Research Article

Fault Location Method Based on SVM and Similarity Model Matching

Chenyu Zhang , Xiaodong Yuan, Mingming Shi, Jinggang Yang, and Huiyu Miao

State Grid Jiangsu Electric Power Co., Ltd. Research Institute, Nanjing 211113, China

Correspondence should be addressed to Chenyu Zhang; zcy530@sina.com

Received 21 March 2020; Revised 2 June 2020; Accepted 16 June 2020; Published 22 September 2020

Academic Editor: Paolo Manfredi

Copyright © 2020 Chenyu Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To locate the fault location accurately and solve the problem quickly is the key to improve the power supply capacity of power grid. This paper presents a fault location method based on SVM fault branch selection algorithm and similarity matching. Firstly, an SVM-based fault branch filter classifier was constructed based on the positive sequence component feature matrix data of each monitoring point, which can accurately select the branch where the current fault is located. Then, based on the positive sequence voltage distribution characteristics, the Euclidean distance and Pearson correlation coefficient (PCC) are used to establish the similarity objective function of fault location. And then, the fault is accurately located by the objective function. Finally, the proposed method is validated by using an IEEE-14 node network. The results show that the proposed method is effective and accurate.

1. Introduction

Fast and accurate location of distribution network faults can effectively reduce the time of troubleshooting and blackout, reduce economic losses, and improve power supply reliability [1–5]. However, the distribution network has many branches with long and short lines and few measuring points, and it is difficult for protection to locate fault points accurately. Therefore, the method of locating fault areas and cooperating with manual patrol are mostly used in practice. With the improvement of automation and intellectualization of the distribution network in Jiangsu power grid, a large number of intelligent monitoring equipment for the distribution network have been put into use, such as FTU (feeder terminal unit) and PMU. These devices can measure the voltage and current of the distribution network in real time and provide data basis for accurate fault location [6, 7]. Therefore, the research in this paper is based on the fact that each node is equipped with monitoring instruments.

At present, scholars have done a lot of work on the accurate fault location of the distribution network. The main fault location methods are mainly divided into the traveling wave method and nontraveling wave method (including

impedance method, node matrix method, and fitting optimization method). The traveling wave method [8–10] determined the fault distance by measuring the propagation time of voltage and current traveling wave to the fault point. However, due to the uneven distribution of line impedance and too many branches in the distribution network, the location accuracy of the traveling wave positioning method was greatly affected. Additionally, the monitoring devices required for traveling wave positioning are expensive, and the practical application of engineering is difficult. The impedance method [11] calculated the impedance of the fault branch by measuring the voltage and current at the fault point, and then calculated the fault distance. Like the traveling wave method, the multibranch of the distribution network affected the accuracy of the method. This method is widely used in the high voltage transmission network. Reference [12] constructed a fault distance function based on the node impedance matrix by using the information of bus voltage drop during fault, and used the matching degree between the calculated value of voltage drop and the measured value to locate the fault, but the robustness to load change was poor and susceptible to transition resistance. In reference [13], the least square method was used to fit the

fault distance distribution function, but the error between the fault distance calculated by the fitting distribution function and the actual fault distance was large. The accuracy of the algorithm needs to be improved. In reference [14], the distribution function reflecting the change of fault distance was established by the node impedance matrix after fault, and the fault distance was determined by solving the corresponding location equation, but the location results are easily affected by the false fault point and measurement error. Reference [15] has been improved on the basis of reference [14]. Considering the influence of measurement error, the probability distribution curve of possible fault section was obtained by Monte Carlo simulation, and the fault location was determined according to the peak value of probability distribution. However, the precondition of this method is to identify the fault type and phase accurately, and the simulation calculation is too large to be applied in large-scale power grid.

In summary, most of the existing location methods either fail to overcome the influence of transition resistance on the location results or need to know the fault branch and fault type to locate, or the location results are greatly affected by measurement errors and fake fault points or need to traverse all locations of the network to search fault points, which results in a huge amount of calculation when the system scale is large.

To solve these problems, a fault location method based on the SVM fault branch selection and similarity model matching was proposed in this paper. The method utilized the distribution characteristics and their interrelationships of positive sequence voltage variations at each monitoring point to construct fault feature modes that were not affected by fault types and transition resistance. A fault branch selection database was established based on the simulation data and use this data to train SVM-based fault branch selecting classifier. And this SVM classifier is used to determine the branch where the current fault is located. Then, the fault location models of each branch were established with the fault distance parameter λ as the only variable. The concept of similarity index was defined based on the Euclidean distance and Pearson correlation coefficient (PCC) to measure the similarity between the current fault and the fault location model. When the similarity index obtained the optimal solution, the accurate fault location results λ can be obtained. The validity and accuracy of this method were verified by an example.

2. Fault Characteristics Based on Positive Sequence Voltage Variation

The location, type, and transition resistance of faults are the three variables that determine the voltage of each monitoring point. The schematic diagram of failure in the system is shown in Figure 1.

In this paper, the symmetrical component method is used to decompose the positive sequence component, and the positive sequence component is used to analyze the fault

FIGURE 1: A fault takes place in a network. A, B : the nodes of fault branch. M : the bus of monitoring point. F : failure point. Z_L : the branch line impedance between node A and node B . λ : the ratio of the distance between F and A to the distance between A and B . R_F : fault transition resistance.

electrical characteristics. In three phases, A phase is 120 degrees ahead of B phase, B phase is 120 degrees ahead of C phase, C phase is 120 degrees ahead of A phase, and the components with the same amplitude of three phases are called positive sequence components. The positive sequence driving point impedance $Z_{MF}^{(1)}$ between monitoring point M and fault point F can be calculated from the impedance matrix of positive sequence nodes, as shown in the following formula:

$$Z_{MF}^{(1)} = (1 - \lambda)Z_{MA}^{(1)} + \lambda Z_{MB}^{(1)}. \quad (1)$$

In formula (1), $Z_{MA}^{(1)}$ is the positive sequence driving point impedance between monitoring point M and node A . $Z_{MB}^{(1)}$ is the positive sequence driving point impedance between monitoring point M and node B . Superscript (1) indicates positive sequence.

Assuming that the positive sequence voltage before system failure of monitoring point M is $\dot{V}_M^{\text{std}(1)}$, the positive sequence component of short-circuit current at fault point F is $\dot{I}_F^{(1)}$. According to the superposition theorem, the positive sequence voltage after system failure of monitoring point M can be expressed as

$$\dot{V}_M^{(1)} = \dot{V}_M^{\text{std}(1)} - Z_{MF}^{(1)} \cdot \dot{I}_F^{(1)}. \quad (2)$$

In formula (2), $\dot{V}_M^{(1)}$ is the positive sequence voltage after system failure of monitoring point M . $\dot{V}_M^{\text{std}(1)}$ is the positive sequence voltage before system failure of monitoring point M . $Z_{MF}^{(1)}$ is the positive sequence driving point impedance between monitoring point M and fault point F . $\dot{I}_F^{(1)}$ is the positive sequence component of short-circuit current at fault point F .

Positive sequence voltage variation at monitoring point is defined as

$$\begin{aligned} \Delta V_M^{(1)} &= \left| \dot{V}_M^{\text{std}(1)} - \dot{V}_M^{(1)} \right| \\ &= \left| Z_{MF}^{(1)} \cdot \dot{I}_F^{(1)} \right|. \end{aligned} \quad (3)$$

It can be seen from the formula above that the change of positive sequence voltage at the monitoring point is only related to the positive sequence impedance between the monitoring point and the fault point and the positive sequence fault current. The positive sequence impedance between the monitoring point and the fault point represents the relative position information between them. For any fault position in the branch, the positive sequence impedance between the monitoring point and the fault position corresponds to it one to one and is not affected by the fault type and transition resistance.

It is assumed that there are two monitoring points in the system and two faults take place at point F successively, as shown in Figure 2. The fault type and transition resistance of fault 1 and fault 2 are different.

According to formula (3), it can be obtained that

$$\begin{aligned} \frac{\Delta V_{M_1 F_1}^{(1)}}{\Delta V_{M_2 F_1}^{(1)}} &= \frac{|Z_{M_1 F}^{(1)} \cdot \dot{I}_{F_1}^{(1)}|}{|Z_{M_2 F}^{(1)} \cdot \dot{I}_{F_1}^{(1)}|} = \frac{|Z_{M_1 F}^{(1)}|}{|Z_{M_2 F}^{(1)}|}, \\ \frac{\Delta V_{M_1 F_2}^{(1)}}{\Delta V_{M_2 F_2}^{(1)}} &= \frac{|Z_{M_1 F}^{(1)} \cdot \dot{I}_{F_2}^{(1)}|}{|Z_{M_2 F}^{(1)} \cdot \dot{I}_{F_2}^{(1)}|} = \frac{|Z_{M_1 F}^{(1)}|}{|Z_{M_2 F}^{(1)}|}, \\ \frac{\Delta V_{M_1 F_1}^{(1)}}{\Delta V_{M_2 F_1}^{(1)}} &= \frac{\Delta V_{M_1 F_2}^{(1)}}{\Delta V_{M_2 F_2}^{(1)}}. \end{aligned} \quad (4)$$

In formula (3), $\Delta V_{M_1 F_1}^{(1)}$ is the positive sequence voltage variation at monitoring point M_1 when fault 1 occurs at F . $\Delta V_{M_2 F_1}^{(1)}$ is the positive sequence voltage variation at monitoring point M_2 when fault 1 occurs at F . $\Delta V_{M_1 F_2}^{(1)}$ is the positive sequence voltage variation at monitoring point M_1 when fault 2 occurs at F . $\Delta V_{M_2 F_2}^{(1)}$ is the positive sequence voltage variation at monitoring point M_2 when fault 2 occurs at F .

From the formulas above, it can be seen that the ratios of positive sequence voltage variations of M_1 and M_2 at monitoring points are equal under two different faults. The characteristics of positive sequence voltage distribution of faults can be generalized: no matter the type and the transition resistance of faults, the positive sequence voltage variations of each monitoring point vary proportionally when different faults occur at the same location. That is to say, the change of positive sequence voltage is only related to the fault location.

3. Fault Location Method Based on SVM Fault Branch Selection and Similarity Matching

3.1. Fault Distance Model of Each Branch. Based on the above analysis, a fault location model based on network monitoring point information can be established according to the change characteristics of monitoring points. Assuming that the first and last nodes of the branch where the fault is located are l and r , the positive sequence voltage variations of all monitoring points can be obtained according to formula

FIGURE 2: Two faults on different branches. M_1 : the bus of monitoring point 1. M_2 : the bus of monitoring point 2. F : failure point. Fault 1: the first failure. Fault 2: the second failure. $Z_{M_1 F}^{(1)}$: the positive sequence impedance between M_1 and fault point F . $Z_{M_2 F}^{(1)}$: the positive sequence impedance between M_2 and fault point F . $\dot{I}_{F_1}^{(1)}$: the short-circuit positive sequence current at F when fault 1 occurs at F . $\dot{I}_{F_2}^{(1)}$: the short-circuit positive sequence current at F when fault 2 occurs at F .

(1) and (3). The sequence of positive sequence voltage variations of monitoring points is as

$$\begin{cases} \Delta \mathbf{V}^{(1)} = [\Delta V_{M_1}^{(1)}, \Delta V_{M_2}^{(1)}, \dots, \Delta V_{M_n}^{(1)}] \\ = |(1 - \lambda)Z_l^{(1)} + \lambda Z_r^{(1)}| \cdot |\dot{I}_F^{(1)}|, \\ \mathbf{Z}_l^{(1)} = [Z_{M_1 l}^{(1)}, Z_{M_2 l}^{(1)}, \dots, Z_{M_n l}^{(1)}], \\ \mathbf{Z}_r^{(1)} = [Z_{M_1 r}^{(1)}, Z_{M_2 r}^{(1)}, \dots, Z_{M_n r}^{(1)}]. \end{cases} \quad (5)$$

$\mathbf{Z}_l^{(1)}$ and $\mathbf{Z}_r^{(1)}$ represent the positive sequence impedance from each monitoring point to the head node and the end node of the branch, respectively.

The calculation formula of positive sequence short-circuit current is standard positive sequence voltage before fault divided by positive sequence driving point impedance at fault position:

$$\dot{I}_F^{(1)} = \frac{\dot{V}_M^{\text{std}(1)}}{Z_{FF}^{(1)}}. \quad (6)$$

In formula (6), $Z_{FF}^{(1)}$, the positive sequence driving point impedance at fault position F can be expressed as follows:

$$\begin{aligned} Z_{FF}^{(1)} &= (Z_{ll}^{(1)} + Z_{rr}^{(1)} - 2Z_{lr}^{(1)} - Z_L^{(1)})\lambda^2 \\ &\quad + \{Z_L^{(1)} - 2(Z_{ll}^{(1)} - Z_{rr}^{(1)})\}\lambda + Z_{ll}^{(1)}, \end{aligned} \quad (7)$$

where $Z_{ll}^{(1)}$ and $Z_{rr}^{(1)}$ are the sequence driving point impedances at nodes l and r , respectively, $Z_{lr}^{(1)}$ is the sequence driving point impedances between nodes l and r , and $Z_L^{(1)}$ is the sequence line impedances between nodes l and r .

From the data information uploaded from the monitoring point and the network topology parameters, it can be seen that the fault location λ is the only variable in formula (5)–(7).

In order to avoid the influence of data amplitude, the sequence of positive sequence voltage variation is standardized in this paper. The standardized processing formula is

$$\Delta V^{(1)N} = \frac{\Delta V^{(1)} - E(\Delta V^{(1)})}{\sqrt{D(\Delta V^{(1)})}}. \quad (8)$$

Among them, $\Delta V^{(1)N}$ is the standardized result, $E(\Delta V^{(1)})$ is the expectation of $\Delta V^{(1)}$ array, and $D(\Delta V^{(1)})$ is the variance of $\Delta V^{(1)}$ array.

After standardizing the positive sequence voltage variation sequence of the monitoring points, the fault location model (LM) of the branch with the fault distance λ as the only variable can be obtained:

$$LM_{b,\lambda} = [M_1^{b,\lambda}, M_2^{b,\lambda}, \dots, M_n^{b,\lambda}]. \quad (9)$$

In the above formula, b represents the number of branch, λ represents fault distance, and $LM_{b,\lambda}$ represents positive sequence voltage variation at n monitoring point when a fault occurs at λ of standardized branch b .

3.2. SVM Fault Branch Selection Method. According to the characteristics of positive sequence voltage distribution, the proportional relationship between positive sequence voltage variations at each monitoring point is fixed for faults occurring at the same location. Therefore, a unique $LM_{b,\lambda}$ corresponds to faults occurring at different locations of the branch. The closer the fault location is, the more similar the value of $LM_{b,\lambda}$ is. For faults occurring on the same branch, the corresponding $LM_{b,\lambda}$ has the highest similarity. Therefore, this paper introduces the SVM-based fault branch selection method for the primary selection of the branch where the fault is located, in order to reduce the amount of calculation needed for fault location.

Support vector machine (SVM) makes the linear non-separable samples in the input space project to the high-dimensional space through nonlinear mapping and becomes the linear separable samples by introducing inner product kernel function [16].

Define the category tag y_i of the sample $\mathbf{x}_i$. In the case of two classifications, if $\mathbf{x}_i \in \omega_1$, then $y_i = 1$; if $\mathbf{x}_i \in \omega_2$, then $y_i = -1$. The training sample set is $\{(\mathbf{x}_i, y_i)\}$. From the classification interface $d(\mathbf{x}) = \omega^T \mathbf{x} + b = 0$, we can get

$$y_i(\omega^T \mathbf{x}_i + b) \geq 1, \quad i = 1, 2, \dots, N. \quad (10)$$

The distance from the sampling point to the classification interface is

$$d_x = \frac{|\omega_0^T + \omega_{n+1}|}{\|\omega_0\|}. \quad (11)$$

The optimal classification interface should satisfy the idea of maximum separation, i.e., $\|\omega\|$ minimization:

$$\begin{cases} \min & \frac{1}{2} \omega^T \omega \\ \text{s.t.} & y_i(\omega^T \mathbf{x}_i + b) - 1 \geq 0, \quad i = 1, 2, \dots, N. \end{cases} \quad (12)$$

By using the extreme value method of inequality constraints, we can get

$$L(\omega, b, \alpha) = \frac{1}{2} \omega^T \omega - \sum_{i=1}^N \alpha_i [y_i(\omega^T \mathbf{x}_i + b) - 1]. \quad (13)$$

The partial derivatives of ω and b are obtained, respectively, and the result is 0. The dual principle is used to transform the problem into and under the condition of

$$\sum_{i=1}^N y_i \alpha_i = 0, \quad \alpha_i \geq 0, \quad i = 1, 2, \dots, N, \quad (14)$$

solving the maximum value of the following functions for α_i :

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j). \quad (15)$$

If α_i^* is the optimal solution,

$$\omega^* = \sum_{i=1}^N \alpha_i^* y_i \mathbf{x}_i. \quad (16)$$

The samples that with $\alpha_i^* \neq 0$ are support vector. In conclusion, the optimal classification function is

$$d(x) = \text{sgn} \left(\sum_{i=1}^n \alpha_i^* y_i (\mathbf{x}_i \cdot \mathbf{x}) + b^* \right), \quad (17)$$

where b^* is the classification threshold. The inner product kernel function $K(\mathbf{x}_i, \mathbf{x})$ is used to replace the inner product $(\mathbf{x}_i \cdot \mathbf{x})$ of the training samples and the samples to be classified. At this point, the objective function becomes

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j). \quad (18)$$

The kernel classification function is expressed as

$$d(x) = \text{sgn} \left(\sum_{i=1}^n \alpha_i^* y_i K(\mathbf{x}_i, \mathbf{x}) + b^* \right). \quad (19)$$

For the linear nonseparable case, SVM introduces the relaxation variable ζ_i and penalty factor C , and the objective function becomes

$$\phi(\omega, \zeta) = \frac{1}{2} \|\omega\|^2 + C \left(\sum_{i=1}^N \zeta_i \right). \quad (20)$$

The kernel function used in this paper is sigmoid function:

$$K(\mathbf{x}_i, \mathbf{x}) = \tanh(v(\mathbf{x}_i \mathbf{x}_j) + c). \quad (21)$$

Based on this, this paper regards the failure of each branch as the same type. At 25%, 50%, and 75% of each branch, the virtual fault points are set up, respectively. The fault simulation is carried out at the virtual fault points. The fault data are processed according to formula (5)~(9). A sufficient number of fault samples $LM_{b,\lambda}$ are generated as training samples of the SVM-based fault branch selection algorithm. The trained SVM classifier is used to judge the branch of the network short-circuit fault, so as to narrow the fault location range and lay the foundation for accurate fault location.

3.3. Fault Location Method Based on Similarity Matching. After determining the branch where the fault is located, this paper uses the similarity matching method to solve the fault location parameter λ . Using the Euclidean distance and Pearson correlation coefficient to establish the target optimization model, the enumeration method is used to obtain the value of the fault location parameter λ that makes the objective function optimal, so as to determine the fault occurrence position.

3.3.1. Euclidean Distance. The Euclidean distance transform is useful for a variety of applications including image processing, computer vision, pattern recognition, shape analysis, and computational geometry [17–19]. There are two vectors $X = [x_1, x_2, x_3, \dots, x_n]$ and $Y = [y_1, y_2, y_3, \dots, y_n]$. Their Euclidean distance is

$$d(X, Y) = \sqrt{\sum_{i=1}^n |x_i - y_i|^2}, \quad (22)$$

where x_i is the i -th component of X , y_i is the i -th component of Y , and n is the number of elements. In this paper, n denotes the number of monitoring points in the network. For the two faults at the same position, the Euclidean distance between their $LM_{b,\lambda}$ is extremely small or even zero, and the similarity between them is high. Conversely, when the two faults are far apart in the network. Then, the Euclidean distance is very large, and the similarity is small.

3.3.2. Pearson Correlation Coefficient (PCC). PCC is a parameter used to measure the linear relationship between distance variables [20, 21]. The larger the absolute value of PCC, the stronger the correlation: the closer the PCC is to 1 or -1 , the stronger the correlation. The closer the PCC is to 0, the weaker the correlation. The PCC is positive for positive correlation and negative for negative correlation.

The PCC between X and Y is defined as

$$\rho_{X,Y} = \frac{N \sum XY - \sum X \sum Y}{\sqrt{N \sum X^2 - (\sum X)^2} \cdot \sqrt{N \sum Y^2 - (\sum Y)^2}} \quad (23)$$

where N is the dimension of X , and X and Y have the same meanings as in formula (22).

Combined with the above Euclidean distance and PCC algorithm, the fault distance objective function established in this paper is shown in the following formula:

$$S_{LM_F, LM_{b,\lambda}} = \frac{\rho_{LM_F, LM_{b,\lambda}}}{1 + \ln(1 + D_{LM_F, LM_{b,\lambda}})}, \quad (24)$$

where LM_F is the positioning information corresponding to the current fault, which is calculated by normalization of formula (5). $S_{LM_F, LM_{b,\lambda}}$ is the similarity index between the current fault information and the faulty branch b positioning model $LM_{b,\lambda}$. $\rho_{LM_F, LM_{b,\lambda}}$ is the PCC between LM_F and $LM_{b,\lambda}$, and $D_{LM_F, LM_{b,\lambda}}$ is the Euclidean distance between LM_F and $LM_{b,\lambda}$. $S_{LM_F, LM_{b,\lambda}}$ has a value range of $[-1, 1]$. The higher the linear positive correlation between the two groups, the closer the $S_{LM_F, LM_{b,\lambda}}$ is to 1; the higher the linear negative correlation between the two groups, the closer the $S_{LM_F, LM_{b,\lambda}}$ is to -1 . When they are identical, $S_{LM_F, LM_{b,\lambda}} = 1$; when the similarity between the two sets of information is lower, $S_{LM_F, LM_{b,\lambda}}$ becomes closer to zero.

3.4. Fault Location Method Execution Steps. In the fault location calculation process, the short-circuit fault is simulated at 25%, 50%, and 75% of each branch in the network, and a fault-based branch selection database is established. Use this database to train the SVM fault branch selection model. Then, based on the current fault information, LM_F , the SVM fault branch selection model is used to determine the branch where the fault is located. Based on the results of the faulty branch, the maximum value of the indicator $S_{LM_F, LM_{b,\lambda}}$ is the final goal. Look for the $LM_{b,\lambda}$ that best matches LM_F . Thus, the optimal solution of the fault distance parameter λ is obtained. Determine the location of the fault at the branch. This paper uses the enumeration method to calculate the value of $S_{LM_F, LM_{b,\lambda}}$, the constraint is $\lambda \in [0, 1]$, and the objective function is as shown in formula (11). The execution flow of the fault location algorithm is shown in Figure 3.

4. Case Analysis

The case analysis is based on the IEEE-14 node typical distribution network as a simulation model. The model topology and the parameters of each branch are shown in Figure 4 and annex. The analysis of the cases in this chapter consists of three parts. Firstly, the effectiveness of the SVM fault branch selection algorithm is verified in Section 4.1. And then, in Section 4.2, the short-circuit fault is simulated on a branch to verify the effectiveness of the proposed algorithm. Finally, the fault simulation experiment is carried out on each branch, and the performance of the fault location algorithm proposed in this paper is verified in Section 4.3.

4.1. Case 1: SVM-Based Fault Branch Selection Algorithm. Fault tests were carried out at 25%, 50%, and 75% of each branch. The fault types included single-phase short-circuit fault, two-phase short-circuit fault, and three-phase

FIGURE 3: Fault location execution flowchart.

short-circuit fault. Table 1 shows the data of each monitoring point when single-phase ground fault occurs at 50% of all 14 branches of the network. The gray area in Table 1 is the monitoring point where the positive sequence voltage value of the node is not affected. The fault sample data of the same branch are classified into one class.

In order to know how many training samples should be sufficient to meet the expected accuracy, the relationship between training set size and test set accuracy was tested before the fault location experiment. The experimental results are shown in Table 2. According to Table 2, when the number of training samples reaches 540 or above, the accuracy of the test set can reach 100%. So, in the preparation of fault location experiment, we simulated each type of fault three times at 25%, 50%, and 75% of each branch and generated 540 training samples in total.

Then, in each branch, three fault points (60 in total, random position, and random fault type) are randomly selected as the test data of the SVM fault branch selection. The process of generating validation samples is as follows:

- A random number from 0 to 1 is generated as the fault distance λ to determine the fault location.
- The fault type is selected randomly, in single-phase fault, two-phase fault, and three-phase fault.
- On branch 1, the failure can be simulated according to the result of random selection, and 1 verification sample is obtained.
- Repeat the random fault simulation three times on branch 1, and then continue the random fault simulation on the next branch. There are 20 branches in total, so 60 verification samples can be obtained.

The correct rate of branch selection test results is shown in Table 3. It can be seen from Table 3 that the SVM-based fault branch selection method can accurately determine the fault branch and lay a foundation for subsequent fault location.

4.2. Case 2 : Fault Location Method Based on SVM Fault Branch Selection and Similarity Model Matching. Assuming that single-phase ground fault occurs in branch 10, $\lambda = 0.8$ from

FIGURE 4: The typical distribution network of IEEE-14 nodes.

TABLE 1: The positive sequence voltage variation information database of monitoring points.

Node	Branch													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1.000	0.988	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000
2	-1.000	1.000	0.992	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000
3	-1.000	0.994	-1.000	1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000
4	1.000	-1.000	-1.000	-1.000	0.949	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000
5	-1.000	0.969	-1.000	-1.000	1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000
6	-1.000	-1.000	1.000	1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000
7	-1.000	-1.000	-1.000	0.960	1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000
8	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	1.000	1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000
9	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	0.100	-1.000	1.000	-0.200	-0.600	-0.900	-1.000	-0.300
10	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-0.683	1.000	-0.109	-0.980	-1.000	-0.881
11	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-0.231	1.000	-1.000	-1.000	-1.000
12	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	1.000	-0.421	-0.818
13	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-0.114	1.000	-0.291
14	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-0.650	-0.875	-1.000	-0.825	-0.450	1.000
15	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-0.921	0.762	1.000	-1.000	-1.000	-0.980
16	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	1.000	0.441	-0.548
17	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-0.385	0.692	1.000
18	-1.000	-1.000	-1.000	-1.000	1.000	1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000
19	-1.000	-1.000	-1.000	1.000	-1.000	-1.000	1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000
20	-1.000	-1.000	-1.000	0.105	-1.000	-1.000	-1.000	-1.000	1.000	-0.196	-0.600	-0.910	-1.000	-0.302

node 9 and $1 - \lambda = 0.2$ from node 10. Name the fault F_1 . The $\Delta V_M^{(1)}$ of fault F_1 is obtained from each monitoring. LM_{F_1} is calculated by formula (7). Input the LM_{F_1} into the SVM classifier, and get the result as branch 10. The calculation results are shown in Table 4.

Establishing the fault distance location model of the No. 10 branch, $LM_{10,\lambda}$, calculate the similarity index $S_{LM_{F_1}, LM_{10,\lambda}}$ between LM_{F_1} and $LM_{10,\lambda}$ according to formula (11). Using the enumeration method to optimize the variable λ , with $\lambda \in [0, 1]$ as the constraint. Step size is 0.001. $\max(S_{LM_{F_1}, LM_{10,\lambda}})$

TABLE 2: The relationship between training set size and test set accuracy.

The number of training samples	Test accuracy (%)
100	8.3
200	38.3
300	81.7
400	96.7
500	100
600	100
700	100

TABLE 3: The test results of the SVM-based fault branch selection algorithm.

Branch number	Accuracy rate (%)
1	100
2	100
3	100
4	100
5	100
6	100
7	100
8	100
9	100
10	100
11	100
12	100
13	100
14	100
15	100
16	100
17	100
18	100
19	100
20	100

TABLE 4: Location information of F_1 .

Node number	2	4	5	7	10	14
$\Delta V_M^{(1)}$	0	0	0	0	1.33	0.03
LM_{F_1}	-1	-1	-1	-1	1	-0.955

is the objective function, seeking the optimal value of the fault distance parameter λ . In order to more intuitively observe the change of $S_{LM_{F_1}, LM_{10}, \lambda}$ with λ . As shown in Figure 5, the similarity changes with the fault distance parameter λ . It can be seen from the figure that the $S_{LM_{F_1}, LM_{10}, \lambda}$ obtains the maximum value 0.981583 when the fault distance $\lambda = 0.788$. The location result is consistent with the position of F_1 , which verifies the effectiveness of the proposed algorithm.

4.3. Case 3: Performance Verification of Fault Locating Algorithm. In each branch, 10 fault points (200 in total, including single-phase short-circuit fault, two-phase short-circuit fault, and three-phase short-circuit fault) are randomly selected for fault location this verification. Establish fault location models for 20 branches. Use the location error

FIGURE 5: Similarity with different fault distances.

TABLE 5: The results of the average location error rate of each branch.

Branch number	The average error rate (%)
1	3.26
2	3.62
3	2.51
4	3.65
5	2.53
6	1.39
7	1.11
8	1.50
9	0.95
10	0.25
11	3.50
12	3.15
13	2.95
14	0.00
15	0.10
16	0.25
17	0.11
18	2.19
19	3.17
20	3.84

rate to describe the performance of the algorithm. As shown in formula (25). In formula (25), the line length is 1. The smaller the error rate, the more accurate the location method is. The results of the average location error rate of each branch are shown in Table 5:

$$e = \frac{|\text{actual fault distance } \lambda - \text{simulated fault distance } \lambda_F|}{\text{line length}} \times 100\%. \quad (25)$$

The average location error rate shown by the experimental results can be well controlled within 4%, which satisfies the requirements for the accuracy of fault location results.

5. Conclusion

The fault location algorithm proposed in this paper only deals with the positive sequence voltage variation of each monitoring point when the fault occurs. The location result is not affected by the fault source type and transition resistance. Firstly, the branch of the fault is determined by the SVM fault branch selection method, which avoids the traversal of the whole network branch, reduces the location range, and reduces the calculation amount. Then, the fault location parameter λ is solved by using the fault location similarity model combining Euclidean distance and PCC.

This paper builds an IEEE-14 node network to verify the effectiveness of the algorithm. The results of the case show that the method can locate various types of faults, and the average location error of each branch can be well controlled within 4%. The algorithm lays a solid theoretical foundation for the rapid processing of grid faults.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to thank the project “Research on Key Technologies of Compact Flexible Loop Closing Device for Medium Voltage Distribution Network” (J2020081) for their support in this research.

References

- [1] M. M. Saha, J. Izykowski, and E. Rosolowski, “Fault location on power networks,” *Power Systems*, Springer, Berlin, Germany, 2010.
- [2] B. Wang, J. Geng, and X. Dong, “High-impedance fault detection based on nonlinear voltage-current characteristic profile identification,” *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 3783–3791, 2016.
- [3] R. Z. Liang, N. Peng, C. Liu, and F. Zare, “Asynchronous fault location in transmission lines considering accurate variation of the ground-mode traveling wave velocity,” *Energies*, vol. 10, no. 12, p. 1957, 2017.
- [4] D. Akmaz, M. S. Mamiş, M. Arkan, and M. E. Tağluk, “Transmission line fault location using traveling wave frequencies and extreme learning machine,” *Electric Power Systems Research*, vol. 155, pp. 1–7, 2018.
- [5] A. Ahmadimanesh and S. M. Shahrtash, “Transient-based fault-Location method for multiterminal lines employing S-transform,” *IEEE Transactions on Power Delivery*, vol. 28, no. 3, pp. 1373–1380, 2013.
- [6] F. Mei, Y. Ren, Q. Wu et al., “Online recognition method for voltage sags based on a deep belief network,” *Energies*, vol. 12, p. 43, 2019.
- [7] H. Sha, F. Mei, C. Zhang, Y. Pan, and J. Zheng, “Identification method for voltage sags based on K-means-Singular value decomposition and least squares support vector machine,” *Energies*, vol. 12, no. 6, p. 1137, 2019.
- [8] M. Korkali, H. Lev-Ari, and A. Abur, “Traveling-wave-based fault-Location technique for transmission grids via wide-area synchronized voltage measurements,” *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 1003–1011, 2012.
- [9] F. V. Lopes, K. M. Silva, F. B. Costa, W. L. A. Neves, and D. Fernandes, “Real-time traveling-wave-based fault location using two-terminal unsynchronized data,” *IEEE Transactions on Power Delivery*, vol. 30, no. 3, pp. 1067–1076, 2015.
- [10] H. Livani and C. Y. Evrenosoglu, “A machine learning and wavelet-based fault location method for hybrid transmission lines,” *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 51–59, 2014.
- [11] K. Jia, Z. Ren, L. Li, Z. Xuan, and D. Thomas, “High-frequency transient comparison based fault location in distribution systems with DGs,” *IET Generation, Transmission & Distribution*, vol. 11, no. 16, pp. 4068–4077, 2017.
- [12] S. Lotfifard, M. Kezunovic, and M. J. Mousavi, “Voltage sag data utilization for distribution fault location,” *IEEE Transactions on Power Delivery*, vol. 26, no. 2, pp. 1239–1246, 2011.
- [13] H. Mokhlis and H. Li, “Non-linear representation of voltage sag profiles for fault location in distribution networks,” *International Journal of Electrical Power & Energy Systems*, vol. 33, no. 1, pp. 124–130, 2011.
- [14] Y. Liao, “Fault location for single-circuit line based on bus-impedance matrix utilizing voltage measurements,” *IEEE Transactions on Power Delivery*, vol. 23, no. 2, pp. 609–617, 2008.
- [15] N. C. Woolley, M. Avendaño-Mora, A. P. Woolley, R. Preece, and J. V. Milanović, “Probabilistic estimation of voltage sags using erroneous measurement information,” *Electric Power Systems Research*, vol. 106, pp. 142–150, 2014.
- [16] T. V. Gestel, J. A. K. Suykens, B. Baesens et al., “Benchmarking least squares support vector machine classifiers,” *Machine Learning*, vol. 54, no. 1, pp. 5–32, 2004.
- [17] C. Arcelli, G. S. di Baja, and L. Serino, “Distance-driven skeletonization in voxel images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 4, pp. 709–720, 2011.
- [18] S. Loncaric, “A survey of shape analysis techniques,” *Pattern Recognition*, vol. 31, no. 8, pp. 983–1001, 1998.
- [19] L. Lam, S.-W. Lee, and C. Y. Suen, “Thinning methodologies-a comprehensive survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 9, pp. 869–885, 1992.
- [20] A. Ly, M. Marsman, and E. J. Wagenmakers, “Analytic posteriors for pearson’s correlation coefficient,” *Statistica Neerlandica*, vol. 72, no. 1, pp. 4–13, 2015.
- [21] H. Xiong, S. Shekhar, P. Tan, and V. Kumar, “Exploiting a support-based upper bound of Pearson’s correlation coefficient for efficiently identifying strongly correlated pairs,” in *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, Seattle, WA, USA, August 2004.