Topology Identification of Low-voltage Transformer Area Based on Improved Particle Swarm Algorithm

Han Liu¹, Xian Huizhu¹, Duan Dapeng¹, Lv Fengming¹, Sun Jian¹, Mao Haofei², Wang Qiang²

¹Beijing Metrology Center, State Grid Beijing Electric Power Research Institute, Beijing, 100080, China
²Smart Industry Division, Beijing Soarrow Space Technology Co., LTD., Beijing 100076, China

*Corresponding author’s e-mail: ZHCYB@htmingdi.com

Abstracts: With the continuous changes in the user-side power environment, the low-voltage distribution network has become more and more complex, which brings great challenges to the line loss management and topology identification of the transformer area. To solve the shortcomings of traditional particle swarm optimization such as poor ergodicity of population initialization and easy to fall into premature convergence, this paper proposed a K-means clustering analysis algorithm based on GA-CPSO. First, on the basis of the traditional particle swarm algorithm, the chaotic shrinkage factor was introduced and the parameters were optimized. Secondly, the genetic algorithm was combined with the chaotic particle swarm optimization algorithm, the crossover and mutation operations of the genetic algorithm were used to establish an information exchange mechanism between particles, and it was combined with the K-means clustering method. The simulation results on the test benchmark function show that the improved particle swarm optimization algorithm in this paper has significantly improved the search speed and optimization accuracy. Finally, taking an actual transformer area as an example, the method was applied to the transformer area topology recognition analysis. Using the electrical parameters such as voltage, current, and active power obtained from the monitoring terminal as sample data, a simulation analysis was carried out to verify the effectiveness and feasibility of the algorithm. The performance parameters of different algorithms had been compared and analyzed through multiple experiments, and it was proved that this method can effectively improve the accuracy of platform topology recognition, and has strong practicability and generalization.

1. Introduction

With the rapid development of power Internet of Things in recent years, it is imperative to build a big data system for power grids. It is gradually becoming a research hotspot in how to use the acquired massive data information to promote power grid management and to improve the stable operation of the power grid in order to improve the ability to deal with problems [1-2]. The low-voltage distribution network field wiring is very complicated, with a large amount of data, and there are many changes in the relationship between households and changes. Correctly obtaining the topological relationship of the low-voltage station area can provide necessary basic support for real-time monitoring of the station state, accurate fault location, lean management of line loss, monitoring of power theft by users, fine management of the low-voltage transformer area and other functions [3-4].

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
At present, the topological relationship identification of low-voltage transformer area distribution network mainly adopts two methods of on-site manual detection and online intelligent monitoring. Manual detection is generally performed by the staff using the transformer area identification device for on-site measurement \([5-6]\). This method has low recognition accuracy, especially when the grid user load changes, it will be difficult to locate the fault because the system is not recognized in time, and it is likely to affect the normal power consumption of customers. Automatic topology recognition currently mostly uses methods such as carrier communication and pulse current based on monitoring terminals and hardware modules of the HPLC communication unit, so its recognition accuracy is relatively high, but the pulse current needs to be controlled within a certain amplitude range, and carrier communication has problems like "station crosstalk", which makes the accuracy to be easily affected, and there are certain difficulties in operation, and a better improvement plan has not yet been obtained. With the continuous expansion of the scale of smart meter users, power companies are more inclined to obtain the operation of household meters through the monitoring terminals of smart meters, laying a research foundation for the use of smart data analysis to identify the online topology of the transformer area.

Compared with the manual recognition method, the system's online intelligent topology recognition method is more efficient and accurate. The method of online recognition is generally based on the multi-information fusion analysis method \([7]\), which has excellent real-time performance and is the mainstream development trend of the current distribution network topology identification and verification. Previous research \([8]\) adopted the technology based on the departure point detection and used the information data of the voltage of the smart meter to change over time to realize the identification and verification of the low-voltage distribution network topology, but the low voltage distribution network topology in a certain area cannot be verified quickly and in large quantities. Literature \([9]\) proposed a smart transformer area user information identification system based on BP neural network and interaction of electricity information with concentrator based on power line carrier communication technology, and comprehensively used the transformer area identifier and handheld devices. However, in the peak period of centralized power consumption and when the grid user load has a short time of sudden change, the identification accuracy will be greatly reduced. Literature started from the relationship between the power loss caused by the line loss and the user's own power monitoring data, and introduced the method of using the Pearson correlation coefficient algorithm to find abnormal electric energy meters, but this method is based on the premise that the line loss factor or the user's power consumption error is far greater than the power loss caused by other factors. Previous research was based on the variance model of the branch voltage deviation measured by the node injected power measurement, which can effectively identify the topological operation structure of the distribution network, but this method required that the node injected power of the distribution network is fully observable.

To improve the accuracy of topology identification, this paper proposed a low-voltage station topology identification method based on cluster intelligent classification algorithm, which is mainly used in the topological identification of the meter box of the transformer area-household meter. The cluster intelligent classification algorithm is composed of two algorithms: improved particle swarm algorithm and K-means clustering analysis. Traditional particle swarm optimization algorithms have been widely used in high-dimensional data optimization due to the advantages of fewer variables, accurate particle position and velocity update, and fast convergence. However, conventional particle swarm optimization algorithms are easy to fall into local optimality and global optimal solution cannot be obtained. This paper proposed an improved chaotic particle swarm algorithm, which introduces a shrinkage factor when the particle speed is updated, and adds a chaotic optimization mechanism during the algorithm operation to avoid premature convergence of the algorithm. On this basis, genetic algorithm was introduced, which uses classic crossover and mutation operations to establish an information exchange mechanism between particles, thereby enhancing the global convergence of the algorithm. In the clustering analysis, this paper designed the improved algorithm aiming at the problems such as the sensitivity of traditional K-means clustering algorithm to the value of the initial clustering center, poor global search ability and low clustering accuracy in the process of classification solution, and the clustering center position in the K-means algorithm was optimized by the improved genetic
chaos particle swarm optimization algorithm. The selection of clustering number was evaluated and verified by the (CH)(Calinski-Haeabasz) index, and the selection of clustering number was optimized.

2. Topology identification of low-voltage transformer area

2.1. Basic structure of low-pressure transformer area
This paper combined the existing meter file information and the actual measurement data of the smart meter to construct a three-layer tree structure of "station transformer-virtual user-user". The "virtual user" was composed of a number of user nodes that are close in electrical distance. The voltage was the center point voltage of the users in the cluster, and its active power was the sum of the power of all users in the cluster. In the topological network, a large number of ordinary user nodes were replaced, and the wiring structure of "transformer - line branch point - table box - user" was fit. On the one hand, it can help to improve the topology identification of the station, and on the other hand, it can relieve the computing load caused by the increase of nodes in the station. The basic composition of the low-pressure transformer area is shown in Figure 1.

![Figure 1. The structural diagram of low-voltage station topology](image)

2.2. Topology identification method of transformer area
Low-voltage transformer area topology recognition generally includes three parts: transformer area recognition, branch recognition, and meter box recognition. This article used a combination of power frequency distortion, pulse current, and GA-CPSO's K-means clustering analysis method to achieve high efficiency, and high-precision topology identification.

(1) Recognition of household change relationship. This paper used the transformer area recognition function based on the broadband power line carrier (HPLC) communication unit. The zero-crossing detection circuit of the HPLC communication unit supports the reception of power frequency distortion signals. The concentrator generates power frequency distortion signals on the side of the transformer area (modulating the main node address + phase of the current transformer area CCO), and after the HPLC communication unit receiving the power frequency distortion signals, based on the received master node address, the network can be selectively accessed the network, thereby quickly solving the
problem of identifying the relationship between households and changes.

(2) Branch recognition. The characteristic current method based on resistance switching is adopted. All terminals in the low-voltage transformer area sequentially generate characteristic current signals according to the commands of the master station and the concentrator, and send them to the low-voltage transformer area power lines, and modulate the monitoring terminal or meter address into the characteristic current signal. Both the upper-level monitoring terminal and the concentrator in the same loop of the low-voltage transformer area can receive the characteristic current signal, and use the method based on time-frequency conversion to extract the characteristics of the current signal, and demodulate and record the monitoring terminal address in the characteristic current signal. After the concentrator reads it, draw the branch topology of the low-voltage transformer area. The maximum emission current of the characteristic current signal is only 420mA, the amplitude is small, and it is not a distorted current and will not affect the power quality.

(3) Meter box identification. The meter box relationship identification usually adopts methods such as carrier communication and pulse current based on the hardware module of the HPLC communication unit. The current technology maturity is not high, and it has not been widely promoted and applied. In this paper, GA-CPSO's K-means cluster intelligent classification algorithm is used, which combines genetic chaotic particle swarm algorithm and K-means clustering algorithm to establish an intelligent classification mathematical model. A variety of parameter data of the user's electricity meter collected by the meter box monitoring terminal is substituted into the model for calculation and solution to realize topological relation identification. After the concentrator completes the identification of the household change relationship, it obtains the current station file, and issues the box identification command in the order of registration through the HPLC channel. After the meter box monitoring terminal receiving the instruction, it performs the multi-parameter data collection work of the household meter and sends the data to the data processing module of the concentrator for analysis and processing. The concentrator will match the membership of the output meter box and household meter with the address information in the transformer area identification, thereby completing the topology identification.

The station topology recognition process is shown in Figure 2.
3. Algorithm research

3.1. Traditional particle population algorithm

Particle Swarm Optimization (PSO) is a swarm search algorithm that simulates the social behavior of a flock of birds. The purpose is to find a pattern that governs the simultaneous flight of a flock of birds and a pattern that suddenly changes direction when the optimal form is reorganized. The expressions of particle velocity and position in the PSO algorithm are as follows.

\[ v_{i}^{k+1} = v_{i}^{k} + c_{1}\cdot r_{1}\left(p_{i}^{k} - x_{i}^{k}\right) + c_{2}\cdot r_{2}\left(p_{g}^{k} - x_{i}^{k}\right) \]  \hspace{1cm} (1)

\[ x_{i}^{k+1} = x_{i}^{k} + v_{i}^{k+1} \]  \hspace{1cm} (2)

Among them, \( v_{i}^{k} \) is the velocity of particle \( i \) in the \( k \)-th iteration; \( x_{i}^{k} \) is the position of particle \( i \) in
the k-th iteration; \( \omega \) is the inertia weight coefficient; c1 and c2 are learning factors, r1 and r2 are random numbers in \([0,1]\); \( p_i^k \) is the individual optimal values found by the i-th particle in k iterations, and \( p_a^k \) represents the historical optimal value of the population after k iterations of optimization.

The basic flow of the algorithm is as follows.

1) Parameters initialization: The initial position and initial velocity of the particles are randomly generated.

2) Evaluate particles: According to the objective function, a suitable fitness function is defined, and then the fitness value of each particle is calculated.

3) Update the optimal position: If it is assumed that the larger the fitness value of a particle, the better the corresponding particle, the optimal position update formula of particle i is.

\[
p_i(t) = \begin{cases} x_i(t), & f_i(t) \geq f_i(t-1) \\ p_i(t-1), & f_i(t) < f_i(t-1) \end{cases}
\]

Where \( x_i(t) \) represents the position of the i-th particle during the t-th iteration, \( p_i(t) \) represents the individual optimal position of the i-th particle during the t-th iteration, \( f_i(t), f_i(t-1) \) respectively represent the fitness value of the i-th particle in the t-th and (t-1)-th iterations.

4) Update the historical optimal position: The formula is as follows.

\[
g(t) = \begin{cases} x_i(t), & f_i(t) \geq f_i(t-1) \\ g_i(t-1), & f_i(t) < f_i(t-1) \end{cases}
\]

(4)

Update the population. The particle velocity is updated according to the following formula.

\[
v_{i+1}^k(t+1) = \omega v_i^k(t) + [c_1 r_1 (p_i^k(t) - x_i^k(t)) + c_2 r_2 (g_i(t) - x_i(t))]
\]

(5) Let \( \phi = c_1+c_2 \), \( c_1 \) and \( c_2 \) are learning factors, and their values are \( c_1=c_2=2.05 \); \( r_1 \) and \( r_2 \) are random numbers that vary within the range of \([-1,1]\), the initial value of \( \phi \) can be taken as 1, and the shrinkage factor can be expressed as (6).

\[
r = \frac{2\kappa}{2-\phi-\sqrt{\phi^2-4\phi}}
\]

(6)

(5) Determine whether the algorithm is terminated: If the fitness value no longer changes or reaches the maximum number of iterations, stop the iteration, otherwise, go to step (2).

3.2. K-means clustering algorithm

The K-means clustering algorithm takes the minimization of errors as the benchmark criterion, uses distance as the similarity evaluation parameter, and initially sets the number of clusters \( k \), adopts distance as the similarity evaluation, and uses the centroid of cluster \( E_i \) \( (i=1,2, \ldots k) \) \( e_i \) represents the cluster. Using \( d(x,e_i) \) to represent the difference between \( x \in E_i \) and the cluster center \( e_i \), the calculation formula is as follows.

\[
d(x_i,x_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{ip} - x_{jp})^2}
\]

(7)

Taking the sum of squared errors as a criterion function, it is usually defined as an objective function to measure the level of cluster classification, which indicates how closely the samples in the cluster surround the cluster center. The smaller the value, the higher the similarity of the sample data in the same class. The expression of the error sum of squares is as follows.

\[
SSE = \sum_{i=1}^{k} \sum_{x \in E_i} dist(x, e_i)
\]

(8)
\[ e_i = \frac{1}{n_i} \sum_{x \in E_i} x \]  \quad (9)

In the above formula, \( x \) represents the sample data object; \( k \) is the number of clusters; \( e_i \) is the cluster center of the cluster \( E_i \); \( n_i \) is the number of sample data in the \( i \)-th cluster.

4. Improvement Design of Particle Swarm Optimization (PSO) Algorithm

4.1. Improvement design of PSO based on chaotic shrinkage factor

The particle swarm optimization (PSO) algorithm can easily fall into the local optimum, though integrating the merits of high update rate and fast convergence in solving nonlinear problems. It has been continuously improved at both home (China) and abroad and widely applied to various fields \[15-16\]. The present improvement research on the PSO algorithm mainly aims at its convergence performance, easy fall into the local optimum and other defects. The improvement method mainly includes the parameter optimization in speed updating formula, enhancement of population diversity, combination with other optimization algorithms, etc.

It should be pointed out that in the solving process using the above mentioned improvement methods, once a particle finds the optimal solution, the other particles will fly towards the optimal solution at full speed, thus resulting in the particle stagnation and premature convergence, which are triggered by the limitations of the PSO algorithm itself. Given this, a chaotic shrinkage factor was proposed in this study to update the particle velocity and further enhance the algorithm convergence. Moreover, the chaotic optimization mechanism was added into the algorithm operation process. The chaos optimization, which was of universality and ergodicity within a certain range, could effectively strengthen the ability of PSO algorithm to seek for the global optimum. When the particles fell into the premature convergence, they could jump out of local optimum by using chaos search, trying to avoid the premature convergence of the algorithm as much as possible.

The chaotic sequence can be generated by many equations, among which, however, logistic mapping is commonly used, with the following iterative formula.

\[ y_{i+1} = R y_i (1 - y_i), \quad i = 1, 2, 3... \]  \quad (10)

Where \( R \) is the control variable, with a value range of \([0,4]\). The value range of \( y_i \) is \([0,1]\). For the position \( x_i \) of the \( i \)-th particle in the particle swarm, the specific chaos optimization steps are as follows.

a) The particle position \( x_i \) was mapped into the interval \([0,1]\), and the following was taken.

\[ y_i = \frac{x_i - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \]  \quad (11)

Where \( x_{\text{min}} \) and \( x_{\text{max}} \) are the lower bound and upper bound of problem space, respectively.

b) \( y_i \) was iterated for \( M \) times according to formula (9), and then a chaotic sequence \((y_{i0}, y_{i1}, y_{i2}, \ldots, y_{iM})\) was generated and mapped into the space of original solution:

\[ g_i = x_{\text{min}} + \left( x_{\text{max}} - x_{\text{min}} \right) y_i \]  \quad (12)

The sequence \((g_{i0}, g_{i1}, g_{i2}, \ldots, g_{iM})\) of feasible solutions can be acquired.

c) The fitness value of \( g_i \) was calculated, and the feasible solution corresponding to the optimal fitness value was used to replace the position \( x_i \) of the \( i \)-th particle in the particle swarm.

4.1.1. Determination of inertia weight value

In the PSO algorithm, the value of inertia weight \( \omega \) has a great bearing on the algorithm convergence rate and algorithm result. It presented a linear reduction with the number of iterations. A large inertia weight was used in the initial iteration phase, in order to accelerate the algorithm convergence. A small inertia weight was selected in the later iteration phase so that the particles could realize the local refined search and further improve the search accuracy.

The updating method of velocity vector for each particle in the improved PSO algorithm is shown in
the following formula.

$$\omega = (\omega_{\text{max}} - \omega_{\text{min}}) \frac{K_{\text{max}} - K}{K} + \omega_{\text{min}}$$

(13)

Where $\omega$ is the value of inertia weight; $K$ is the current number of iterations; the subscript $\text{max}$ is the maximum value and $\text{min}$ is the minimum value.

4.1.2. Selection of control parameters

The parameter $\kappa$ decides the population development ability. When its value is close to 1, the search scope will be enlarged, thus reducing the algorithm convergence rate. When it is approximate to 0, fast convergence can be realized. In general, $\kappa$ is a constant, but in order to facilitate the fast convergence in the later iteration phase, here $\kappa$ was made to gradually decline with the number of iterations, namely.

$$\kappa = \exp \left( \frac{t}{N} \right)$$

(14)

4.2. Introduction of genetic algorithm (GA) strategy

The PSO algorithm with chaos optimization is capable of effectively strengthening the global optimal search ability of particle populations and shows a certain universality and ergodicity within a certain range, but it solves the optimal solution by relying upon the continuous update rate of particles, lacking an information exchange mechanism between particles in the whole population. In order to enhance the algorithm search ability and global convergence, the genetic algorithm (GA) was introduced to realize the exchange of population information by operations such as selection, crossover and mutation of particles in the population, and finally the parameters conforming to the optimization object were generated. Next, the operators in GA were introduced to facilitate the better exchange between populations, so as to effectively accelerate the optimization rate and optimize the search result.

The crossover and mutation operators usually have fixed values in the traditional GA. It has been found by many domestic (Chinese) scholars that the optimization ability of this algorithm can be effectively improved by using adaptive probability value. In this study, the adaptive crossover and mutation probability values were used, as follows.

$$P_c = \begin{cases} P_{c1} \frac{\gamma(f' - f_{\text{avg}})}{f_{\text{max}} - f_{\text{avg}}}, & f' \geq f_{\text{avg}} \\ P_{c1}(1 + k), & f' \leq f_{\text{avg}} \end{cases}$$

$$P_m = \begin{cases} P_{m1} \frac{\gamma(f_{\text{max}} - f)}{f_{\text{max}} - f_{\text{avg}}}, & f' \geq f_{\text{avg}} \\ P_{m1}(1 + k), & f \leq f_{\text{avg}} \end{cases}$$

(15)

(16)

Where $\gamma$ and $k$ are constants with the interval of [0,1]; $f_{\text{max}}$ and $f_{\text{avg}}$ are the maximum value and average value of fitness, respectively; $f'$ is the optimal fitness value among the individuals under the crossover operation; $f$ is the fitness value of individuals requiring the mutation operation.

4.2.1. Crossover operation

The individual crossover operation was introduced into the algorithm iteration and implemented according to the set crossover probability $P_c$. In this process, a certain number of particles were designated and placed into the hybridization pool, in which every two particles were randomly paired and crossed over to generate the new generation of offspring particles with the same number and replace the parent particles. This operation could keep the total number of particles unchanged in the particle swarm, increased the information exchange and sharing between particles, and effectively improved the algorithm convergence rate. The position and velocity of new individuals were updated through formulas (17) - (18).

$$x_{k+1} = p_c \cdot x_i + (1 - p_c) \cdot x_j$$

(17)
\[ v_{i}^{t+1} = \frac{v_{i}^{t} + v_{j}^{t}}{|v_{i}^{t}| + |v_{j}^{t}|} \]

\( x_{i}^{t} \) and \( v_{i}^{t} \) denote the position and velocity of the \( i \) (th) particle at the \( t \) (th) iteration, respectively; \( P_c \) is the crossover probability.

### 4.2.2. Mutation operation

In order to enhance the population diversity and prevent the particle swarm from falling into local optimum in the later search phase, the mutation operation in GA was introduced into the PSO algorithm during its iteration process, namely, the particles were mutated at a certain mutation probability of \( P_m \), the particle velocity was updated as seen in formula (18), and the position updating was described as follows.

\[ x_{k}^{t+1} = x_{k}^{t} + x_{k}^{t} \cdot P_m \]  

(19)

\( x_{i}^{t} \) is the position of the \( i \) (th) particle at the \( t \) (th) iteration. In consideration of dynamically constrained optimization region, the particles were initialized in the dynamically constrained region, namely, the individual particles needing the mutation operation were mutated to randomly generate new individual particles.

### 4.3. Comparison of algorithm performance

In order to verify the superiority of genetic algorithm based chaotic particle swarm optimization (GACPSO) algorithm, it was comparatively simulated with PSO and CPSO using the testing benchmark function, as shown in Figure 3.

![Simulation Curves of Different Optimization Algorithms](image)

Figure 3: Simulation Curves of Different Optimization Algorithms

By analyzing Figure 2, it could be known that relative to the CPSO algorithm and PSO algorithm, the GACPSO algorithm integrated the merits of high search rate and high convergence precision, along with more ideal optimization and convergence effects.

### 4.4. Improvement of K-means clustering algorithm

The optimal cluster number should be determined before the K-means algorithm was used. This parameter was denoted by \( k_0 \), which was selected in the following way.

1) The search scope of cluster number was selected as \( k_{\text{min}} \leq k_0 \leq k_{\text{max}} \). In general, \( k_{\text{min}} = 2 \) and \( k_{\text{max}} = \text{int} \sqrt{n} \), where \( n \) is the number of samples.

2) The cluster number was evaluated by selecting the Calinski-Haeabasz (CH) index. When it comes to the physical significance of this index, it expresses the joint degree between data by the dispersion
matrix of different classes, namely, the separation degree between data is expressed by the dispersion matrix of classes. The higher the CH value, the higher the data difference between classes, and the better the clustering effect. The CH will reach the maximum value under an optimal cluster number, and it is defined as below:

\[ CH(k) = \frac{trB(k)}{trW(k)} \cdot \frac{(k-1)}{(n-k)} \]

Where \( k \) is the present cluster number; \( trB(k) \) is the trace of dispersion matrix between classes; \( trW(k) \) is the trace of dispersion matrix between classes; \( n \) is the number of samples.

5. Case Analysis of Clustering Identification Method Based on the Improved PSO Algorithm

5.1. Calculation of optimal cluster number

Based on the data acquired by electric meters in several test transformer areas on one day in October, 2020, the voltage, current and power data were used as the samples. The clustering analysis scheme 1 was designed by taking single voltage data as the sample, the scheme 2 was set by taking voltage and current parameters as the sample data, and the scheme 3 was formed by taking voltage, current and power parameters as the sample data. The related program was compiled via MATLAB, and the curve of CH index under different cluster numbers was drawn.

![Figure 4. Evaluation Results of CH Index under Different Sample Data](image)

The evaluation results of CH index under different schemes are presented in Figure 4. It could be known that the optimal cluster number under the three different schemes was 6, 4 and 3, respectively. The detailed classification accuracy under different data combinations is listed in Table 1.

| Scheme  | Sample data                     | Accuracy (%) | Maximum value of CH | Cluster number |
|---------|---------------------------------|--------------|---------------------|----------------|
| Scheme 1| Voltage                         | 75           | 560                 | 6              |
| Scheme 2| Voltage and current             | 87.5         | 458                 | 4              |
| Scheme 3| Voltage, current and active power | 100          | 621                 | 3              |

As seen in Table 1, the clustering analysis was done by the scheme 3 based on three electric parameters: voltage, current and power. When the CH index reached the maximum value, the corresponding classification accuracy was the highest, thus providing a basis for selecting the data samples in the topology identification of transformer areas.
5.2. Case analysis of topology identification in transformer areas

In this study, the voltage, current and active power data (at the same time point) of users within a transformer area in a school in recent 12h were selected to form a sample sequence. Each intelligent electric meter acquired the parameter data once every other 10 min, and the number of data points acquired within 24h was 144. The clustering analysis was performed using the 12h data points in the algorithm test, namely, the sample size was N=72, specifically as follows.

1) The population and algorithm parameters were initialized. Here, the measured U values of different parameters were taken as the sample data, and the sample size was N=72. The objective function and fitness function were defined. An initial chaotic variable was randomly generated. N chaotic variables were obtained according to formula (10) and mapped into the original problem space, thus obtaining N feasible solutions. The N feasible solutions served as the initial population \( U_i = (u_1, u_2, \ldots, u_i) \) (i=1, 2, ..., N).

2) The initial cluster number \( k \) was given. The number of meter boxes subordinate the selected branch was unknown. According to the agreement reached by the CH index on the selection of cluster number, \( 2 \leq k_0 \leq 6 \), and the number of meter boxes subordinate to household meter was assumed as 3, namely, \( k=3 \).

3) Each particle was regarded as clustering center, the initial particle position was defined as the initial optimal position of individual particle, and the initial velocity was randomly generated and denoted as \( V_i= (v_1, v_2, \ldots, v_i) \).

4) The fitness value of objective function for the initial population was calculated, the maximum value and average value were output, and the individual optimal value and global optimal value of the population were recorded. The individual optimal position and global optimal position were updated according to formulas (3) and (4), respectively, and the local optimal solution was output.

5) Crossover operation. The crossover probability \( P_c \) was calculated through the formula (15). Two particles were randomly selected and hybridized. The position and velocity were updated using formulas (17) and (18).

6) Mutation operation. The mutation probability \( P_m \) was calculated according to the formula (16). Two particles were randomly selected and hybridized, and the velocity and position were updated using the formulas (18) and (19). The particle position was searched for M2 times, then M2 feasible solutions were obtained, the distance from each feasible solution to all samples was calculated, the M2 distances were compared, and the feasible solution with the minimum distance was chosen for the sake of replacement of particle position.

7) The iteration was stopped when the maximum number of iterations was reached or the clustering center was no longer changed, or otherwise, the operation should be done from the Step 4). If the algorithm iteration was stopped under the present cluster number, the effective objective functional value \( G_k \) under this cluster number was calculated.

8) The cluster number was updated, and \( k=k+1 \) was set. The effective functional values \( G_k \) under all cluster numbers were compared. The cluster number corresponding to the minimum effective functional value was the optimal cluster number, and the identification result of the corresponding transformer area was acquired.

The voltage, current and power data acquired in a transformer area within 12h were taken as the sample data, the clustering analysis was done through different clustering methods, and the simulation results are as shown in Figure 5.
The clustering analysis results of K-means analysis and CPSO-KM algorithm are displayed in Figure 6 (a) and (b), respectively. In the node simulation diagram (Figure 4 (a)), different classes had a fuzzy boundary and were even interlaced, so the identification could hardly be realized by this clustering method. As shown in Figure 4 (b), individual nodes of cable box 1 and cable box 3 were crosslinked, with a poor identification effect. Aggregated plaques were manifested in the simulation result, but no too large cluster appeared. Therefore, it could be preliminarily judged that the acquired household meters were mostly the users of one or multiple meter boxes subordinate to the same branch box.

The identification results realized by the GACPSO-KM algorithm are as shown in Figure 4 (c). It could be clearly seen that the boundary between different classes was clear, which was consistent with the membership relationship between the meter box actually labelled by the power supply bureau-household meter.

The classification result of meter box could be obtained as 3 through the algorithm proposed in this study. By analyzing the voltage curve chart of user meters, meter 1 and meter 2 were subordinate to the meter box 1, meter 4 and meter 5 to the meter box 2, and meter 3, meter 6 and meter 7 to the meter box 3. The simulation results were completely consistent with the actual topological relation. The meter box-household meter membership relationship output by the simulation calculation was matched with the address of meter box saved in the identification of transformer area, so as to complete the meter box-
household meter topology identification. The topological relation in the whole transformer area could be acquired by combining the household meter-transformer relation identification and branch identification result.

6. Comparative Analysis of Algorithm Performance

In order to analyze the classification performance of the clustering algorithm—GAPSO-KM, the several commonly used classification methods were used to perform 30 independent tests, and the simulation calculation results were output. The GAPSO-KM was compared with the traditional K-means algorithm, K-medoids algorithm, CPSO-KM and Gaussian mixture model (GMM). Through the simulation calculation, the accuracy, precision, silhouette coefficient, fitness and calculation time of these algorithms were obtained, as seen in Table 2.

| Method     | Accuracy | Precision (%) | SC  | Fitness | Calculation time (s) |
|------------|----------|---------------|-----|---------|----------------------|
| K-means    | 87.23    | 83.14         | 0.67| 77.89   | 27.4                 |
| K-medoids  | 90.34    | 88.09         | 0.73| 65.76   | 34.8                 |
| PSO-KM     | 93.56    | 86.46         | 0.78| 75.48   | 17.1                 |
| GACPSO-KM  | 98.45    | 91.27         | 0.89| 78.89   | 25.2                 |
| CPSO-KM    | 95.79    | 86.76         | 0.87| 77.76   | 23.5                 |
| GMM        | 90.78    | 84.43         | 0.75| 56.67   | 22.7                 |

As seen in Table 2, the K-means algorithm had a low clustering accuracy, which was mainly ascribed to its strong reliance on the initial clustering center. The accuracy of K-medoids algorithm was improved somehow in comparison with K-means algorithm, but the program operation time was long due to the large calculated quantity. Through the analysis using the PSO-KM algorithm, the calculated result of each index was obviously improved compared with the traditional K-means algorithm, proving that the clustering effect was improved to a certain extent by using the intelligent swarm optimization. The CPSO-KM algorithm deeply integrated the intelligent ant colony algorithm with the clustering analysis, gave full play to the global search ability of CPSO and GACPSO, set the global optimal position searched as the clustering center, and enhanced the randomness of the next-generation population. In the meantime, it took full advantages of strong global search ability of K-means and conducted another more accurate local search nearby the particles, so it could hardly fall into local extremum. Comparatively speaking, the GACPSO-KM algorithm had a higher accuracy than the CPSO-KM algorithm with a better fitness value, indirectly verifying that the optimization effect of particles was more ideal.

7. Conclusion

A K-means clustering algorithm based on genetic particle swarm optimization (GAPSO-KM) was put forward in this study, in an effort to improve the accuracy and reliability of topology identification in the low-voltage transformer area. The traditional PSO algorithm was improved. The GA, PSO and clustering analysis were combined. The fitness functional values were compared through the data simulation of optimization process so as to comparatively analyze the algorithm performance. The optimal value of CH index was simulated using test data, and the optimal combination of sample data was determined. In the end, the algorithm verification was performed using the electric parameters acquired in the actual transformer area, which was then compared with the actual physical topology of meter box in the transformer area. The results showed that the accuracy of clustering identification result obtained by the GAPSO-KM was high, verifying the effectiveness of the proposed method. The GAPSO-KM algorithm was only applied to identify the meter box-household meter membership relationship in this study, so it may be extended to transformer area identification and other intelligent classification system models in the follow-up study.
Acknowledgments
This paper is supported by the science and technology project (52022319004a) of State Grid Corporation of China.

References
[1] Guo B, Shao J. (2019) Construction of monitoring data model for transmission and transformation equipment based on big data technology[J]. Computer Applications and Software, 36(2): 135-138.
[2] Xiong X, Zhou S, and Song K. (2019) Key technology of user-side power big data based on Ubiquitous power IoTs[J], Information Technology, 36(2): 135-138.
[3] Tang D, Liu Y, Xiong Z, et al. (2020) Early warning method of electricity anti-theft in distribution station area based on spatiotemporal correlation matrix[J], Automation of Electric Power Systems, 44(19): 168-181.
[4] Liu J. (2019) Research on anti-Stealing electricity considering the identification of the relationships between users and transformers[D]. Kunming: Kunming University of Science and Technology.
[5] Li J, Zhao H. (2014) The Development and Manufacture of a Multi-Function Equipment for Low Voltage Area Identified[J], Electrical Measurement & Instrumentation, 51(13): 107-111.
[6] Xu Z. (2009) Application Analysis of Area Users Identify Apparatus[J], Metrology & Measurement Technique, 36(11): 26-28.
[7] Lin J, Cao C, Zhang B. (2002) Topology identification of distribution network based on hierarchical model[J]. Journal of Chongqing University (Natural Science), 25(2): 71-74.
[8] Geng J, Wu B, Wan D, et al. (2017) Low-Voltage Transformer Area Topology Verification Based on Outlier Detection[J]. Electric Power Information and Communication Technology, 15(5) : 61-65.
[9] Li Y. (2017) Research on the intelligent transformer area recognition method based on BP neural network[J]. Electrical Measurement &Instrumentation, 54(3): 25-30.