenced by the function and performance of AMI, inevitably existed noise data in acquired voltage measurement data. The noise data can not only increase the storage space, but also affects the selection of the initial cluster centers, which resulting in clustering errors. So, it is necessary to process noise data before clustering. A noise processing method is proposed based on density of points according to [7].

Firstly, calculating the Euclidean distance between a certain object  $p$  selected in initial data and the other data. Then, the data object  $p$  is chosen as a center, with a radius of half the mean of all distances as a neighbour. Counting the number of data whose distance from the center is smaller than the radius of the neighbour. Regard those data whose counting numbers are larger than half of whole number of data as high density points, or as noise data.

$$d_{ij} = \left( \sum_{a=1}^m (x_{ia} - x_{ja})^2 \right)^{1/2} \quad (2)$$

Here,  $d_{ij}$  denotes the Euclidean Distance between two curves.  $m$  is the demission of vectors.  $x_{ia}$ ,  $x_{ja}$  are vector coordinates.

##### 4.2 Clustering metrics

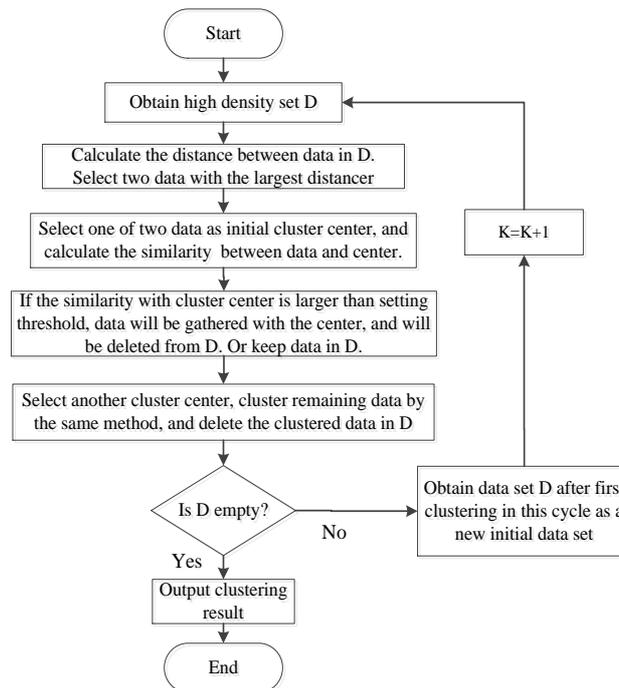
The similarity of curves can be better measured by Pearson Coefficient, compared to Euclidean Distance which is fit more for measuring two or three demission vectors. The two curves can be clustered together if their Pearson Coefficient are larger than a certain threshold.

$$r = \frac{\sum_{i=1}^m (x_i - \bar{X})(y_i - \bar{Y})}{\left( \sum_{i=1}^m (x_i - \bar{X})^2 \right)^{1/2} \left( \sum_{i=1}^m (y_i - \bar{Y})^2 \right)^{1/2}} \quad (3)$$

Here,  $x_i$ ,  $y_i$  are the sample voltage data of two users, and  $\bar{X}$ ,  $\bar{Y}$  are means of two users' voltage.  $r$  is Pearson Coefficient limited at  $[0,1]$ . The larger  $r$  means the higher similarity.

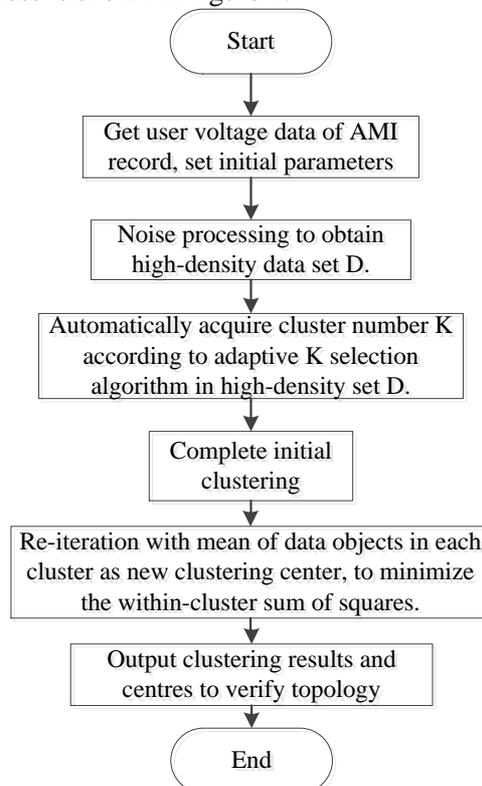
##### 4.3 Adaptive $k$ value selection

The clustering result of the k-means clustering algorithm is greatly affected by  $k$ . The value of  $k$  cannot be obtained before clustering as it is impossible to determine whether the data exists or how many types of users from other stations. An adaptive  $k$  value selection is proposed in this paper to dynamically obtain the cluster numbers. The process is shown in figure 1.

Figure 1. Adaptive  $k$  value selection process.

#### 4.4 Topology verification process

The improved  $k$ -means clustering algorithm is applied to the topology verification of low-voltage distribution network. The process is shown in figure 2.

Figure 2. Topology verification based on  $k$ -means clustering

**5. Test result**

A practical transformer substation A and its neighbour transformer stations are selected for topology verification. 38 customers are posted into right position in GIS after verification. The study is divided into two part to testify the algorithm’s availability. In the progress of testing, threshold of similarity is 0.9, and the initial value of  $k$  is 2.

*5.1 Verification of different lines’ users under same transformer*

Customers may be transferred between two lines under the same transformer station area, but the connections may not be updated timely in GIS when it occurred. For that topology changes, the algorithm proposed in this paper is used to verify and update the line of the user in the GIS system. The initial topology in GIS is shown in Figure 3. Customers on all lines are configured AMI smart meter M1-M16, and line 1 is chosen to be verified.

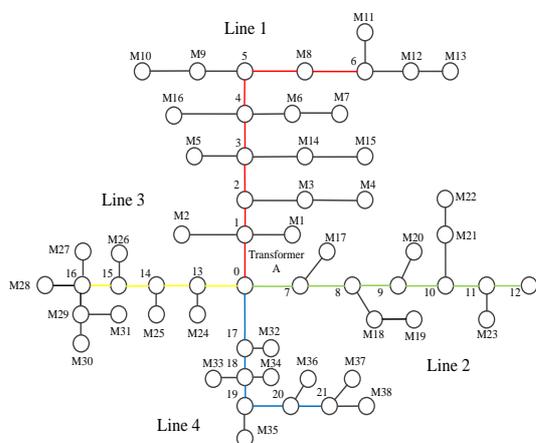


Figure 3. Initial topology of transformer A in GIS

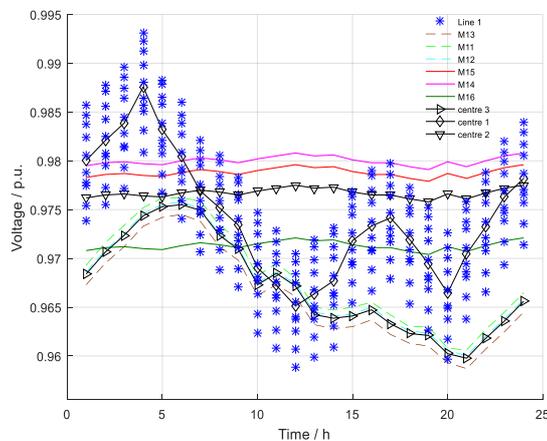


Figure 4. Clustering result of line 1

Firstly, the smart meter measurement data of users on line 1 is selected, and 16 users’ voltage data recorded on line 1 are accurately obtained after noise processing, which are actually high density data. The  $k$  value of clustering is 3 which is selected automatically by the adaptive  $k$ -value selection algorithm after  $k$ -means clustering. The value of  $k$  indicating that there are two types of users from other lines which have smaller counting, except the line 1 users. The result is shown in figure 4.

The figure 4 shows that M1-M10 are users of line 1 which have great similarity among the cluster, and M11-M13 and M14-M16 show distinct difference from line 1 users. After clustering, M1-M10 perform similarly with clustering centre 1, as M11-M13 and M14-M16 are similar to clustering centre 2 and centre 3, all of them are recorded in wrong line in GIS. The similarity analysis was performed between the two clusters of users and the users under Line 2 and Line 3. The results are shown in Table 1, Figure 5, and Figure 6.

Table 1. Similarity analysis

Meter ID	M17	M18	M19	M20	M24	M25	M26	M27
M11	0.9882	0.9845	0.9832	0.9843	0.7651	0.7542	0.7689	0.7643
M12	0.9983	0.9912	0.9923	0.9985	0.7517	0.7523	0.7656	0.7668
M13	0.9878	0.9876	0.9877	0.9872	0.7714	0.7775	0.7765	0.7761
M14	0.7531	0.7552	0.7589	0.7601	0.9772	0.9745	0.9743	0.9744
M15	0.7601	0.7612	0.7603	0.7615	0.97758	0.9736	0.9756	0.9734
M16	0.7589	0.7588	0.7593	0.7578	0.9845	0.9872	0.9816	0.9853

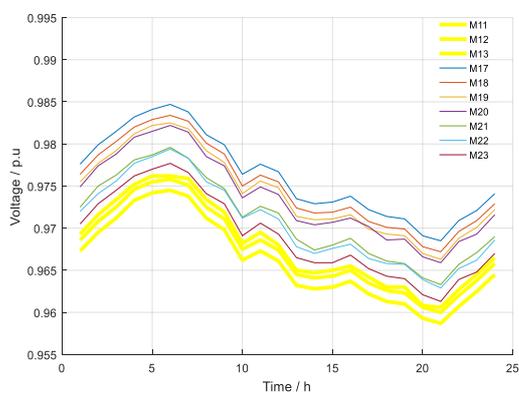


Figure 5. Line 2 users and M11-M13 voltage curves

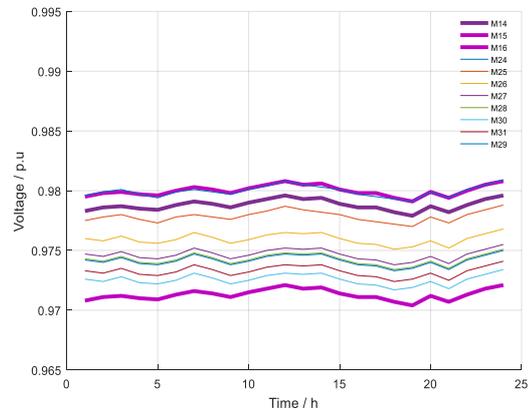


Figure 6. Line 3 users and M14-M15 voltage curves

Table 1 show that the users' voltage of M11-M13 has higher similarity with the users' voltage of M17-M20 all are greater than 0.9, in contrast with the users' voltage of M24-M27 are all about 0.75, which means they are probably from line 2. It has been ensured by figure 5 because they have the coincident voltage fluctuation of 24 hours. Eventually, the position can be judged from the magnitude of M11-M13, which are lowest, means that they belong to downstream end of line 2. In the same way shown in figure 6, the right position of M14-M16 can also be judged into line 3, because of the similarity and voltage fluctuation. Finally, topology structure in GIS is shown in figure 7 after verification.

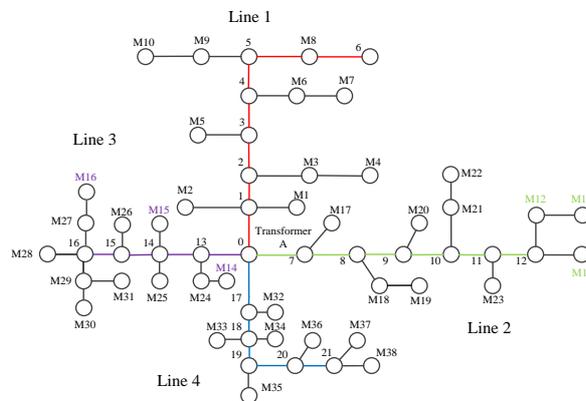


Figure 7. Topology structure of transformer A after verification

### 5.2 Users under different transformers

The low-voltage distribution network is usually a hand-in-hand network structure for multi-terminal power supply, which connect users and the transformer feeders by tie switches. Users are often connected to another transformer's power supply ranges by tie switches, due to the limitation of transformer capacity or users' line fault, in low-voltage distribution. However, these changes won't be updated by GIS timely. Aiming for this situation, all users' voltage data in transformer area A is used as the observation object, and the clustering is performed after the noise processing. Power supply ranges of all transformer is shown in figure 8.

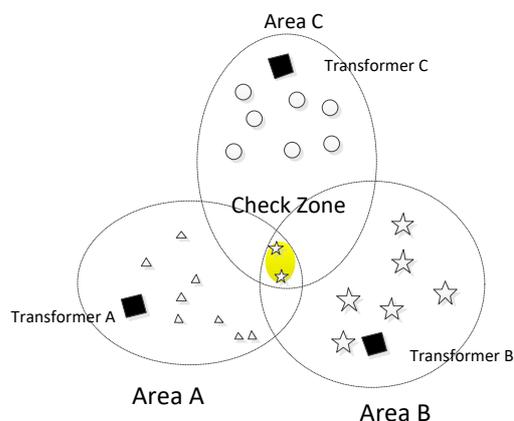


Figure 8. Power supply ranges

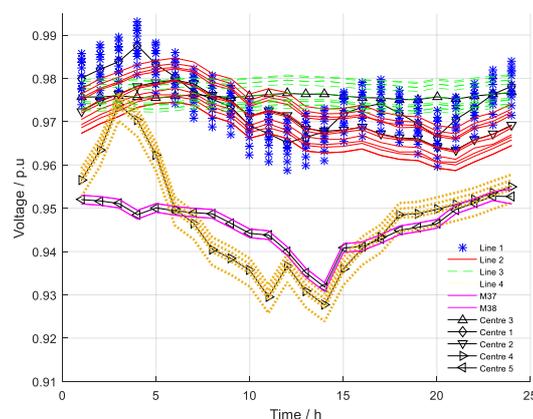


Figure 9. Clustering result

According to the clustering results in figure 9, the users' voltage curves of station area A are clustered into five groups automatically, excepting for users of four lines under the transformer, one type contains two users M37 and M38, which are showing obviously differences from the other four, presenting connected in another transformer. The similarity analysis is performed on the identified users M37 and M38 with the users under the transformer B and transformer C. It behaved high similarity with users under transformer B. After field visits, it can be determined that the users M37 and M38 are connected from the original station A to the station B due to the line fault.

## 6. Conclusion

A topology verification algorithm was proposed based on improved k-means clustering, addressing the problem of error connection exist in GIS of low-voltage distribution network. By performing noise processing on the initial data, a high-density set is obtained, and an adaptive k-value selection algorithm is used to obtain the optimal cluster number. Two kinds of topology error in GIS can be verified quickly and feasibly.

## References

- [1] Chen, Y.B, Zhou, J.Y, et. (2010) Transfer power flow approach to topology error identification. *Automation of Electric Power Systems*. 34:20-24.
- [2] Luan, W.P, Peng, J, Maras, M, et. (2015) Smart Meter Data Analytics for Distribution Network Connectivity Verification. *IEEE Transactions on Smart Grid*. 6:1964-1971.
- [3] Yan, W.G, Wang, G.M, et. (2018) A Method for LV Distribution Network Topology Verification based on AMI Metering Data. *Electric Power*. <http://kns.cnki.net/kcms/detail/11.3265.TM.20181206.1452.006.html>.
- [4] Geng, J.C, Wu, B, et.(2017) Low-Voltage Transformer Area Topology Verification Based on Outlier Detection. *Electric Power Information and Communication Technology*.15:61-65.
- [5] Geng, J.C, Zhang, X.F, et.(2017)Topology verification of low-voltage transformer areas based on discrete Frechet distance and editing nearest neighbors method. *Electrical Measurement & Instrumentation*. 54:50-55.
- [6] Mohammad, B, Marcelo, G.S, Michael, B.W, et. (2016) Smart Grid Topology Identification Using Sparse Recovery. *IEEE Transactions on Industry Applications*.52:4375-4384.
- [7] Ji, X.Q, Han, G.Z, et. (2016) Application of Improved K-means Clustering Algorithm Based on Density in Distribution Network Block Partitioning. *Journal of Shandong University (Engineering Science)*.46:41-46.

- [8] Sharon, Y, Motto, A.L, et. (2012) Topology Identification in Distribution Network with Limited Measurements. In: IEEE PES Innovative Smart Grid Technologies. Washington.
- [9] Mohammad, B, Marcelo, G.S, Michael, B.W, et. (2016) Compressive Sensing-Based Topology Identification for Smart Grids. IEEE Transactions on Industry Informatics.12:532-543.
- [10] Bolognani, B, Bof, N, Michelotti, D. (2013) Identification of Power Distribution Network Topology Via Voltage Correlation Analysis. In: 52nd IEEE Conference on Decision and Control. Florence, Italy.
- [11] Luan W, Peng J, Maras M, et al. (2013) Distribution network topology error correction using smart meter data analytics. In: Power and Energy Society General Meeting. Vancouver, BC, Canada
- [12] Kanungo, T, Mount, D.M, Netanyahu, N.S. (2002) An efficient k-means clustering algorithm: analysis and implementation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 24:881-892.
- [13] Zhou, S.B, Xu, Z.Y, Tang, X.Q. (2010) Method for determining optimal number of clusters in K-means clustering algorithm. Journal of Computer Applications. 30:1995-1998.
- [14] Zhang, K.S, He, Y.Q, Liu, F.C, et. (2014) Load Clustering Method and Its Application Based on the Similarity of Curves. Power System and Clean Energy. 30: 32-38.
- [15] Rasanen, T, Voukantsis, D. (2010) Data-based Method for Creating Electricity Use Load Profiles Using Large Amount of Customer –specific Hourly Measured Electricity Use Data. Applied Energy.87 :3538-3545.