Fault location of distribution network with distributed generation based on Karrenbauer transform and support vector machine regression

SIMING WANG, KAIKAI ZHAO✉

School of Automation and Electrical Engineering, Lanzhou Jiaotong University
China
e-mail: ✉*1104556123@qq.com*

Abstract: As the capacity and scale of distribution networks continue to expand, and distributed generation technology is increasingly mature, the traditional fault location is no longer applicable to an active distribution network and "two-way" power flow structure. In this paper, a fault location method based on Karrenbauer transform and support vector machine regression (SVR) is proposed. Firstly, according to the influence of Karrenbauer transformation on phase angle difference before and after section fault in a low-voltage active distribution network, the fault regions and types are inferred preliminarily. Then, in the feature extraction stage, combined with the characteristics of distribution network fault mechanism, the fault feature sample set is established by using the phase angle difference of the Karrenbauer current. Finally, the fault category prediction model based on SVR was established to solve the problem of a single-phase mode transformation modulus and the indistinct identification of two-phase short circuits, then more accurate fault segments and categories were obtained. The proposed fault location method is simulated and verified by building a distribution network system model. The results show that compared with other methods in the field of fault detection, the fault location accuracy of the proposed method can reach 98.56%, which can enhance the robustness of rapid fault location.

Key words: distributed generation, distribution network fault location, fault type, Karrenbauer transform, SVR agent prediction model

1. Introduction

With the increasingly tense energy situation in the world, distributed generation technology has developed rapidly in recent years. It has the advantages of cleanness, high efficiency, environmental friendliness, energy diversification, etc., and conforms to the national sustainable

www.czasopisma.pan.pl   PAN   www.journals.pan.pl
POLSKA AKADEMIA NAUK

462   S. Wang, K. Zhao   Arch. Elect. Eng.

development strategy [1]. More and more distributed generation (DG) is connected, which leads to problems such as limited fault current, unclear fault characteristics and complex harmonics in fault location of distribution networks, and seriously affects the correctness of fault diagnosis. Therefore, with the access of DG, the distribution network has put forward higher requirements for fault location technology.

The existing method simplifies the protection device by using the traditional single ended three section current protection configuration [2], but faces the following problems:

1. The distribution network has uneven distribution of line parameters. There are branches on the main feeder, and the main feeder and side branch contain intermediate loads. After the fault occurs, the fault impedance is large and the fault current level is low [3].

2. It is difficult to extract and analyze fault features. DG in an active distribution network is usually connected to the distribution network through power electronic devices. This will lead to the limitation of fault current, unclear fault characteristics and complex harmonics in the active distribution network, which will seriously affect the correctness of fault diagnosis and location.

3. Traditional fault location methods are difficult to apply. Due to the mutual coupling between distribution network faults and power quality problems, the weak characteristics and high-frequency transient characteristics of distribution network faults become more obvious, and the difficulty of fault diagnosis and processing increases, which makes fault location with DG a technical problem [4].

In recent years, many scholars have carried out a lot of research on fault location methods of distribution networks with DG. Divided into two categories: a direct method and indirect method [5]. The direct method is mainly combined with the topology to analyze the over-current and voltage loss when the fault occurs. The transient traveling wave-based method [6], matrix algorithm [7], optimization algorithm [8], are widely used. This kind of method has a simple principle, strong universality, and high accuracy, but it also requires a high adoption rate of measurement and data processing devices, which is no longer applicable to a modern intelligent power grid. The indirect method mainly uses artificial intelligence technology to summarize and calculate the information uploaded based on the feeder terminal unit (FTU) and build a historical fault knowledge model of multiple factors [9]. Such methods include the Petri net [10], linked list method [11], genetic algorithm [12] and Bayesian network [13]. This method has a direct modeling mode and high fault tolerance, but it is easy to cause information loss and distortion, so it requires high communication and computing.

In modern distribution networks, a micro-synchronous phasor measurement unit (μ-PMU) provides a comprehensive and reliable data base for distribution network positioning technology based on artificial intelligence [14, 15]. The limited utilization of the active distribution network fault location method of the μ-PMU has good adaptability to fault types including a short circuit and broken line, but it is mainly aimed at the single-fault location of distribution networks, not multiple-fault location [16]. In the distribution network fault location based on machine learning (ML), convolutional neural networks (CNNs) first used the original and sampled data of three-phase voltage and current signals of the fault category and non-fault category for fault detection, showing better performance in terms of accuracy and computation [17]. The CNN fault area location model based on migration learning can quickly complete accurate fault area location in the case of small samples and is not affected by fault factors such as transition

resistance [18]. The method of Graph Attention Network (GAT) has a good application effect in different distribution network topology changes and scenarios, but it cannot be applied to distribution networks with completely different topology structures, and its universality is not high enough [19, 20]. The Bayesian network organically combines causal knowledge and prior probability information, uses probability theory to deal with the uncertainty caused by conditional correlation between different knowledge components, and uses the Bayesian theorem to calculate a posterior probability, which can calculate the fault probability of each component under the current fault symptom, to achieve fault diagnosis and location [21]. The above methods do not need to establish an accurate mathematical model, but there are still the following challenges in the actual research process:

1. It is difficult to obtain data samples.
2. Model precision training is difficult.

Based on the above, this paper presents a fault location method for distribution networks with DG based on traditional phase-mode transformation and ML. Karrenbauer transformation can transform the phase domain system into a module domain system without electromagnetic coupling, eliminating the difficulties in analyzing the transient process of single-phase ground fault directly in the phase domain. The calculated Karrenbauer current phase angle difference is used to establish the fault feature sample set to solve the problem of data sample acquisition. According to the fault type and area location, multiple binary classification problems can be superimposed into multiple classification problems, and SVR is used as the training sample set of an agent prediction model to realize the diagnosis of fault sections and their types. Finally, a distribution network system model is built, and the comparison analysis and verification show that the proposed method can quickly identify fault sections and types.

2. Distribution network fault location method based on Karrenbauer transform

A PMU is a device for measuring, outputting and dynamically recording the synchronous phase. The device uses the global positioning system (GPS) to measure the amplitude and phase information of electrical quantities at nodes with high accuracy, synchronization and real-time. The amplitude and phase errors are only 0.5% and 0.01° [22, 23]. To isolate faults quickly and accurately in the active distribution network with effectively grounded neutral points and diagnose faults in a timely and comprehensive manner, this paper uses μ-PMU-measured current and phase angle information data at both ends of the section for calculation.

2.1. Fault judgment principle of distribution network section with DG

In the distribution network, the distribution line is short. After the DG is connected, the active distribution network can be divided into several double ended power supply sections without branches according to the line nodes. This section can be divided into two situations, one is that there is power supply access at both ends of the section, and the other is that there is power supply access at one end and load connection at the other end. As shown in Fig. 1, the node currents at the nodes p and q at both ends of the section are I_p and I_q, respectively (the reference positive

direction defining the current in the system is from the node p to the node q).

$$\Delta\theta(p, q) = \theta_p - \theta_q , \tag{1}$$

where θ_p and θ_q are the phase angles of the single-phase full current at the nodes p and q, respectively.
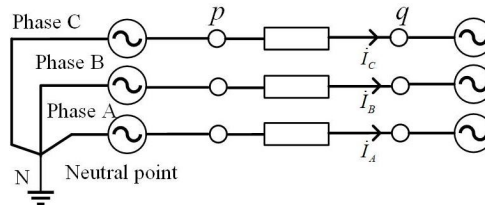

Fig. 1. Schematic diagram of three-phase network in a section
(both ends relate to power supply)

Let the three-phase currents at the node p of the section be $\dot{I}_A$, $\dot{I}_B$ and $\dot{I}_C$, respectively,

$$\begin{cases} \dot{I}_A = \sqrt{2}I_A \sin(\omega t + \theta_A) = I_A \angle \theta_A \\ \dot{I}_B = \sqrt{2}I_B \sin(\omega t + \theta_B) = I_B \angle \theta_B \\ \dot{I}_C = \sqrt{2}I_C \sin(\omega t + \theta_C) = I_C \angle \theta_C \end{cases} , \tag{2}$$

where: I_A, I_B and I_C are the effective values of the three-phase currents, θ_A, θ_B and θ_C are the phase angles of the three-phase current phasors.

During normal operation, the phase angle difference of the single-phase full current in the section is 0. In case of fault, the full current phase angle difference of the fault phase is not 0, but it is difficult to distinguish the ground short-circuit fault and the phase to phase short-circuit fault in two phases, and it is necessary to provide criteria for further judgment.

In order to overcome the above problems and realize the rapid preparation and positioning judgment of the fault section, the current phase angle difference under phase-mode transformation is used to realize the fault location of the section.

2.2. Distribution network fault criterion based on Karrenbauer transform

In the distribution network, it is difficult to directly analyze the transient process of single-phase grounding fault in the phase domain because of the asymmetry of line parameters in the medium and low-voltage distribution network and the electromagnetic coupling between the three-phase systems. Therefore, it is necessary to transform the phase domain system into a model domain system without electromagnetic coupling through coordinate transformation. The commonly used symmetric component transformation (C.L. Fortescue transformation) has complex factors, but it is not suitable for asymmetric systems. The result of instantaneous symmetric component transformation is complex and has no definite meaning [24]. Biaxial transformation (R.H. Park transformation) is mostly used for synchronous motor analysis and the analysis of symmetrical faults, while E. Clarke transformation has a complex fault model when single-phase

grounding is applied [25]. Therefore, the Karrenbauer transform is adopted in this paper to reduce the fault judgment amount and change the three-phase system to a 0-mode α Mold β Mold system.

For the balanced line, its parameter matrix is not affected, while for the unbalanced line, based on Karrenbauer transformation, the high-precision decoupling is realized through matrix diagonalization. The process is as follows:

Karrenbauer transformation matrix is

$$T_m = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{bmatrix}. \tag{3}$$

The relationship between phase current and mode current is expressed as:

$$I_m = T_m^{-1} I_P . \tag{4}$$

Simplification

$$\begin{bmatrix} \dot{I}_0 \\ \dot{I}_\alpha \\ \dot{I}_\beta \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} \dot{I}_A \\ \dot{I}_B \\ \dot{I}_C \end{bmatrix}, \tag{5}$$

where: $\dot{I}_0$ is the ground modulus obtained by the three-phase current of the node through Karrenbauer transformation, $\dot{I}_\alpha$ and $\dot{I}_\beta$ is the two line moduli obtained by the three-phase current of the node through Karrenbauer transformation.

Combining existing Eq. (5) and Eq. (2) above:

$$\begin{cases} \dot{I}_\alpha = \frac{1}{3} \left(\dot{I}_A - \dot{I}_B \right) \\ \dot{I}_\beta = \frac{1}{3} \left(\dot{I}_A - \dot{I}_C \right) \end{cases}. \tag{6}$$

So, we can get that after Karrenbauer transformation, the α Mold β phase angle of mode current is:

$$\begin{cases} \theta_\alpha = \arctan \left(\dfrac{I_A \sin \theta_A - I_B \sin \theta_B}{I_A \cos \theta_A - I_B \cos \theta_B} \right) \\ \theta_\beta = \arctan \left(\dfrac{I_A \sin \theta_A - I_C \sin \theta_C}{I_A \cos \theta_A - I_C \cos \theta_C} \right) \end{cases}. \tag{7}$$

2.3. Characteristic analysis before and after fault

The load rate of the main line in the distribution network with DG is often low. For normal operation, the distributed capacitance current is negligible compared with the load current on the line. Therefore, after Karrenbauer transformation, the section line model in normal operation is the α Mold sum β Phase angle difference analysis of mode current. The system model is shown in Fig. 2.

During normal operation, the current phasor flowing through the section is $\dot{I}$, the current phasors at the nodes p and q at both ends of the section are, respectively, $\dot{I}_p$ and $\dot{I}_q$, and the
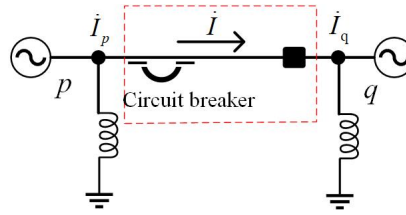
Fig. 2. System model of a section under normal conditions

voltage phasors at the corresponding two nodes are:

$$\begin{cases} \dot{U}_p = U_p\angle\theta_p \\ \dot{U}_q = U_q\angle\theta_q \end{cases}. \tag{8}$$

When the three-phase line is in normal operation without fault, it can be obtained that

$$\dot{I} = \dot{I}_p = \dot{I}_q. \tag{9}$$

The equivalent impedance of this section is:

$$\Delta\dot{Z}_s = R_s + j \cdot X_s = Z_s\angle\theta_s, \tag{10}$$

where R_s is the equivalent resistance of the section and X_s is the equivalent reactance of the section. By combining Eqs. (8), (9) and (10) we get:

$$\begin{cases} \dot{I} = \dfrac{\dot{U}_p - \dot{U}_q}{\Delta\dot{Z}_s} = \dfrac{U_p\angle\theta_p - U_q\angle\theta_q}{Z_s\angle\theta_s} = \sqrt{\dfrac{U_p^2 + U_q^2 - 2U_pU_q\cos(\theta_p - \theta_q)}{R_s^2 + X_s^2}}\angle(\theta_{\Delta U} - \theta_s) \\ \theta_{\Delta U} = \arctan\left(\dfrac{U_A\sin\theta_A - U_B\sin\theta_B}{U_A\cos\theta_A - U_B\cos\theta_B}\right) \\ \theta_s = \arctan\left(\dfrac{X_s}{R_s}\right) \end{cases}, \tag{11}$$

where: $\theta_{\Delta U}$ is the phase angle of the voltage drop from the node p to the node q, and θ_s is the impedance angle. In combination with Eq. (9), it can be obtained that the current phase angle under normal conditions is:

$$\Delta\theta(p, q) = \theta_{\Delta U} - \theta_s = \theta_p - \theta_q = 0°. \tag{12}$$

To sum up, when the actual section operates normally, after Karrenbauer transformation α Mold sum β the phase angle differences of the mode currents are all $0°$. Similarly, ignore the parallel admittance of the line, and analyze the section equivalent line model in the case of fault, as shown in Fig. 3.

After the fault, the current phasors flowing through the nodes p and q at both ends of the section are $\dot{I}_{pf}$ and $\dot{I}_{qf}$, respectively, and the voltage phasors at the corresponding two nodes are

$$\begin{cases} \dot{U}_{Pf} = U_{Pf}\angle\theta_{pf} \\ \dot{U}_{qf} = U_{qf}\angle\theta_{qf} \end{cases}. \tag{13}$$
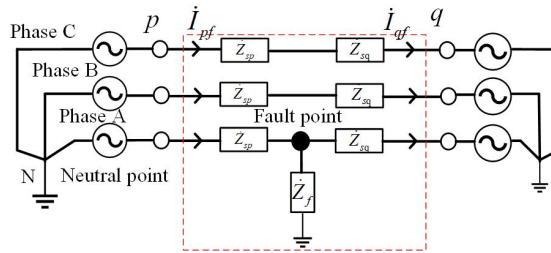
Fig. 3. System model of a section after phase A grounding fault

Its equivalent impedance is divided into two parts $\dot{Z}_{sp}$ and $\dot{Z}_{sq}$, where:

$$\Delta \dot{Z}_{sf} = \dot{Z}_{sp} + \dot{Z}_{sq} = R_{sf} + j \cdot X_{sf} = Z_{sf} \angle \theta_{sf} . \tag{14}$$

Similarly, the fault currents flowing through the nodes p and q at both ends of the section in the case of fault are corresponding to $\dot{I}_{pf}$ and $\dot{I}_{qf}$, and the expressions of the currents flowing in the section are, respectively:

$$\begin{cases} \dot{I}_{pf} = \sqrt{\dfrac{U_{pf}^2 + U_f^2 - 2U_{pf} U_f \cos(\theta_{pf} - \theta_f)}{R_{sf}^2 + X_{sf}^2}} \angle(\theta_{\Delta Uf} - \theta_s(p)) \\[2mm] \theta_{\Delta U}(p) = \arctan\left(\dfrac{U_{pf} \sin\theta_{pf} - U_f \sin\theta_{sf}}{U_{pf} \cos\theta_{pf} - U_f \cos\theta_{sf}}\right) \\[2mm] \theta_s(p) = \arctan\left(\dfrac{X_{sf}}{R_{sf}}\right) \end{cases}, \tag{15}$$

$$\begin{cases} \dot{I}_{qf} = \sqrt{\dfrac{U_{qf}^2 + U_f^2 - 2U_{qf} U_f \cos(\theta_{qf} - \theta_f)}{R_{sf}^2 + X_{sf}^2}} \angle(\theta_{\Delta Uf} - \theta_s(q)) \\[2mm] \theta_{\Delta Uf}(p) = \arctan\left(\dfrac{U_{qf} \sin\theta_{qf} - U_f \sin\theta_{sf}}{U_{qf} \cos\theta_{qf} - U_f \cos\theta_{sf}}\right) \\[2mm] \theta_s(q) = \arctan\left(\dfrac{X_{sf}}{R_{sf}}\right) \end{cases}. \tag{16}$$

To sum up, the expression of the current phase angle difference $\Delta\theta_{If}(pq)$ in this section after fault is:

$$\Delta\theta_{If}(p,q) = \Delta\theta_{If}(p) - \Delta\theta_{If}(q) = \left(\theta_{\Delta Uf}^p - \theta_s(p)\right) - \left(\theta_{\Delta Uf}^q - \theta_s(q)\right). \tag{17}$$

In the above formula, it should be noted that $\theta_s(p) = \theta_s(q)$ holds in the actual line, it can be obtained that:

$$\Delta\theta_{If}(p,q) \neq 0°. \tag{18}$$

According to the analysis of the characteristics before and after the fault, the faults of the active distribution network can be divided into four categories: single phase ground short-circuit fault

(AG, BG, CG), two-phase ground short-circuit fault (ABG, BCG, ACG), two-phase short-circuit fault (AB, BC, AC) and three-phase short-circuit fault (ABCG). When Karrenbauer transform is used to analyze α Mold β, the phase angle difference of the mode current can detect various faults in the section, especially the α module can detect a ground fault, the β module can detect the interphase fault.

2.4. Simulation analysis

To verify the effectiveness of the Karrenbauer transformation criterion for the fault location of distribution networks in this paper, the standard microgrid proposed by CIGRE in document [26] is selected.

The network used in this paper is a reference low-voltage (LV) grid developed in Microgrids, which was later adopted by GIGRE TF C6.04.02 as a benchmark LV system, consisting of LV feeders. The benchmark network maintains the important technical characteristic of real utility grids, whereas, at the same time, it dispenses with the complexity of actual networks, to permit efficient modeling of microgrid operation for steady state and transient simulations. The power grid model is built with Matlab/Simulink software for verification. The distribution network system includes 10 double-ended sections. To reflect the fault location advantages of the algorithm in this paper in the multi-source distribution network, 4 DGs are added to the distribution network, and the simulation model is shown in Fig. 4.
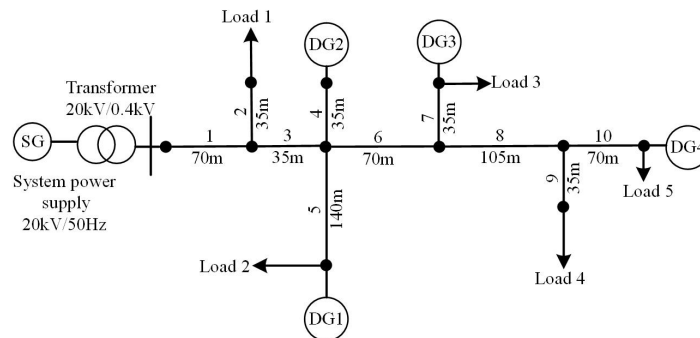

Fig. 4. Distribution network simulation system model

2.4.1. System parameter

In this model, the total installed capacity of DG is 485.25 kW, the permeability is 40%. Different types of DG access are considered in the system. In this paper, the sampling frequency is 50 Hz, the simulation time is 0.2 s, and the fault occurrence time is 0.065 s. Among them, since Section 6 is a three-phase balanced double-ended power supply type section, and the distance between DG at both ends is relatively close, which can better verify the effectiveness of this method. Therefore, the fault point is set in Section 6, and the transition resistance of the fault is 0.01 Ω. The specific parameters are shown in Table 1.

Figure 5 is a comparison diagram of the change of zero sequence current in time zone 6 when the system is in normal operation and when the phase A ground short-circuit fault occurs. When

Table 1. Specific parameters of the distribution network simulation system model

Parameter name		Specific values	
SG		20 kV/50 Hz	
Transformer		20/0.4 kV, 50 Hz, 400 kVA	
DG type	DG1	Batteries	50 kW
	DG2	Microturbine	31.25 kW
	DG3	Doubly fed induction generator (DFIG)	100 kW
	DG4	Photovoltaics (PV)	100 kW
Load	Load 1	15 kW	
	Load 2	72 kW	
	Load 3	55 kW	
	Load 4	15 kW	
	Load 5	47 kW	

the neutral point is not grounded, since the output fault current of each circuit is caused by the short circuit between the grounding capacitor and the fault point, it can be clearly seen that the fault line current changes when the fault occurs in 0.065 s.
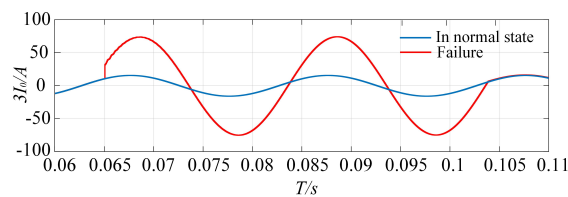

Fig. 5. Zero-sequence current diagram under normal state and fault state

2.4.2. Analysis of simulation results

The Karrenbauer transformation criterion is used to locate the fault of distribution networks, and the influence of various fault types, different transition resistances and line parameters on the system model is verified. At the same time, the robustness under different types of DG is verified.

1. Multiple fault types

The same fault location (Section 6) and transition resistance (0.01 Ω) are set in the system model, the fault types of single-phase ground short circuit, two-phase ground short circuit, phase-to-phase short circuit and three-phase short circuit are simulated and verified, respectively. To ensure accurate identification and analysis of simulation results, the threshold value of the phase angle difference of Karrenbauer transform mode current is set to 0°. Figure 6 is a simulation waveform diagram of the phase angle difference of different fault types based on the Karrenbauer transformation criterion.
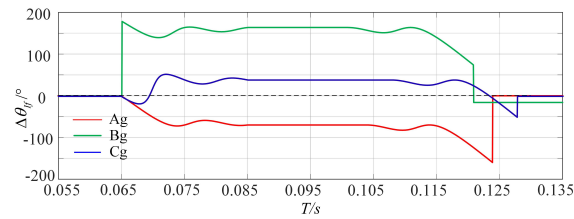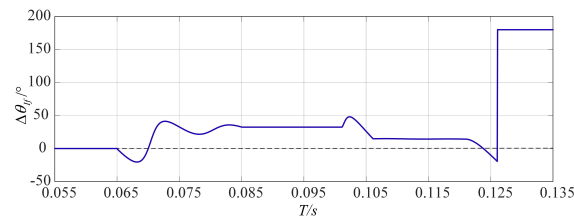
(a) When a single-phase grounding fault occurs
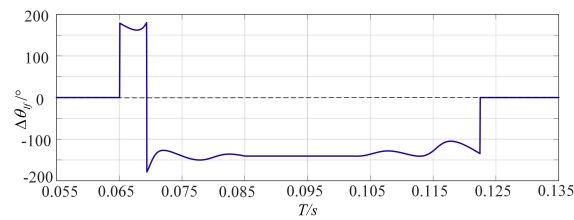
(b) When AB two-phase ground short-circuit fault occurs
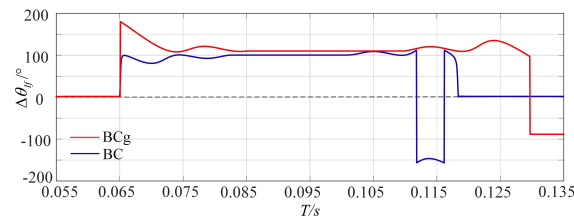
(c) When BC two-phase ground short-circuit and BC phase fault occurs

(d) AC phase-to-phase short circuit fault

Fig. 6. Simulation waveform diagram under various fault states

Figure 6(a) is a waveform diagram of phase angle difference when a single-phase ground short-circuit fault (AG, BG, CG) occurs in combination with Eq. (7). Figure 6(b) is a waveform diagram of phase angle difference when phase AB is short circuited to ground. In combination with the above analysis of Karrenbauer transform mode current, the threshold value of phase angle difference fluctuates greatly. Figure 6(c) is a waveform comparison diagram when BC phase-to-phase fault and BC phase-to-ground short-circuit fault occur. It is known that the phase angle difference fluctuates above the threshold value when two-phase to ground fault occurs, and the waveform is obviously unstable when phase-to-phase fault occurs. Figure 6(d) shows that the

phase angle difference fluctuates above and below the threshold value when AC phase-to-phase short-circuit fault occurs, and the fluctuation range is small.

To sum up, the phase angle difference of Karrenbauer transform mode current changes significantly after the fault occurs in Section 6 at 0.065 s. The simulation result analysis is consistent with the previous theory, and multi type fault location can be realized.

2. Effect of transition resistance

When the fault location is set at the A-phase grounding short-circuit fault of Section 6, the fault location results are affected when the transition resistance is set to 0.01 Ω, 50 Ω, 100 Ω and 300 Ω, respectively. As shown in Table 2, the positioning accuracy is above 98%, indicating that the proposed method is not easily affected by the size of transition resistance.

Table 2. Analysis of fault location results of different transition resistors

Transition resistance/Ω	Positioning accuracy/%	Fault type
0.01	99.5	Ag
50	98.98	Ag
100	98.5	Ag
300	98.37	Ag

3. Influence of changing line parameters

The long term operation of the distribution network will be the affected by natural environment, climate, and other factors, which will make the actual value of the saving impedance in the system model deviate from the stored value set in advance in the system model. Therefore, considering the influence of line parameter changes, all parameter values stored in advance during simulation are adjusted to 0.9 to 1.1 times of the original value. As shown in Table 3, the results show that the fault location is not affected by the change of line parameters.

Table 3. Fault location results under different fault types when line parameters change

Change times of line parameters	Fault type	Error compared with normal line parameters
0.9	Single-phase grounding short circuit	0.0015
	Two-phase grounding short circuit	0.0048
	Phase-to-phase short circuit	0.0023
	Three-phase grounding short circuit	0.0035
1.1	Single-phase grounding short circuit	0.003
	Two-phase grounding short circuit	0.0045
	Phase-to-phase short circuit	0.0012
	Three-phase grounding short circuit	0.0067

3. A distribution network fault location model based on SVR

Although the Karrenbauer transformation-based method in the previous section can quickly and accurately identify the type of fault, a single-phase modulus cannot reflect the characteristics of all faults, which is not conducive to the calculation of a large amount of data in the smart distribution network, while the SVR-based agent prediction model can solve the rapid calculation of complex features in the smart distribution network.

When a single-phase ground fault occurs in the distribution network, the three-phase voltage and current transient values will show nonlinear and non-stationary signals. Support vector machine regression (SVR) is a new general functional machine learning method. Compared with other traditional learning methods, it has better generalization ability in solving small sample analysis and nonlinear problems. It has significant advantages and is widely used in binary classification, pattern recognition and regression methods. The SVR-based fault location flowchart of the DG-containing distribution network is shown in Fig. 7.
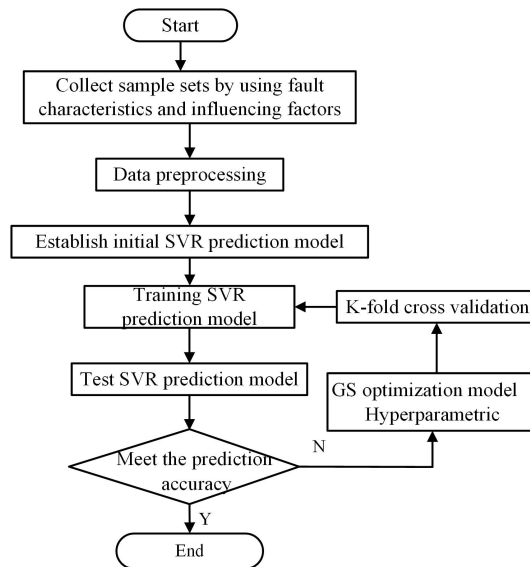

Fig. 7. SVR based fault location process of distribution network with DG

3.1. Data preprocessing

Due to the information correlation between features and the appearance of overlapping data, if all of them are used as the feature input of training sets and test sets for fault classification and identification and fault location, the data volume will be large, and the fault type identification and location will be inaccurate. Therefore, feature selection and extraction shall be carried out for samples before training. Three-phase zero-sequence voltage, three-phase zero sequence current and the phase angle difference of Karrenbauer mode current are used as input for the sample set, and the fault category as well as the fault location section are used as sample output.

To improve the accuracy of the fault classification and location prediction model training of the distribution network with DG, the sample output data shall be normalized as shown in Eq. (19) before training.

$$y_i = \frac{\Delta_i - \Delta_{\min}}{\Delta_{\max} - \Delta_{\min}}, \quad (i = 1, 2, \ldots, n),$$ (19)

where: $\Delta_{\max}$ and $\Delta_{\min}$ are the maximum and minimum values of the sample output data, respectively, Δ_i is the fault category and location under the DG distribution network prediction model, y_i represents the sample response value of x_i after normalization.

3.2. Fault location prediction model based on SVR

As a statistical machine learning theory, support vector machines (SVMs) were introduced by Vapnik and Corte in 1995 as an alternative technology for polynomials, radial functions, and multi-layer perceptual classifiers. The weights of neurons are determined by solving quadratic programming (QP) problems with linear, inequality and equality constraints rather than solving nonconvex problems. It is an unconstrained minimization problem. Hu Wei, Liu Keyan, and others proposed an intelligent distribution network fault location method based on local anomaly factors and the SVM classification of voltage data. A supervised learning algorithm is used to analyze data in classification and regression analysis. The fault data in this paper are trained based on an SVM algorithm, and different fault types in the fault area are reasonably classified [27, 28]. However, at this stage, the topology of intelligent distribution networks is complex, and the selected fault input features are relatively simple.

The basic principle of SVR is to find the optimal classification surface, so that for the linear inseparable problem, the kernel function is usually introduced to map the low-dimensional nonlinear inseparable samples to the high-dimensional linear separable space, to find the optimal classification surface in the transformed high-dimensional space. The introduction of the kernel function realizes the linear classification after the nonlinear transformation, but the computational complexity is not increased. The modeling and solution process is as follows [29]:

1. Given training sample set

$$\boldsymbol{X} = \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_i, y_i), \ldots, (\boldsymbol{x}_n, y_n)\},$$ (20)

where x_i and y_i are the phase angle difference of the group i zero sequence current, zero sequence voltage, Karrenbauer mode current and their fault type, fault location section, i.e., input and output of fault feature samples; n is the number of training samples.

2. Construct regression model

$$f(\boldsymbol{X}) = w^T \varphi(\boldsymbol{X}) + b,$$ (21)

where: $f(X)$ is the regression estimation function of the support vector, $\varphi(X)$ and X is the nonlinear function that maps the sample X to the high-dimensional feature space; w and b are the undetermined parameters of the model, which are also the key to SVR model training.

3. Model parameter solution

As shown in Fig. 8, a spacing band with width 2ε is constructed with $f(X)$ as the center. It is assumed that the maximum deviation of ε between $f(\boldsymbol{x}_i)$ and y_i can be tolerated, that is, when

the absolute value of the difference between $f(\boldsymbol{x}_i)$ and y_i is greater than ε, the ε insensitive loss function l_ε is introduced (ε – insensitive loss).

$$l_\varepsilon(\boldsymbol{x}_i, y_i) = \begin{cases} 0, & |y_1 - f(\boldsymbol{x}_i)| \le \varepsilon \\ |y_1 - f(\boldsymbol{x}_i)|, & -|y_1 - f(\boldsymbol{x}_i)| > \varepsilon \end{cases}. \tag{22}$$
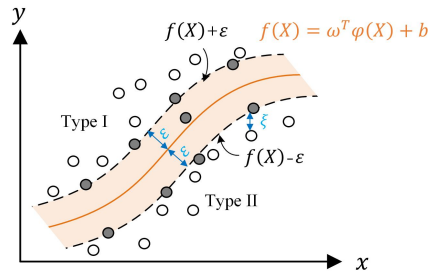

Fig. 8. Schematic diagram of SVR

When the training sample point falls into the red interval band, when the difference between the return value $f(\boldsymbol{x}_i)$ of the regression function and the sample output value y_i is less than or equal to ε, the loss is considered to be 0, that is, the training sample is correctly predicted, and the sample point falling outside the interval band is also called the support vector. Therefore, the SVR problem can be formalized as:

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} l_\varepsilon(f(\boldsymbol{x}_i) - y_1). \tag{23}$$

Relaxation variables ξ_i and $\hat{\xi}_i$ are introduced to characterize the degree to which the sample does not satisfy the constraint. The above formula can be rewritten as:

$$\begin{cases} \max\limits_{\xi_i, \hat{\xi}_i, b, \omega} L = \min \dfrac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} f\left(\xi_i + \hat{\xi}_i\right) \\[2mm] \text{s.t.} : \begin{cases} f(\boldsymbol{x}_i) - y_i \le \varepsilon + \xi_i \\ f(\boldsymbol{x}_i) - y_i \le \varepsilon + \hat{\xi}_i \\ \xi_i \ge 0, \ \hat{\xi}_i \ge 0, \ i = 1, 2, \ldots, n \end{cases} \end{cases}, \tag{24}$$

where: L is the objective function, C is the regularization constant, also known as the penalty coefficient. The larger C, the greater the penalty for the sample whose error exceeds ε; ε specifies the error limit of the SVR function.

However, Eq. (24) is still a quadratic programming problem. In order to facilitate the solution and introduce the Lagrange function according to the structural risk minimization criterion, we

convert L into a dual form, as shown in Eq. (25).

$$
\begin{cases}
\max\limits_{\xi_i,\hat{\xi}_i,b,\omega} L = \sum_{i=1}^{n} \left(y_i\left(\hat{a}_i - \alpha_i\right) - \varepsilon\left(\hat{a}_i + \alpha_i\right)\right) - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\hat{a}_i - \alpha_i\right)\left(\hat{a}_j - \alpha_j\right)K\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) \\
\text{s.t.}: \begin{cases} \sum_{i=1}^{n}\left(\hat{a}_i - \alpha_i\right) = 0 \\ 0 \le \alpha_i, \quad \hat{a}_i \le C \end{cases}
\end{cases}
, \quad (25)
$$

where $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \varphi^T(\boldsymbol{x}_i)\varphi(\boldsymbol{x}_j)$ is the kernel function. α_i, $\hat{a}_i$ is the Lagrange multiplier. The optimal solution obtained by solving Eq. (25) is:

$$
\begin{cases}
\boldsymbol{a} = [\alpha_1, \alpha_2, \ldots, \alpha_n] \\
\hat{\boldsymbol{a}} = [\alpha_1, \alpha_2, \ldots, \alpha_n]
\end{cases}. \quad (26)
$$

Therefore, it can be further concluded that the parameters w and b are

$$
\begin{cases}
w = \sum_{i=1}^{n}\left(\alpha_i - \hat{a}_i\right)\varphi(\boldsymbol{x}_i) \\
b = \frac{1}{n_{sv}}\sum_{0<\alpha_i<C}\left[y_i - \sum_{x_i \in sw}\left(\alpha_i - \hat{a}_i\right)K\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) - \varepsilon\right] \\
\quad + \frac{1}{n_{sv}}\sum_{0<\alpha_i<C}\left[y_i - \sum_{x_j \in sw}\left(\alpha_j - \hat{a}_j\right)K\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) + \varepsilon\right]
\end{cases}. \quad (27)
$$

where n_{sv} is the number of support vectors. Therefore, the regression function of the support vector machine can be obtained by bringing Eq. (27) into Eq. (21):

$$
f(\boldsymbol{x}) = w^T\varphi(\boldsymbol{x}) + b = \sum_{i=1}^{n}\left(\hat{a}_i - \alpha_i\right)\varphi^T(\boldsymbol{x}_i)\varphi(\boldsymbol{x}) + b = \sum_{i=1}^{n}\left(\hat{a}_i - \alpha_i\right)K\left(\boldsymbol{x}_i, \boldsymbol{x}\right) + b. \quad (28)
$$

4. Selection of kernel function

To construct an SVR with good performance, the selection of a kernel function is the key. The selection of a kernel function includes two parts: one is the selection of a kernel function type, and the other is the selection of relevant parameters after determining the kernel function type. The Gaussian radial basis kernel function (RBF) selected in this paper is a kernel function with strong locality, few hyperparameters relative to a polynomial kernel function, high accuracy, and good approximation characteristics.

$$
K(\boldsymbol{x}_i, \boldsymbol{x}) = \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}\|^2}{2\sigma^2}\right) = \exp\left(-\gamma\|\boldsymbol{x}_i - \boldsymbol{x}\|^2\right), \quad (29)
$$

where $\gamma = 1/2\sigma^2$, that is, the width of the RBF γ. The larger the Gaussian distribution is, the thinner and longer it is, and it is easy to over fit γ. The smaller the Gaussian distribution is, the shorter and fatter the Gaussian distribution is, and the smoothing effect is too large to obtain a high accuracy on the training set, and it is easy to under fit.

3.3. SVR hyperparametric optimization based on GS-CV

To make full use of the limited data to obtain higher prediction accuracy, the super parameter optimization of the SVR model is carried out in combination with grid search and cross validation, and the best super parameter combination is determined. The steps are as follows:

Step 1: Determining the hyperparameters (C, γ) in the value range. The commonly used value range is $[2^{-10}, \ 2^{10}]$.

Step 2: Hyperparametric gridding. Due to the large span of the parameter range, take the logarithm of the value range of the super parameter with 2 as the base to obtain $(\log_2 C, \log_2 \lambda) \in [-10, 10]$. Cross values are taken at intervals of 0.25, and a total of $81 \times 81 = 6561$ kinds of super parameter combinations.

Step 3: K-fold cross validation. From 100 sets of sample data, 85 were taken as training data sets and 15 were taken as test data sets. Let $k = 5$, then divide the training data set into 5 pieces, and each piece of data is taken as the verification set in turn. Evaluate the quality of the model trained by the remaining 3 pieces of data, and finally take the hyperparameter with the smallest mean square prediction error as the optimal hyperparameter in the cross verification.

Step 4: Hyperparametric combined mesh search. Select the next group of hyperparametric combinations and repeat step 3 until the hyperparametric combination with the minimum mean square error is found as the optimal SVR model hyperparametric under the current sample set. When different parameter combinations (C, γ) correspond to the same cross validation accuracy, a set of parameters with smaller C is generally selected to improve the generalization ability of the model.

3.4. Accuracy evaluation standard

The prediction accuracy of an SVR proxy model needs to rely on the corresponding error criteria for performance evaluation. In this paper, the mean square error and correlation coefficient are selected as the prediction accuracy evaluation indicators of the proxy model.

1. The mean square error (MSE) is often used to describe the error between the predicted value and the sample value, and its expression is

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \,, \tag{30}$$

where: n is the number of tests samples, y_i is the theoretical response value of the source function, $\hat{y}_i$ is the predicted value of the source function proxy model.

The closer the mean square error is to 0, the higher the prediction accuracy is.

2. Coefficient of determination (R^2) is often used to describe the fitting degree of the regression model to the observed value, and its expression is:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \hat{y})^2} \,. \tag{31}$$

The value range of the determination coefficient is $[-\infty, 1]$. The closer R^2 is to 1, the more significant the regression equation is, indicating the poor prediction ability of the model.

4. Result analysis

4.1. Accuracy evaluation and result analysis of proxy model

1. Fault location analysis

Taking the distribution network model in Fig. 4 as an example, the prediction performance is evaluated based on the simulation data under Karrenbauer transformation. The abscissa is the layout number, and the ordinate is the 10 sections included in the model. As can be seen from 9(a), in the case of faults in different sections, the real value is close to the predicted value, except for minor errors. In the training set prediction accuracy evaluation, the MSE is 0.00196, close to 0. The determination coefficient R^2 is about 0.92941, close to 1, and the positioning result is accurate. Input the test sample into SVR, output the label of the sample, and determine the fault section corresponding to the label. The fault area location accuracy is 98%, as shown in 9(b).
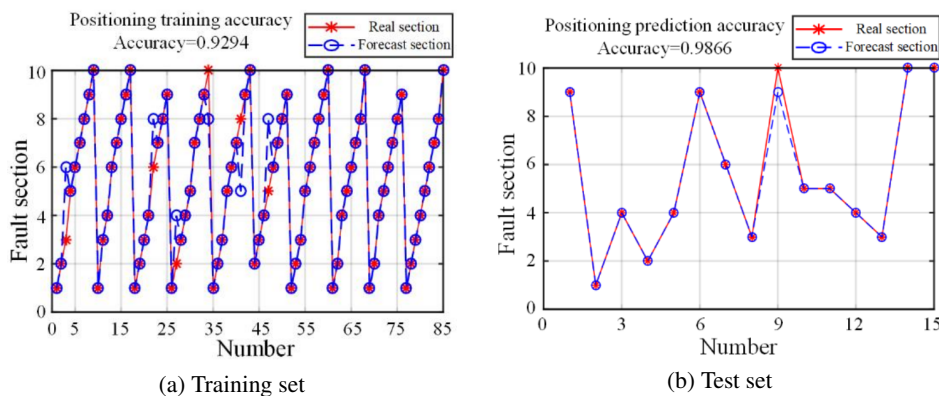

(a) Training set   (b) Test set

Fig. 9. Model prediction results of fault location training set and test set

2. Fault type analysis

According to the fact that the fault type is a multi classification problem superposed by multiple binary classification problems, the fault types in Section 2 are effectively used for SVR model training and testing. The abscissa is the layout number, and the labels of ordinates 1∼10 are 10 fault categories. As shown in Fig. 10(a), after training 85 samples, the sample points of measurement data are gathered in different areas, and the R^2 of the fault type recognition training set can reach 0.94118, so the training performance is very superior. It can be seen from Fig. 10(b) that the predicted value basically coincides with the real value, the prediction accuracy of the fault type identification test set can reach 93.3%, the MSE is 0.00123, approaching 0, and a good classification effect can be obtained.
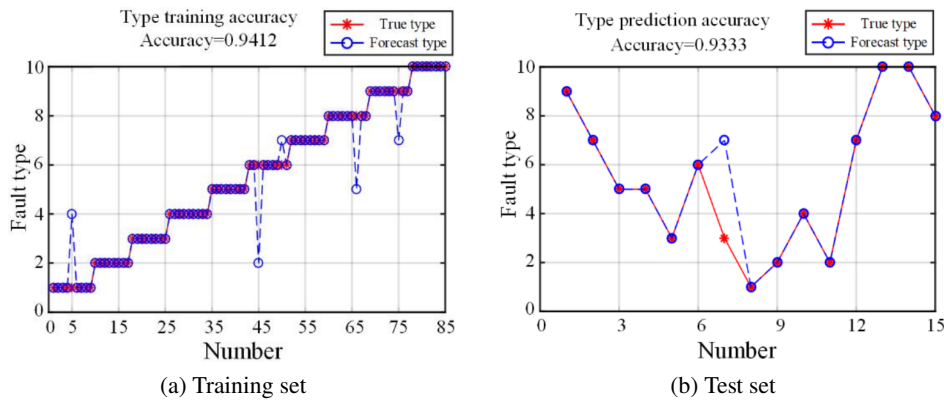
(a) Training set   (b) Test set

Fig. 10. Model prediction results of fault type set training and test set

4.2. Accuracy comparison of fault location in distribution network under different methods

To prove that the method proposed in this paper has high accuracy in fault location, this section also applies several common machine learning and neural network algorithms to fault location. The details are as follows:

1. The gradient lifting tree is used to establish the fault diagnosis model of low-voltage intelligent distribution networks, and a fixed number of interpolations are used to replace the measured values of specific branches, which can adapt to the change of network topology to a certain extent but cannot accurately locate the fault section.

2. A neural network algorithm is widely used in the transient protection and fault line selection of smart grids. The CNN can effectively avoid the over fitting of the training process in the recognition of fault sections and has a good feature extraction ability. But the algorithm is easy to fall into a local minimum, and the required samples are large, and the convergence speed is very slow.

3. The graph neural network (GNN)-based method is extended to the graph domain to mine the topological relationship between nodes, which improves the accuracy of location. However, the universality and practicability of the model under the change of network topology are not considered.

4. A Bayesian probability model evaluates the fitting degree between the possible fault section and the actual over-current information of the switch, and judges that the feeder section that can best explain the over-current information is the fault location result. This method has high fault tolerance. However, this method uses a single information source, which is limited by the problem that the fault section is misjudged when the current information is wrong.

5. Random forest (RF) is an integrated algorithm composed of decision trees, which can train and predict samples. Only consider the case where the fault occurs at the node, while the fault usually occurs in the distribution network circuit.

In conclusion, SVR requires fewer samples and has better convergence speed and generalization ability than other methods. In addition, SVR combined with traditional Karrenbauer

transform can fully mine the characteristic information of voltage and current during fault and establish a good sample library. Therefore, the SVR method has great advantages. The comparison results are shown in Table 4.

Table 4. Comparison of fault location results of different methods

Method	Suitability of power distribution system	Model training duration	Time required for positioning	Accuracy
Gradient lifting tree	95.68%	92.24s	1.43s	85.24%
CNN	93.54%	85.52s	0.97s	95.03%
GNN	82.93%	87.68s	1.05s	93.5%
Bayesian estimation	94.29%	90.44s	1.23s	82.13%
RF	80.75%	88.21s	0.98s	92.43%
SVR	96.16%	93.66s	0.91s	98.56%

5. Conclusions

In the face of difficulties in sample acquisition and model accurate training in distribution network fault location, this paper uses Karrenbauer transformation and an SVR agent prediction model to realize fault type identification and location of distribution networks with distributed generation. The following conclusions are drawn.

1. In view of the asymmetry of LV distribution network line parameters and the electromagnetic coupling between each phase line of the three-phase system, Karrenbauer transformation is used to calculate, deduce, and compare the current, voltage, phase angle and other information data measured before and after the fault at both ends of the section, obtain the phase angle difference of Karrenbauer transformation mode current, and preliminarily detect the fault section.

2. Because a single-phase mode transformation modulus cannot reflect the characteristics of all faults, using the electrical quantity information after Karrenbauer transformation, a sampling data set based on SVR is established to build a fast agent prediction model. The test results show that the decision coefficient R2 of the regression model is close to 1, which has a high prediction accuracy, and is conducive to the complex calculation of a large amount of data in the intelligent distribution network.

3. Through simulation verification, and compared with other methods, this method is not affected by the change of transition resistance and line parameters and can effectively use the electrical characteristics of the distribution network to better achieve the accurate location and type identification of different faults in the distribution network, with high accuracy and robustness.

References

[1] Sachin S., Niazi K.R., Verma K., Tanuj R., *Impact of battery energy storage, controllable load and network reconfiguration on contemporary distribution network under uncertain environment*, IET Generation, Transmission and Distribution, vol. 14, no. 21, pp. 4719–4727 (2020), DOI: 10.1049/iet-gtd.2020.0369.

[2] Jia K., Ren Z.F., Li L., Xuan Z.W., Thomas D., Steentjes S., *High-frequency transient comparison based fault location in distribution systems with DGs*, IET Generation, Transmission & Distribution, vol. 11, no. 16, pp. 4068–4077 (2017), DOI: 10.1049/iet-gtd.2017.0426.

[3] Liu L., *Faulted feeder identification and location for a single line-to-ground fault in ungrounded distribution system based on principal frequency component*, Archives of Electrical Engineering, vol. 69, no. 3, pp. 695–704 (2020), DOI: 10.24425/aee.2020.133926.

[4] Zhan H.Y., Liu K.Y., Sheng W.X., Meng X.L., *Literature Review and Prospects of Fault Diagnosis in Active Distribution Network*, High Voltage Engineering, pp. 1–12 (2022), DOI: 10.13336/j.1003-6520.hve.20211604.

[5] Liu J., Zhang X.Q., Tong X.Q., Zhang Z.H., Du H.W., Chen Y.K., *Fault location for distribution systems with distributed generations*, Automation of Electric Power Systems, vol. 37, no. 2, pp. 36–42+48 (2013), DOI: 10.7500/AEPS201208181.

[6] Deng F., Li X.R., Zeng X.J., Li Z.W., Guo J., Tang X., *A novel multi-terminal fault location method based on traveling wave time difference for radial distribution systems with distributed generators*, Proceedings of the CSEE, vol. 38, no. 15, pp. 4399–4409+4640 (2018), DOI: 10.13334/j.0258-8013.pcsee.172480.

[7] Xie L.W., Li Y., Luo L.F., Chen C., Cao Y.J., *Fault location method for distribution networks based on distance matrix and branch coefficient*, Proceedings of the CSEE, vol. 40, no. 7, pp. 2180–2191+2397 (2020), DOI: 10.13334/j.0258-8013.pcsee.182509.

[8] He S.M., Yuan Z.Y., Lei J.Y., Xu Q., Lin Y.H., Liu Y.L., Lin X.H., *Optimal setting method of inverse time over-current protection for a distribution network based on the improved grey wolf optimization*, Power System Protection and Control, vol. 49, no. 18, pp. 173–181 (2021), DOI: 10.19783/j.cnki.pspc.201351.

[9] Atencia-De la Ossa J., Orozco-Henao C., Marín-Quintero J., *Master-slave strategy based in artificial intelligence for the fault section estimation in active distribution networks and microgrids*, International Journal of Electrical Power and Energy Systems, vol. 148 (2023), DOI: 10.1016/j.ijepes.2022.108923.

[10] Yang F.R., Yu Y.J., *Fault diagnosis of distribution network based on time constrained hierarchical fuzzy Petri nets*, Power System Protection and Control, vol. 48, no. 2, pp. 99–106 (2020), DOI: 10.19783/j.cnki.pspc.190200.

[11] Zhang X.F., Lü F.P., Lü W.C., Zhang Y.Y., Deng F.Q., *Research on the on-line locating of single-phase-to-earth fault based on chain table algorithm*, Power System Protection and Control, vol. 40, no. 12, pp. 31–34+40 (2012).

[12] Dashtdar M., Bajaj M., *Fault location in distribution network by solving the optimization problem using genetic algorithm based on the calculating voltage changes*, Soft Computing, vol. 26, no. 17, pp. 8757–8783 (2022), DOI: 10.1007/S00500-022-07203-8.

[13] Jia K., Li L., Yang Z., Zhao G.K., Bi T.S., *Research on Distribution Network Fault Location Based on Bayesian Compressed Sensing Theory*, Proceedings of the CSEE, vol. 39, no. 12, pp. 3475–3486 (2019), DOI: 10.13334/j.0258-8013.pcsee.180705.

[14] PASAD S., KUMAR D.M.V., *Trade-offs in PMU and IED deployment for active distribution state estimation using multi-objective evolutionary algorithm*, IEEE Transactions on Instrumentation and Measurement, vol. 67, no. 6, pp. 1298–1307 (2018), DOI: 10.1109/TIM.2018.2792890.

[15] Pinte B., Quinlan M., Reinhard K., *Low voltage micro-phasor measurement unit (μPMU)*, In Proceedings of the 2015 IEEE Power and Energy Conference at Illinois (PECI), Champaign, IL, USA, pp. 1–4 (2015), DOI: 10.1109/PECI.2015.7064888.

[16] Zhang J.L., Gao Z.J., Wang Z.Y., Sun X.R. *et al.*, *Fault Location Method for Active Distribution Based on Finite μPMU*, Power System Technology, vol. 44, no. 7, pp. 2722–2731 (2020), DOI: 10.13335/j.1000-3673.pst.2019.2607.

[17] Rai P., Londhe N.D., Raj R., *Fault classification in power system distribution network integrated with distributed generators using CNN*, Electric Power Systems Research, p. 192, 106914 (2021), DOI: 10.1016/j.epsr.2020.106914.

[18] Meng Z.C., Du W.J., Wang H.F., *Distribution Network Fault Area Location Based on Deep Convolution Neural Network with Transfer Learning*, Southern Power System Technology, vol. 13, no. 7, pp. 25–33 (2019), DOI: 10.13648/j.cnki.issn1674-0629.2019.07.004.

[19] Mo H.J., Peng Y.G., Wei W., *SR-GNN Based Fault Classification and Location in Power Distribution Network*, Energies 16, vol. 60, no. 1, 433 (2023), DOI: 10.3390/en16010433.

[20] Li J.W., Wang X.J., He J.H., Zhang Y.J., Zhang D.H., *Distribution network fault location based on graph attention network*, Power System Technology, vol. 45, no. 6, pp. 2113–2121 (2021), DOI: 10.13335/j.1000-3673.pst.2020.2222.

[21] Cai B., Huang L., Xie M., *Bayesian Networks in Fault Diagnosis*, IEEE Transactions on Industrial Informatics, vol. 13, no. 5, pp. 2227–2240 (2017), DOI: 10.1109/TII.2017.2695583.

[22] Wang X.J., Ren X.Y., He J.H. *et al.*, *Distribution network fault location based on μPMU information*, Power System Technology, vol. 43, no. 3, pp. 810–818 (2019), DOI: 10.13335/j.1000-3673.pst.2018.2963.

[23] Von Meier A., Stewart E., McEachern A. *et al.*, *Precision micro-synchro phasors for distribution systems: a summary of applications*, IEEE Transactions on Smart Grid, vol. 8, no. 6, pp. 2926–2936 (2017), DOI: 10.1109/TSG.2017.2720543.

[24] Li L., Lu D.K., Guo J.N., *Cross-correlation Fault Positioning Based on Phase-mode Transformation*, Guangdong Electric Power, vol. 27, no. 10, pp. 68–73 (2014), DOI: 10.3969/j.issn.1007-290X.2014.10.014.

[25] Niu G., Zhou L., Pei W. *et.al.*, *Fault location method for low voltage active distribution network based on phase-angle differences of the Clark currents*, Proceedings of the CSEE, vol. 35(S), pp. 15–24 (2015), DOI: 10.13334/j.0258-8013.pcsee. 2015.S.003.

[26] Stavros P., Nikos H., Kai S., *A benchmark low voltage microgrid network*, CIGRE Symposium: Power Systems with Dispersed Generation: Technologies, Impacts on Development, Operation and Performances, Athens, Greece (2005).

[27] Hu W., Li Y., Cao Y.J. *et al.*, *Fault identification based on LOF and SVM for smart distribution network*, Electric Power Automation Equipment, vol. 36, no. 6, pp. 25–33 (2016), DOI: 10.16081/j.issn.1006Ł-6047.2016.06.002.

[28] Liu K.Y., Dong W.J., Xiao R.W., Wei J., Zhao W., *Fault identification and location of active distribution network based on SVM classification of voltage data*, Power System Technology, vol. 45, no. 1, pp. 2369–2379 (2021), DOI: 10.13335/j.1000-3673.pst.2020.0516.

[29] Zhou Z.H., *Machine learning*, Beijing, China: Tsinghua University Press, pp. 121–137 (2016).