Hindawi

Research Article

Terminal Waveform Similarity Measurement Method Based on the Improved Dynamic Time Warping Algorithm

Xiong Yang,[1] Jiahao Guo,[1] Xuhui Zhang,[2] and Chenyang Zhu [2]

[1]*Distribution Network Technology Department, State Grid Jiangsu Electric Power Co., Ltd., Research Institute (JSEPRI), Nanjing 211103, China*
[2]*School of Mechanical Engineering, Nanjing Institute of Technology, No. 1 Hongjing Road, Nanjing 211167, China*

Correspondence should be addressed to Chenyang Zhu; zzccyy120323@163.com

The dynamic time warping algorithm (DTW) has problems such as high computational complexity and "ill-conditioned matching." Aiming at the above two main problems, this paper proposes an improved DTW algorithm for the final wave recording of the primary and secondary deep fusion equipment detection platform. The terminal recorded waveform and the waveform with non-Gaussian noise added as the research object, the two sets of waveforms are divided into frames and windowed, and the short-term energy entropy ratio of the two sets of waveforms is input into the DTW as the test vector. Using the optimal matching paths and distances of the two input vectors, the common substring lengths of the two sets of short-term energy entropy ratio sequences are calculated. Then, we define the optimal matching coefficient and correct the waveform similarity. The experimental data show that the improved DTW algorithm can accurately quantify the similarity between terminal waveforms, which can provide effective data support for the health status assessment of power distribution terminals.

1. Introduction

With the acceleration of distribution network construction, the coverage rate of distribution terminals is also increasing year by year. The power distribution terminal is an important sensing unit, and its health status is evaluated. The staff can discover the potential safety hazards of the terminal in time so as to arrange a reasonable maintenance plan to ensure the safe operation of the power grid [1]. The sampling waveform can reflect the problems existing in the test system during the actual recording process. In practical applications, the system will be disturbed by the electromagnetic and noise signals of primary switch opening and closing. By analyzing the waveform recorded by the terminal and the standard waveform of the power source, the health status can be evaluated [2–4]. Therefore, analyzing the waveform characteristics is a key step in maintaining the normal operation of the system.

Similarity measurement theory is used to compare the similarity relationship between certain forms, images,

textual information, or various data. Based on the wide application of similarity algorithms in various fields, it has surpassed dozens of classical similarity calculations [5–10]. Waveform similarity measure inherits and develops the similarity measure theory. Its application fields are very wide, such as speech recognition, radar detection and recognition, electrocardiogram intelligent detection, traditional Chinese medicine fingerprint identification, and other fields, and the waveform similarity algorithm has been well applied and developed [11, 12].

Dynamic time warping (DTW) is a classic optimization algorithm. We use a time warping function that meets specific requirements to define the relative time-to-moment correlation between the test module and the reference module. Thus, the minimum cumulative distance when the two modules are matched is calculated [13]. When the two sequences as a whole have very similar shapes, when the sequences are not aligned on the x-axis, before comparing the similarity, warping one (or both) of the sequences under the time axis helps to achieve a better align effect. The DTW

algorithm is an effective method to achieve warping, which calculates the similarity between two sequences by lengthening and shortening the time series to obtain the shortest distance between them [14, 15]. For a long time, scientific researchers and technicians in various fields have carried out a lot of research on the DTW algorithm. Ran et al. [16] introduced a weighting function to improve the DTW algorithm to suppress the edge effect, but this method is only optimized for the head and end, but it is not suitable for the global alignment of the end waveform. Li et al. [17] proposed a piecewise linear fitting dynamic time warping similarity measure, using the DTW algorithm to measure the time series after multidimensional piecewise fitting, which has a relatively large data scale and continuous variable sequence. It has a good effect, but the experimental results are greatly affected by the choice of parameters. In terms of computational efficiency and to improve the communication rate, it utilizes the event-triggered schedule to relieve the communication burden, which is between sensors and the remote filter [18]. In article [19], it develops an event-triggered strategy with a sampling scheme and an arrangement to reduce data transmission significantly, in which it sends measurements when it meets the event-triggered conditions. It reduces the transmission cost of data to reduce the communication burden and also reduces the burden of sampling units to improve the computational efficiency of the algorithm.

The waveforms detected in this paper are generated by the primary and secondary fusion equipment of State Grid Corporation of China. At present, the primary and secondary equipment of the distribution network are still independent of each other, but there have been distribution switchgears with some secondary intelligent units in the primary part. In order to continuously improve the lean level and operational efficiency of line loss management, State Grid Corporation of China proposed a technical solution for the integration of primary and secondary equipment for power distribution. The deep integration of primary body equipment, high-precision sensors, and secondary terminal equipment can achieve the goals of "high reliability, miniaturization, platformization, versatility, and economy." The primary and secondary deep integration equipment can meet the requirements of interface standardization and complete set of bidding and procurement. In other words, the device can meet the requirements of high integration, interchange, maintenance, etc. Solve the problems of insulation coordination, electromagnetic compatibility, and life matching of complete sets of equipment. In the future, the intelligent equipment composed of primary and secondary equipment will not only have the main equipment body for transmitting and distributing electric energy but also have functions such as measurement, control, protection, and metering. The physical form of each function is embodied in the form of intelligent components, and the traditional division of primary and secondary equipment is no longer emphasized. In the aspect of signal measurement, based on the AC power flow model, the theory of admittance weighted topology, and complex network centrality, it

proposes a sampling key node identification approach [20]. The primary and secondary equipment of the distribution network are involved in this paper, which takes the sampling key node as the main point and selects 14 cycles in total, which contains the first five and the last nine cycles in the sampling period.

tIn order to solve the problem of the accuracy of the sampling waveform at the terminal of the primary and secondary fusion equipment, this paper proposes an improved DTW algorithm combining common substring and energy entropy ratio. Firstly, a method for preprocessing waveforms is proposed to solve the problems of high computational complexity and insufficient characteristic events. The method divides the terminal waveform and the waveform with added non-Gaussian noise into frames and calculates the short-term energy entropy ratio of the two sets of waveforms as input. Then, aiming at the problem of "ill-conditioned matching" in the waveform comparison, a method to correct the similarity is proposed, which is achieved by calculating the penalty coefficient based on the length of the longest common substring. Finally, the improved DTW algorithm is verified and applied, and Matlab is used to compare the waveforms recorded by the terminal. By comparing the terminal recording waveform with the power source signal waveform, this algorithm can provide data to evaluate the health status of the terminal. It can improve the efficiency and address the "ill-conditioned matching" issue, which is too complicated to adjust through the local adjustment of the waveform. The experiments show that the algorithm can effectively improve the calculation efficiency and accuracy of waveform similarity.

2. Principle Analysis of an Improved DTW Algorithm

2.1. Traditional DTW Algorithm. The two waveform sequences that need to be judged for similarity are set as X and Y, respectively, and the lengths are $|X|$ and $|Y|$. The shape of the normalized path is $W = w_1, w_2, \ldots, w_k$, where $\max(|X|, |Y|) \leq K \leq |X| + |Y|$. The shape of w_k is (i, j), where i represents the i coordinate system in X, and j represents the j coordinate system in Y. The normalization path starts from $w_1 = (1, 1)$ and ends at $w_k = (|X|, |Y|)$, ensuring that W can contain all positions in X and Y. In addition, the i and j values of $w(i, j)$ in the default W are monotonically increasing to ensure that their curves do not cross. Monotonically increasing means that

$$w_k = (i, j), w_{k+1} = (i', j') i < = i' < = i + 1 j < = j' < = j + 1. \quad (1)$$

The path with the least cost of regularization is

$$DTW = \left\{ \sqrt{\sum_{k=1}^{K} \frac{w_k}{K}} \right\}. \quad (2)$$

The goal regularization path takes the one with the shortest relative distance as

$$D(i, j) = \text{Dist}(i, j)$$
$$+ \min[D(i-1, j), D(i, j-1), D(i-1, j-1)].$$
(3)

The process of solving the twisted curve is to find a path from the lower left corner to the upper right corner, which minimizes the sum of the element values traversed by the path [21]. In the process of calculating the DTW distance, the sequence points need to be self-replicated before alignment and matching. This approach can measure nonequal-length series, which is beneficial to support the bending and extension of time series [22–24].

The planning of the path satisfies the constraints such as from the previous square $(i-1, j-1)$ or $(i-1, j)$ or $(i, j-1)$ to the next square (i, j). If you start planning from $(i-1, j)$ or $(i, j-1)$, the distance is $d(i, j)$. If you start planning from $(i-1, j-1)$, the distance is $2d(i, j)$.

$$g(i, j) = \min \left\{ \begin{array}{c} g(i-1, j) + d(i, j) \\ g(i-1, j-1) + 2\, d(i, j) \\ g(i, j-1) + d(i, j) \end{array} \right\}.$$
(4)

In the formula (4), $g(i, j)$ means that both templates are matched successively from the initial component, plan the path to the i component in X and the j component in Y, and calculate the distance between the two templates. Add $d(i, j)$ or $2d(i, j)$ to the result of the previous match and take the minimum value Dist_m.

2.2. Short-Time Energy Entropy Ratio.

The short-term analysis method is widely used in the field of speech signal processing. The waveform signal and the speech signal have the same characteristics as nonlinearity. When performing short-term analysis, the analyzed signal needs to be divided into frames, and several feature parameter time series composed of "frames" are obtained. Then, follow-up analysis and processing is performed on the characteristic parameter time series, which greatly reduces the computational complexity of the computer.

The short-term energy entropy ratio is a time-domain analysis algorithm, which can improve the calculation efficiency by framing and windowing. Not only that, the algorithm can also significantly improve the signal-to-noise ratio, which is beneficial to distinguish abrupt events in the waveform signal. The DTW algorithm has the problems of high computational complexity and low computational efficiency. It is necessary to preprocess the waveform signal through the short-term energy entropy ratio so as to reduce the complexity of the DTW algorithm and improve the calculation efficiency of the DTW algorithm.

Its calculation process is as follows:

(1) Let the waveform sequence of length N be $x(n)$, $n = 1, 2, \ldots, N$. Divide the DC component and normalize the amplitude of the waveform sequence.

(2) Select the appropriate frequency and frame shift to frame and window the processed waveform sequence. Common windowing functions are as follows:

Rectangular window:

$$\omega(n) = 1.$$
(5)

Haining window:

$$\omega(n) = 0.5 - \left(1 - \cos\left(\frac{2\pi n}{(L-1)}\right)\right).$$
(6)

Hamming window:

$$\omega(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{(L-1)}\right),$$
(7)

where L is the length of the window, and $0 \leq n \leq L - 1$.

After the windowing function $\omega(n)$ is divided into frames, the vibration signal of the ith frame is obtained as $y_i(n)$, and then $y_i(n)$ satisfies

$$y_i(n) = \omega(n)x((i-1)i_{nc} + n$$
$$1 \leq n \leq w_{len}, 1 \leq i \leq f_n.$$
(8)

In the formula, $y_i(n)$ is the value of one frame; w_{len} is the frame length; i_{nc} is the frame shift length; and f_n is the total number of frames after the signal is divided into frames.

(3) After Fourier transform is performed on $y_i(n)$, the energy spectrum of the frequency component f_k of the kth spectral line is $Y_i(k)$. Then the normalized spectral probability density function $p_i(k)$ of each frequency component is defined as

$$p_i(k) = \frac{Y_i(k)}{\sum_{i=0}^{N/2} Y_i(k)}.$$
(9)

In the formula, $p_i(k)$ is the probability density corresponding to the kth frequency component f_k of the ith frame.

(4) The spectral entropy and energy of frame i are H_i and E_i, respectively, given as

$$H_i(k) = \sum_{i=0}^{N/2} p_i(k) \ln p_i(k),$$

$$E_i(k) = \lg\left(\frac{1 + \sum_{n=1}^{N} x_i(n)}{a}\right).$$
(10)

In the formula, a is a constant. It can be adjusted according to different waveform energy change degrees, which is conducive to analyzing the mutation points in the waveform signal.

(5) Calculate the value of the energy-entropy ratio to complete the preprocessing of the signal:

$$R_i = \sqrt{1 + \left|\frac{E_i}{H_i}\right|}.$$
(11)

The short-term energy entropy ratios of the two groups of waveforms are input into the DTW in turn, and the corresponding dynamic paths and similarity are obtained.

2.3. Similarity Correction Method.

In the process of data similarity matching operation based on the DTW algorithm, the waveform has the phenomenon of time series fluctuation and amplitude variation, and the curve direction and the monotonic transition speed are different. When calculating the distance of the time curve, it is easy to produce the phenomenon of "ill-conditioned matching" in which the peaks and valleys are not in the same direction. In the generalization process of step mode operation, the path direction can only maintain a short-term monotonic characteristic, but the adjustment process of ill-conditioned matching is complicated. Considering that the voltage amplitude during the recording process may have a large sudden change and the terminal is disturbed by noise signals, this paper optimizes the algorithm in a targeted manner. Assuming a moderate shift in the waveform phase, it is possible to achieve regional trend similarity, thereby eliminating the negative impact of "ill-conditioned matching" on similarity. The length of the longest common substring of the two waveform sequences has a greater impact on the similarity; that is, the longer the longest common substring of the two waveform sequences, the smaller the error and the smaller the required adjustment range. In order to facilitate the adjustment of ill-conditioned matching, a coefficient of optimal matching is defined as a penalty coefficient. Use the penalty coefficient to adjust the spacing of the waveform and multiply the penalty coefficient by the distance to obtain a new distance.

2.4. The Specific Method

(1) Calculate the maximum standard deviation sd_{max}. Let x be the average of all digital sequences X, and n be the number of digital sequences X, then calculate the value of the standard deviation sd as

$$sd_X = \sqrt{\frac{\sum_{i-1}^{n}(x_i - x)^2}{n}}, \tag{12}$$

$$sd_Y = \sqrt{\frac{\sum_{i-1}^{n}(y_i - y)^2}{n}}. \tag{13}$$

The maximum standard deviation is the one with the larger standard deviation of the data in the two time series:

$$sd_{max} = \max(sd_X, sd_Y). \tag{14}$$

(2) The longest common substring to be solved and its length l: Since X and Y are both numerical sequences, when finding the longest common substring, the maximum standard deviation can be set as offset tolerance; that is, both numbers are within this

maximum standard deviation and can be considered to be part of the male substring.

Let the length of the sequence X be a, and let the length of the sequence Y be b. Define matrix $dq[i][j]\,(0 \le i < a),\,(0 \le j < b)$ as

$$dq[i][j] = \begin{cases} 0 & X[i] - Y[i] \ge sd_{max}, \\ 1 & X[i] - Y[i] \ge sd_{max}, \\ dq[i-1][j-1] + 1 & X[i] - Y[j] < sd_{max} i, j > 0. \end{cases} \tag{15}$$

Through the principle of dynamic programming, the optimal path from the lower left corner to the upper right corner in the matrix can be obtained. Find the longest common substring and its length l.

(3) Define the penalty coefficient a as

$$\alpha = 1 - \frac{l}{\min(\mathrm{len}(X), \mathrm{len}(Y))}. \tag{16}$$

(4) According to formulas (17) and (18), the distance algorithm is optimized as

$$\mathrm{Dist}_n = \alpha \times \mathrm{Dist}_m, \tag{17}$$

$$\omega_n = \frac{1}{1 + \mathrm{Dist}_n}. \tag{18}$$

Under the improved DTW algorithm, waveform matching can be regionalized, reducing the similarity distance of common parts. The minimum distance is adjusted by the proportion of common substrings in the overall sequence, which effectively solves the impact of "ill-conditioned matching" on the overall accuracy. Thus, the accuracy of waveform similarity calculation is improved.

2.5. Algorithm Flow.

To sum up, this paper proposes an optimization method of DTW combining energy entropy ratio and common substring. The algorithm flow is shown in Figure 1. The specific steps are as follows:

(1) Select the appropriate window function and frame length to calculate the energy-entropy ratio sequence of the two sets of waveforms, and complete the preprocessing of the waveform sequence

(2) Input the two sets of energy entropy ratio sequences into DTW, construct a matrix, and obtain the distance and similarity of the two sets of waveforms

(3) Calculate the common substring of the energy-entropy ratio sequence, obtain the optimal matching coefficient, complete the correction of the waveform similarity, and output the result

3. Simulations Analysis

3.1. Modeling of Non-Gaussian Noises.

There are many sources of noise. In practical application, the rain-induced corona, switching pulse and lightning,
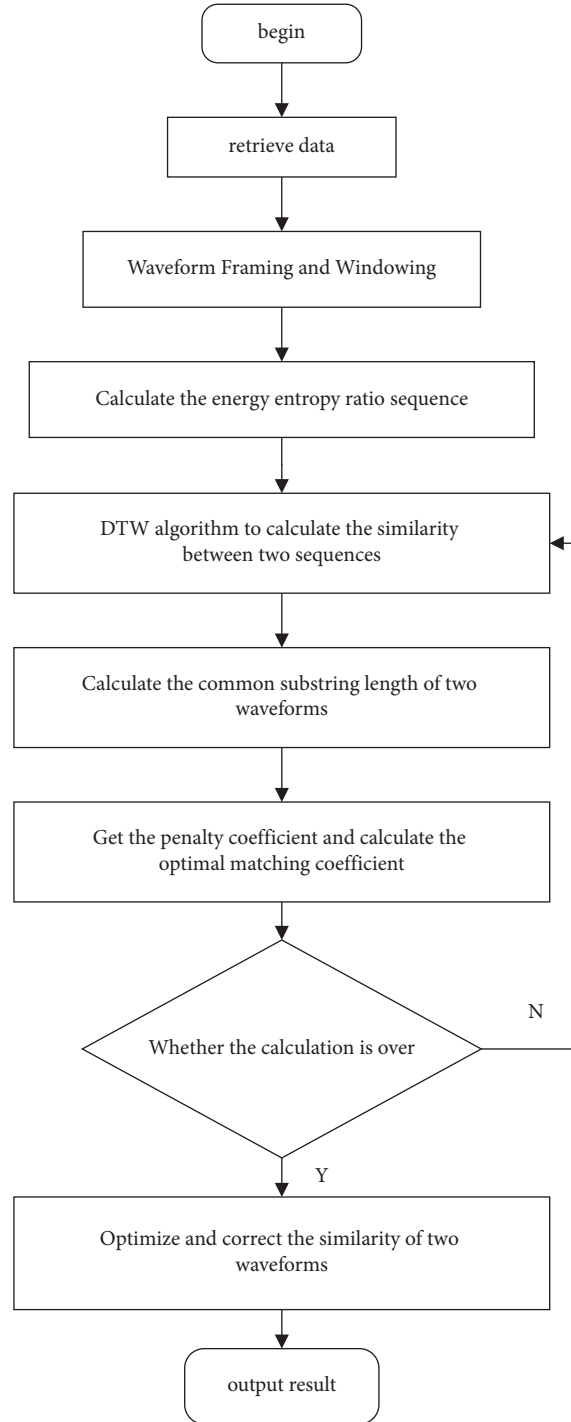
FIGURE 1: Algorithm flowchart.

communication channel noise, ambient temperature variation, and other reasons may make the process noise and measurement noise to follow non-Gaussian or even unknown distributions [19]. For non-Gaussian noise, it is modeled according to the widely used ε-contaminated model:

$$P = (1 - \varepsilon)N(0, k1) + \varepsilon N(0, k2), \qquad (19)$$

where P is the probability distribution of the noise, $\varepsilon \in (0, 1)$ denotes the pollution degree, and $N(\mu, \kappa)$ represents the Gaussian distribution of mean μ and covariance.

The modeling simulation of non-Gaussian noise under Matlab is shown in Figure 2, where the black curve represents the probability distribution of noise in the case of white noise, and the red curve is the probability distribution of non-Gaussian noise.
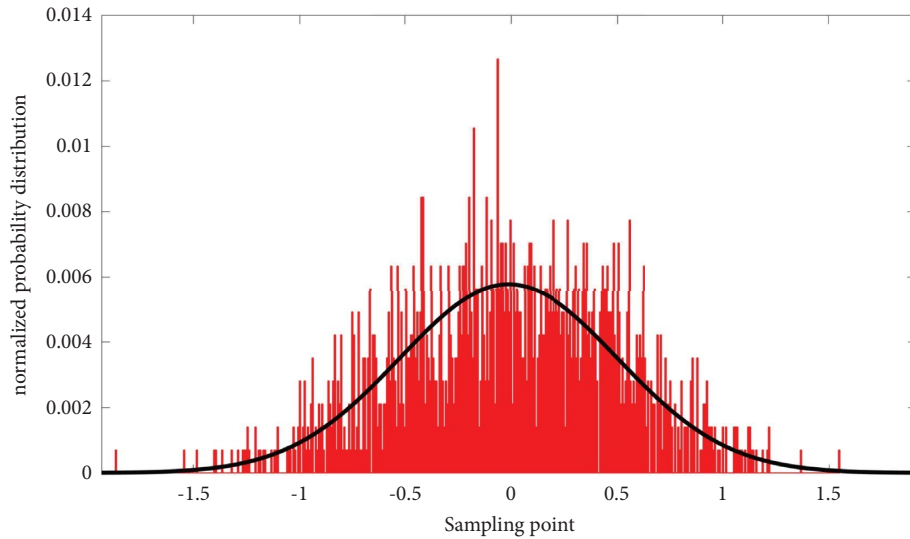
FIGURE 2: Non-Gaussian noise probability distribution.

3.2. Waveform Comparison Based on the Traditional DTW Algorithm. The waveform was fetched from the actual tested terminal in this paper. Figure 3(a) shows the waveform collected by the terminal in the healthy state, and Figure 3(b) shows the waveform of the terminal after adding non-Gaussian noise signal. The purpose of adding non-Gaussian noise is to simulate the background noise of the system. Due to the background noise and other related problems of electromagnetic interference, the collected data contain noise or outliers. Background noise can affect the similarity between waveforms. A total of 14 cycles from the first five cycles and the last nine cycles were selected as sampling periods, and the feasibility of the algorithm was verified by comparing the waveforms under different noise intensity.

Figure 4(a) shows the point matching diagram of the terminal waveform in the ideal state and the two original waveform sequences after adding the non-Gaussian noise signal. Figure 4(b) is the dynamic path planning diagram of the original waveform sequence of the two waveforms. In the process of dynamic path planning, abrupt events of time cannot be highlighted.

3.3. Waveform Comparison Based on the Improved DTW Algorithm. For the problems of low matching efficiency and high computational complexity, the short-term energy entropy ratio can effectively extract the features contained in the original signal. The method can remove the characteristics of redundant components in the original signal, reduce the complexity of input features, and improve computational efficiency.

The frame length is the determinant of the short-term energy entropy ratio resolution. When the signal sampling frequency is constant, the smaller the frame length is, the higher the temporal resolution of the short-term energy entropy ratio is, but too short frame length is not conducive to exerting the advantage of short-term energy entropy ratio to improve the signal-to-noise ratio. Therefore, the selection

of the frame length should take into account the sampling frequency of the signal, the time resolution requirements, and the signal-to-noise ratio requirements to select an appropriate value. By observing the collected waveform signal, it can be found that the vibration signal within 30 ms is relatively stable. When the frame length is small, the smoothing effect of the short-term analysis processing method is not obvious. When the frame length is larger, its smoothing effect is better, which is beneficial to the DTW algorithm for path regulation. Setting the frame length to 25 ms preserves characteristic events well. Set the frame length $w_{len} = 25$ ms and frame shift $i_{nc} = 20$. Figure 5(a) shows the energy-entropy ratio curve of the terminal wave recording waveform in the ideal state, and Figure 5(b) shows the energy-entropy ratio curve of the waveform disturbed by noise. By processing the short-time energy entropy ratio of the two sets of waveform sequences, the short-time energy entropy ratio characteristic sequence is obtained.

Figure 6(a) is the point matching diagram of the terminal waveform under the ideal state and the terminal waveform after adding the non-Gaussian noise signal after short-term energy entropy ratio processing under the DTW algorithm. Combined with the analysis in Figure 4(a), the terminal wave recording will produce an instantaneous mutation, the mutation point is around the seventh cycle of sampling, and the corresponding sampling point is around 100. The difference between the original waveform and the middle position of the stretched waveform is obvious, indicating that there are shock events and sudden changes in the waveform. Using the short-term energy entropy ratio to process the signal can effectively extract the characteristics of the events with small shocks contained in the waveform signal and can better represent the sudden change of the waveform.

Figure 6(b) is a dynamic path planning diagram of the energy-entropy ratio sequence of the two waveforms. In the process of dynamic path planning, there is a situation of ill-
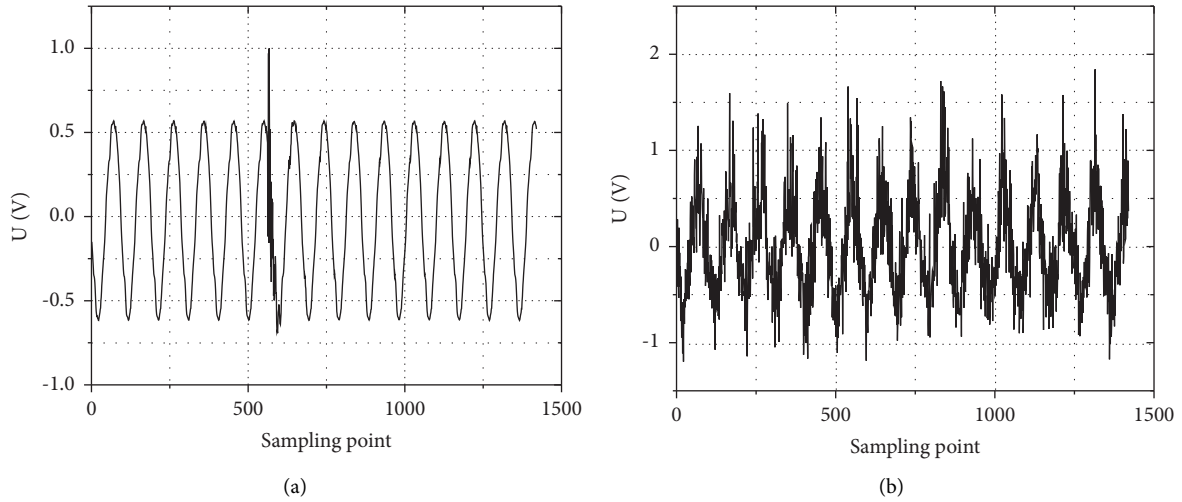
(a)

(b)

FIGURE 3: Terminal output waveform. (a) Voltage waveform of terminal sampling. (b) Voltage waveform with noise add.
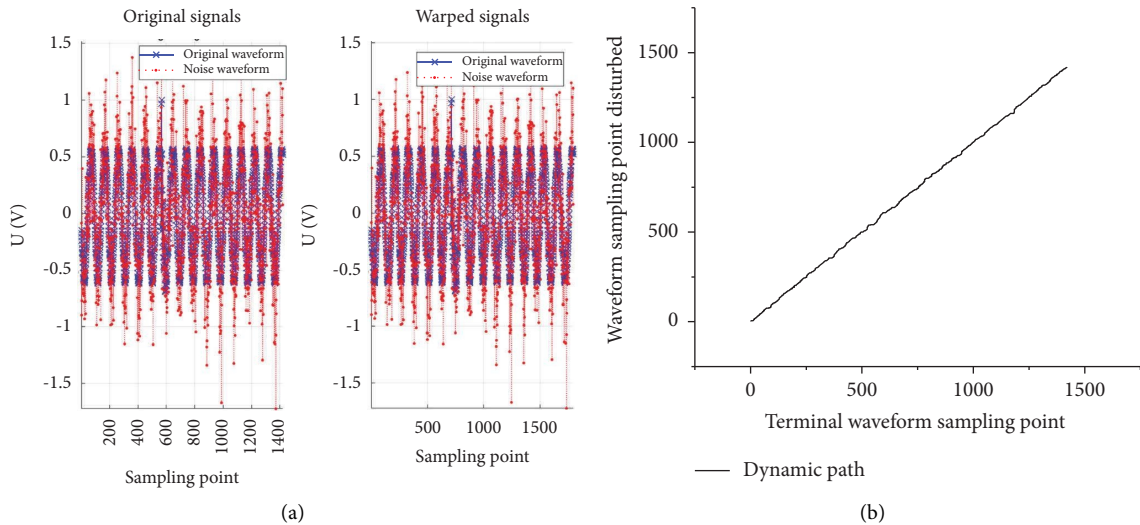

(a)

(b)

FIGURE 4: Initial waveform sequence analysis diagram. (a) Point matching diagram of the initial waveform sequence. (b) Path planning of the initial waveform sequence.

conditioned matching. When the sampled point of the disturbed waveform is 100, there is a situation where one point matches with multiple consecutive points. The same problem exists when the sampling point of the terminal is 2, and the sampling point of the disturbed waveform is 173. Therefore, the similarity is modified in combination with the penalty coefficient based on the common substring.

The data of parameters and results in the operation process of the improved DTW algorithm are shown in Table 1:

$Dist_m$ and ω_m are the distance and similarity preprocessed by the energy entropy ratio. $Dist_n$ and ω_n are the distance and similarity of the waveform under the DTW algorithm based on the common substring after preprocessing by the energy entropy ratio. By analyzing the distance and similarity, the DTW algorithm combining energy entropy ratio and common substring can effectively improve the accuracy of waveform similarity calculation. As

the variance of the noise signal increases, its similarity decreases step by step, but when the variance is 0.2 and 0.3, the similarity before optimization is 0.0796 and 0.0809. The "ill-conditioned matching" situation of the waveform sequence is corrected by means of a common substring, and the corrected similarity is 0.1391 and 0.1359, which are in line with expectations.

3.4. Algorithm Performance Analysis.

In the process of path planning, there is an "ill-conditioned matching" phenomenon in point matching. The common substring can make the waveform phase shift moderately so that the characteristics of the waveform's area toward similarity can be corrected. The ratio of "ill-conditioned matching" is used as the criterion for evaluating the quality of the algorithm, and the accuracy of the algorithm is verified by a numerical value as
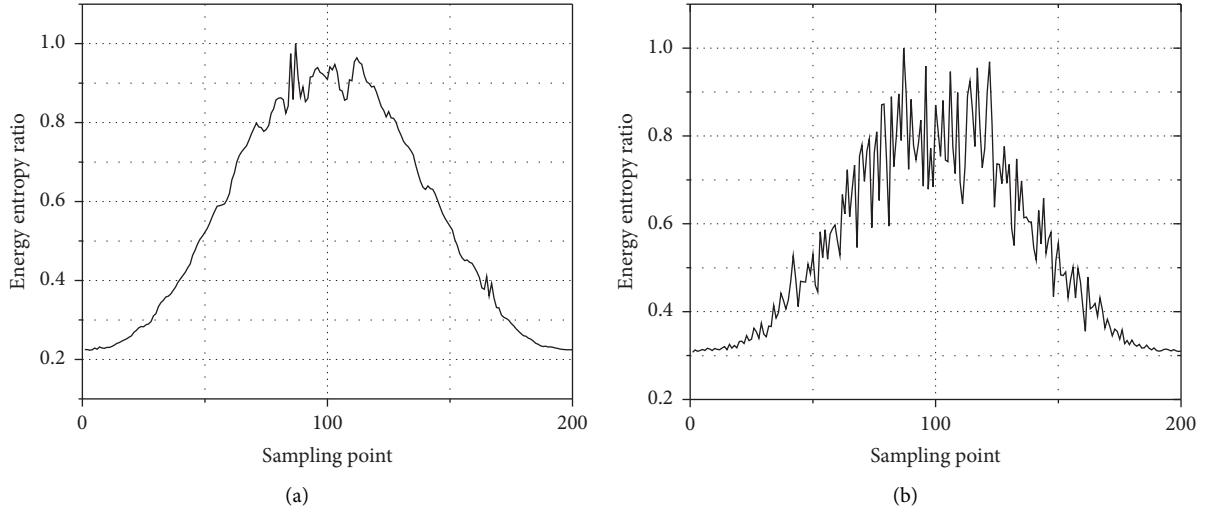
(a)

(b)

FIGURE 5: Schematic diagram of energy entropy ratio. (a) Terminal waveform energy entropy ratio. (b) Disturbed waveform energy entropy ratio.
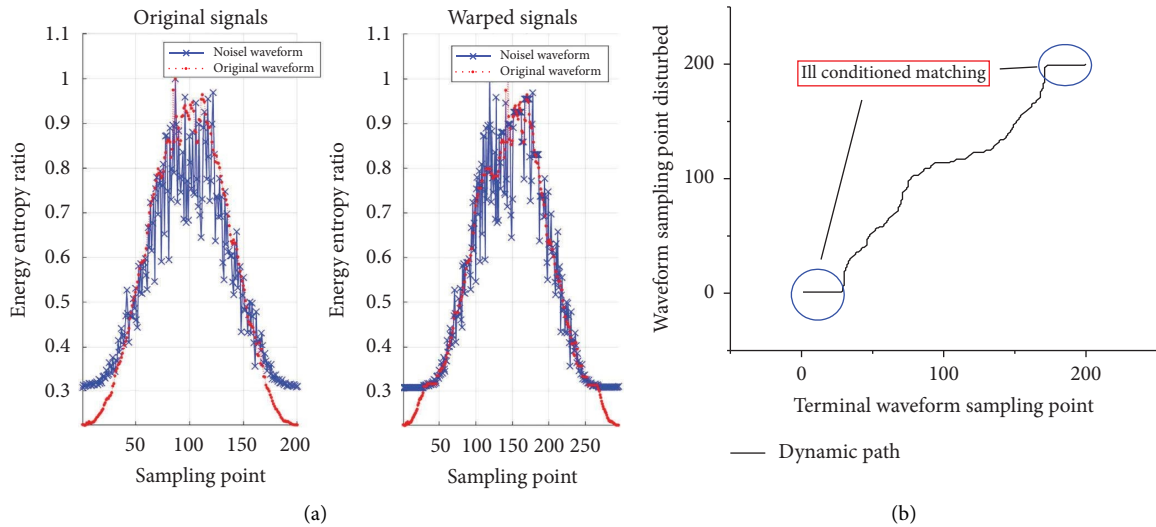

(a)

(b)

FIGURE 6: Energy entropy ratio sequence analysis diagram. (a) Point matching graph of energy-entropy ratio sequences. (b) Path planning of energy-entropy ratio sequences.

$$\text{rate} = \frac{\ln c - (ff - xx)}{\ln c} \times 100\%. \tag{20}$$

In the formula (20), l_{nc} is the total length of matching points, ff is the original "ill-matched" point length, and xx is the length of the "ill-conditioned matching" point corrected under the common substring algorithm.

The three different algorithms are compared experimentally in different non-Gaussian noise environments, and the accuracy results are shown in Table 2. These three algorithms are the traditional DTW algorithm, the DTW based on the energy entropy ratio, and the improved DTW algorithm.

By comparing the accuracy of different algorithms, the accuracy of the traditional DTW algorithm is about 80%.

The accuracy of the DTW algorithm after processing the energy entropy ratio is only about 70%. The accuracy of the DTW algorithm combining the energy entropy ratio and the common substring can reach about 95%. It shows that the energy entropy ratio and common substring proposed in this paper can have higher accuracy, and the DTW algorithm combined with the two algorithms can effectively solve the problem of "ill-conditioned matching."

4. Experiments

4.1. Experiment Platform. This paper takes the primary and secondary deep fusion equipment detection platform built by the Electric Power Research Institute of State Grid Jiangsu Electric Power Co., Ltd. as the research object. Figure 7

TABLE 1: Waveform similarity of DTW algorithm combining common substring and energy entropy ratio.

Noise variance	$Dist_m$	sd_{max}	l	α	$Dist_n$	ω_m	ω_n
0.1	9.5484	0.1692	86	0.570	5.4426	0.0948	0.1552
0.2	**11.5672**	0.1759	93	0.535	**6.1885**	0.0796	0.1391
0.3	**11.3567**	0.1404	88	0.560	**6.3598**	0.0809	0.1359
0.5	15.5511	0.1673	86	0.570	8.8641	0.0604	0.1014

TABLE 2: Comparing the accuracy of different algorithms.

Noise variance	Traditional DTW	DTW based on energy entropy ratio	Improved DTW
0.1	80.9	70.2	95.4
0.2	77.5	72.7	98.2
0.3	80.6	71.2	97.2
0.5	80.0	70.0	95.8
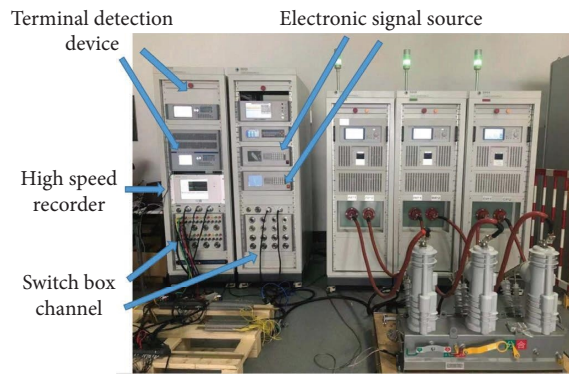
FIGURE 7: Field application of complete sets of automatic detection devices for primary and secondary fusion equipment.

shows the field application of a complete set of automatic detection devices for primary and secondary fusion equipment. The equipment was tested in the Electric Power Research Institute of State Grid Jiangsu Electric Power Company and successfully connected with the quality inspection and control system. The system can have terminal detection, transformer detection, and primary and secondary complete detection functions.

The terminal wave recording system is shown in Figure 8: The PC sends the waveform to the electronic signal source through the protocol, and the electronic signal source outputs the waveform to the terminal. The PC reads the waveform data through the 104 protocol, and the sampling rate is set to 200 kS/s. When the trigger mode of the recorder is set on the PC, the recorder automatically collects the waveform output from the electronic signal source. After reading the waveform data of the recorder and 104, the PC generates the time series of the waveform. The waveform of the power source is compared with the sampled waveform of the terminal to obtain the similarity of the two waveforms so as to provide data support for measuring the health status of the terminal.

5. Results and Discussion

Based on the above experimental platform, the source signal waveform recorded by the wave recorder and the terminal sampling waveform are played back through the software on the PC side, and the terminal waveform disturbed by noise is filtered and decomposed. The playback waveforms are shown in Figures 9 and 10.

Figure 9 is the playback waveform of the source signal which is sampled by the wave recorder in the software of the PC side. The experimental site determines the benchmark comparison window according to the sampling point of the test waveform, and the benchmark comparison window of the test waveform is the sampling period. Figure 10 shows the actual noise signal which is obtained by filtering the sampling waveform from the actual terminal waveform, which contains noises. The noise signal is detected by the power distribution terminal. After the PC terminal completes the sampling and the playback of the waveform signal, a discrete time series is obtained. Finally, the time series of the source signal waveform and the terminal sampling waveform obtained on the PC terminal benchmark comparison window are extracted for Matlab waveform comparison.

Select four groups of waveforms for on-site testing. After sampling is completed, compare the two waveform time series of each group in Matlab; that is, the source signal waveform sequence and the terminal sampling waveform sequence. Figures 11 and 12 are schematic diagrams of energy entropy ratios of first two groups of waveforms sampled on-site. Figure 13 shows the matching path, and Table 3 shows parameters obtained by each group of waveforms under the algorithm of this paper.
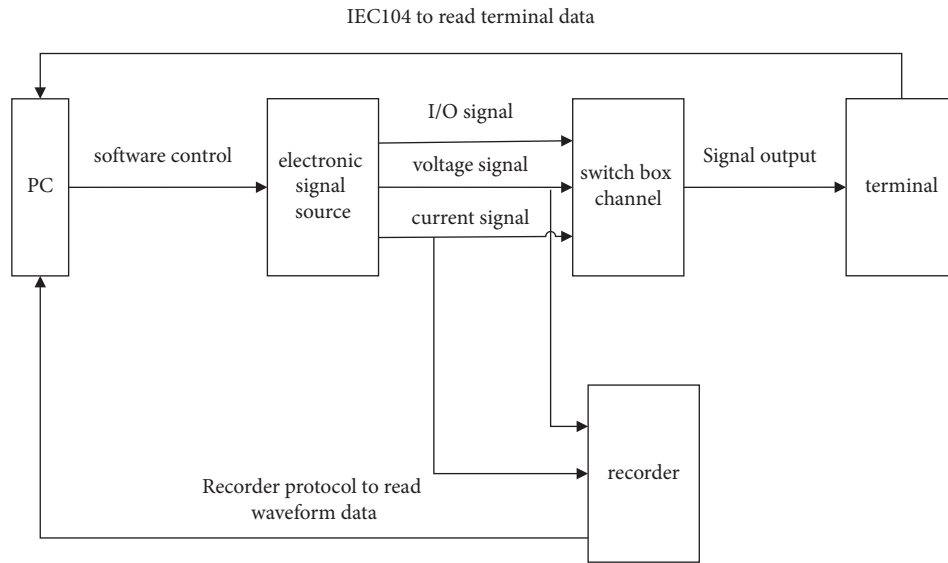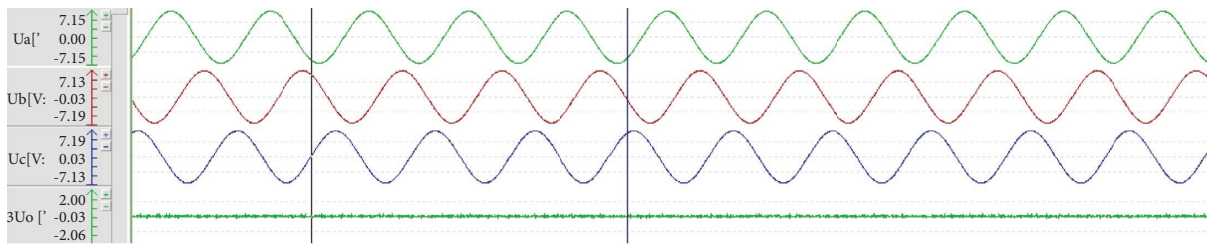
FIGURE 8: Terminal wave recording system.

FIGURE 9: The playback waveform of the power source waveform on the PC side.
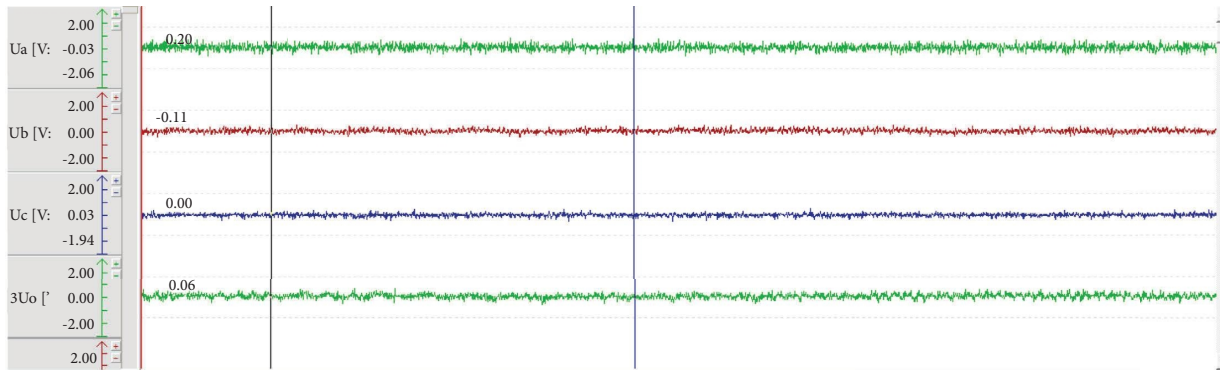

FIGURE 10: The noise interference waveform attached to the terminal sampling waveform in practical application.

Since the longest common substring is an optimization of the original algorithm in terms of data, Table 3 records the four groups of waveforms with the longest distance before and after the common substring algorithm processing and related parameters, where Dist_m and ω_m are the waveform distance and similarity before correction, $\text{sd}_{\max}$ and l are the maximum standard deviation and length obtained under the longest common substring algorithm, α is the penalty

coefficient based on $\text{sd}_{\max}$ and l, the waveform distance and similarity are corrected by α, and the obtained Dist_n and ω_n are the corrected waveform distance and similarity.

Experiments show that the improved DTW algorithm can effectively measure the similarity between the terminal waveform and the source signal waveform and provide effective data for the health status assessment of power distribution terminals.
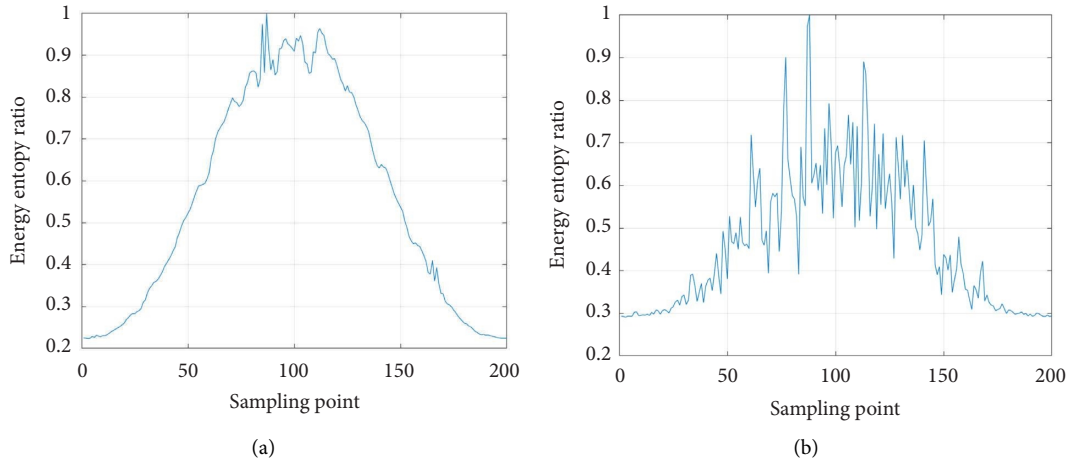
(a)

(b)

FIGURE 11: Schematic diagram of energy entropy ratio of the first set of waveforms sampled in the field. (a) Source waveform energy entropy ratio. (b) Disturbed terminal waveform energy entropy ratio.
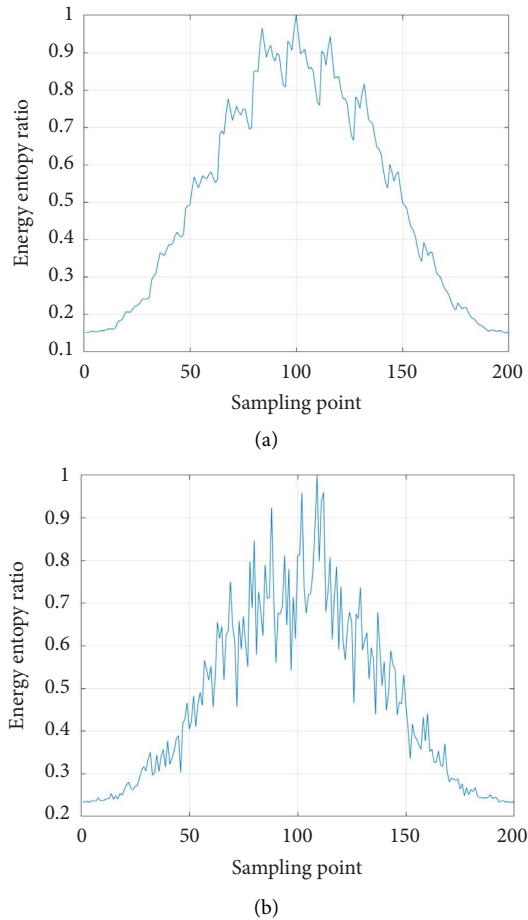

(a)

(b)

FIGURE 12: Schematic diagram of energy entropy ratio of the second set of waveforms sampled in the field. (a) Source waveform energy entropy ratio. (b) Disturbed terminal waveform energy entropy ratio.

The comparison between the energy entropy ratio and the response time of the DTW algorithm of the common substring is shown in Figure 14. The main time difference is in the calculation of the energy entropy ratio. The overall performance increases w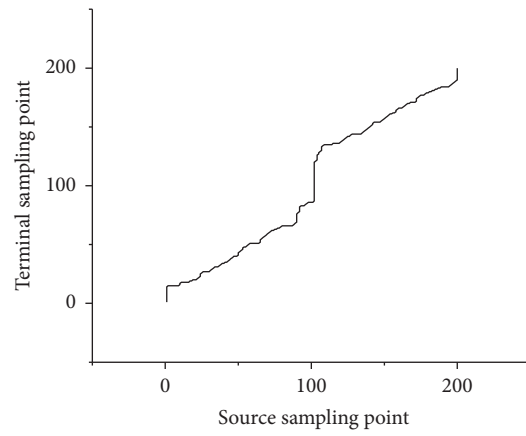ith the number of nodes because the response time differs very little after the data are preprocessed. There is no significant change in response time, and the advantage is more obvious. It shows that the preprocessing of the waveform through the short-term energy entropy ratio can effectively improve the calculation efficiency.
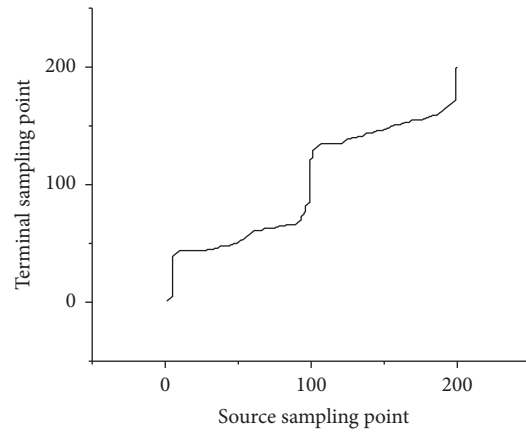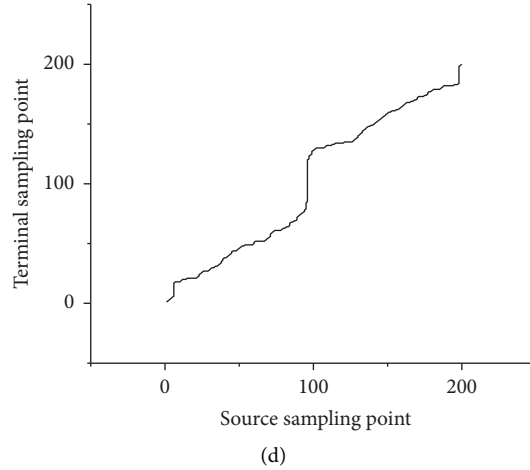
(a)
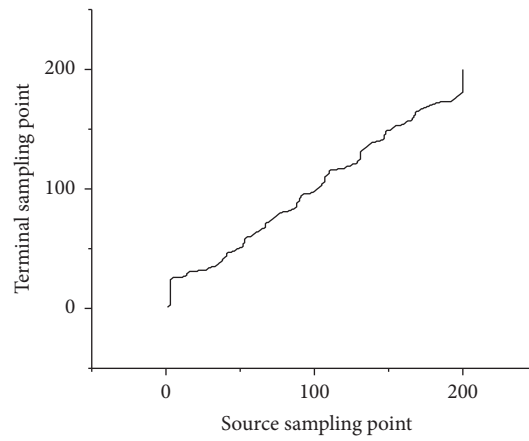
(b)

(c)

Figure 13: Continued.

(d)

FIGURE 13: Dynamic paths of four groups of field measured waveforms. (a) Path planning of the first set of waveforms sampled in the field. (b) Path planning of the second set of waveforms sampled in the field. (c) Path planning of the third set of waveforms sampled in the field. (d) Path planning of the fourth set of waveforms sampled in the field.

TABLE 3: Experimental waveform similarity under the improved DTW algorithm.

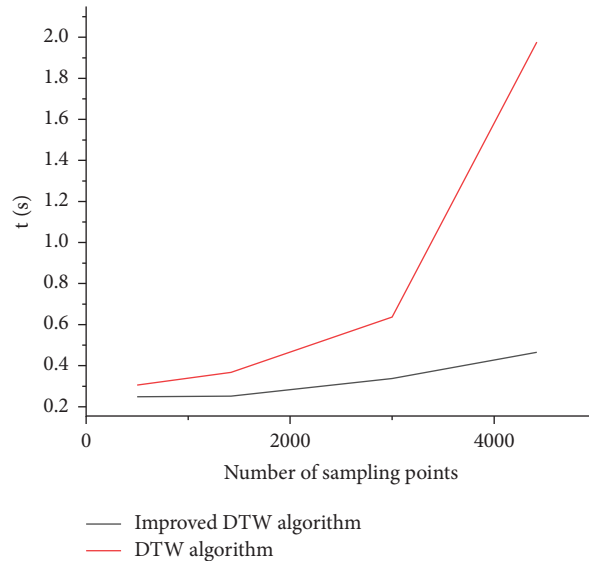| $Dist_m$	sd_{max}	l	α	$Dist_n$	ω_m	ω_n
20.9409	0.1325	70	0.6500	13.6116	0.0456	0.0684
8.6882	0.1973	67	0.6550	5.7777	0.1032	0.1475
9.1907	0.2015	69	0.6550	6.0199	0.0981	0.1425
5.6425	0.2733	87	0.5650	3.1880	0.1505	0.2388

FIGURE 14: Response time of the algorithm.

6. Conclusions

It is also of great significance to study the similarity theory of voltage and current waveforms for the field of electrical science research. In this paper, the similarity data analysis of the time series of the terminal waveform under different noises and sampling periods is carried out, and the effects achieved after optimization are as follows:

(1) Using the short-term energy entropy ratio to effectively extract the features contained in the original signal, the waveform data are preprocessed, and the characteristic events are highlighted

(2) The complexity of input features is reduced, and the similarity matching operation efficiency in waveform data processing is improved

(3) By adopting the matching method of the longest common substring, the DTW optimized distance is calculated, which overcomes the "ill-conditioned matching" problem in the distance calculation of periodic phase mismatch and improves the accuracy of the algorithm

This method provides data support and decision-making basis for terminal health status assessment. The algorithm analyzes massive data, explores the difference in waveform similarity under different interference conditions, and can realize the function of preliminarily judging the fault type. It has certain reference significance for realizing terminal health status assessment.

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

References

[1] X. Xing, H. Xu, X. Cao, and M. Liao, "A comprehensive assessment method of distribution terminal health status based on cloud model and D-S evidence theory," *Power System Protection and Control*, vol. 49, no. 13, pp. 72–81, 2021.

[2] R. Han, *Time series analysis method based on network theory and its application*, East China University of Science and Technology, Shanghai, China, 2020.

[3] Z. He, C. Wu, Y. Tian, X. Zhang, Q. Chen, and G. Liu, "A review of similarity measurement and indexing methods for big data in geoscience time series," *Geological Science and Technology Bulletin*, vol. 39, no. 4, pp. 44–50, 2020.

[4] Q. Huang and R. Liu, "Overview of similarity measurement of cloud models," *Data Communications*, vol. 6, pp. 43–49, 2019.

[5] D. Gao, H. Kao, and B. Wang, "Misconception of waveform similarity in the identification of repeating earthquakes," *Geophysical Research Letters*, vol. 48, no. 13, 2021.

[6] A. Tsuchiyama and J. Nakajima, "Diversity of deep earthquakes with waveform similarity," *Physics of the Earth and Planetary Interiors*, vol. 314, Article ID 106695, 2021.

[7] L. Zhang, T. Zhou, Y. Shao, H. Long, and Q. Du, "Research on audio similarity comparison algorithm based on physical features," *Television Technology*, vol. 41, pp. 110–114, 2017.

[8] L. Zhao, Y. Zheng, N. Tong, F. Wei, S. Xiao, and X. Lin, "Fault-adaptive reclosing scheme of multi-terminal flexible and direct power grid based on waveform similarity characteristics," *High Voltage Technology*, pp. 1–14, 2022.

[9] Y. Du, *Optimization of Coalbed Methane Drainage and Production System Based on Time Series Analysis*, China University of Petroleum, Beijing, China, 2018.

[10] H. Yang, X. Lu, C. Lou, X. Xiang, and L. Hu, "Analysis of key waveforms of weather radar using Fréchet distance algorithm," *Meteorological, Hydrological and Oceanographic Instruments*, vol. 38, no. 2, pp. 5–8, 2021.

[11] X. Pan, A. Ma, L. Wu, F. Liu, and J. Guo, "Research and application review of trajectory data similarity measurement method based on time series," *Journal of Yanshan University*, vol. 43, no. 6, pp. 531–545, 2019.

[12] X. Ao Liu, Ri-G. Zhou, A. El-Rafei, F. X. Li, and R. Q. Xu, "Similarity assessment of quantum images," *Quantum Information Processing*, vol. 18, no. 8, p. 244, 2019.

[13] Z. Yang, L. Tian, G. Zhuang, and J. Xu, "Application of mathematical model DTW and KNN algorithm in risk division of infectious disease areas," *China Science and Technology Information*, vol. 1, pp. 69-70, 2021.

[14] T. Zhu, Z. Peng, Z. Zhu, Y. Xie, W. Wang, and Z. Zhang, "A closed-loop detection algorithm based on dynamic time warping," *Control and Information Technology*, vol. 4, pp. 83–88, 2021.

[15] K. Ye, Z. Chen, J. Shi, C. Hu, and R. Wang, "Optimization of dynamic time warping algorithm," *Software Guide*, vol. 20, no. 1, pp. 132–135, 2021.

[16] Q. Ran, Z. Zhang, X. Yin, Y. Rehman, B. Liu, and J. Han, "Vertical protection scheme of DC distribution network lines based on improved dynamic time bending distance algorithm," *Electric Power Automation Equipment*, pp. 1–14, 2022.

[17] Z. Li, J. Guo, Y. Gao, and H. Mao, "Similarity measurement method for multivariate time series," *Control and Decision*, vol. 32, no. 2, pp. 368–372, 2017.

[18] S. Li, Z. Li, J. Li et al., "Application of event-triggered cubature kalman filter for remote nonlinear state estimation in wireless sensor network," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 6, pp. 5133–5145, 2021.

[19] Z. Li, S. Li, B. Liu, S. S. Yu, and P. Shi, "A stochastic event-triggered robust cubature kalman filtering approach to power system dynamic state estimation with non-Gaussian measurement noises," *IEEE Transactions on Control Systems Technology*, pp. 1–8, 2022.

[20] B. Liu, Z. Li, X. Chen, Y. Huang, and X. Liu, "Recognition and vulnerability analysis of key nodes in power grid based on complex network centrality," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 65, no. 3, pp. 346–350, 2018.

[21] X. Meng, X. Hu, and B. Shen, "Anomaly analysis method of spatial behavior trajectory of community elderly people," *Science Technology and Engineering*, vol. 21, no. 9, pp. 3676–3681, 2021.

[22] C. Chen, W. Li, and Y. Chen, "Similarity analysis of hydrological element time series data based on optimized DTW algorithm," *Hydrology*, vol. 41, no. 3, pp. 44–48, 2021.

[23] Y. Lu, J. Cao, X. Lv, Y. He, and R. Tian, "Seismic random noise suppression based on dynamic time warping ICA algorithm," *Petroleum Geophysical Survey*, vol. 57, no. 5, pp. 697–704, 2018.

[24] M. Liu, C. Zhou, and T. Zhang, "DTW similarity measurement based on segmentation features and adaptive weighting," *Computer Engineering*, vol. 47, no. 8, pp. 62–68, 2021.