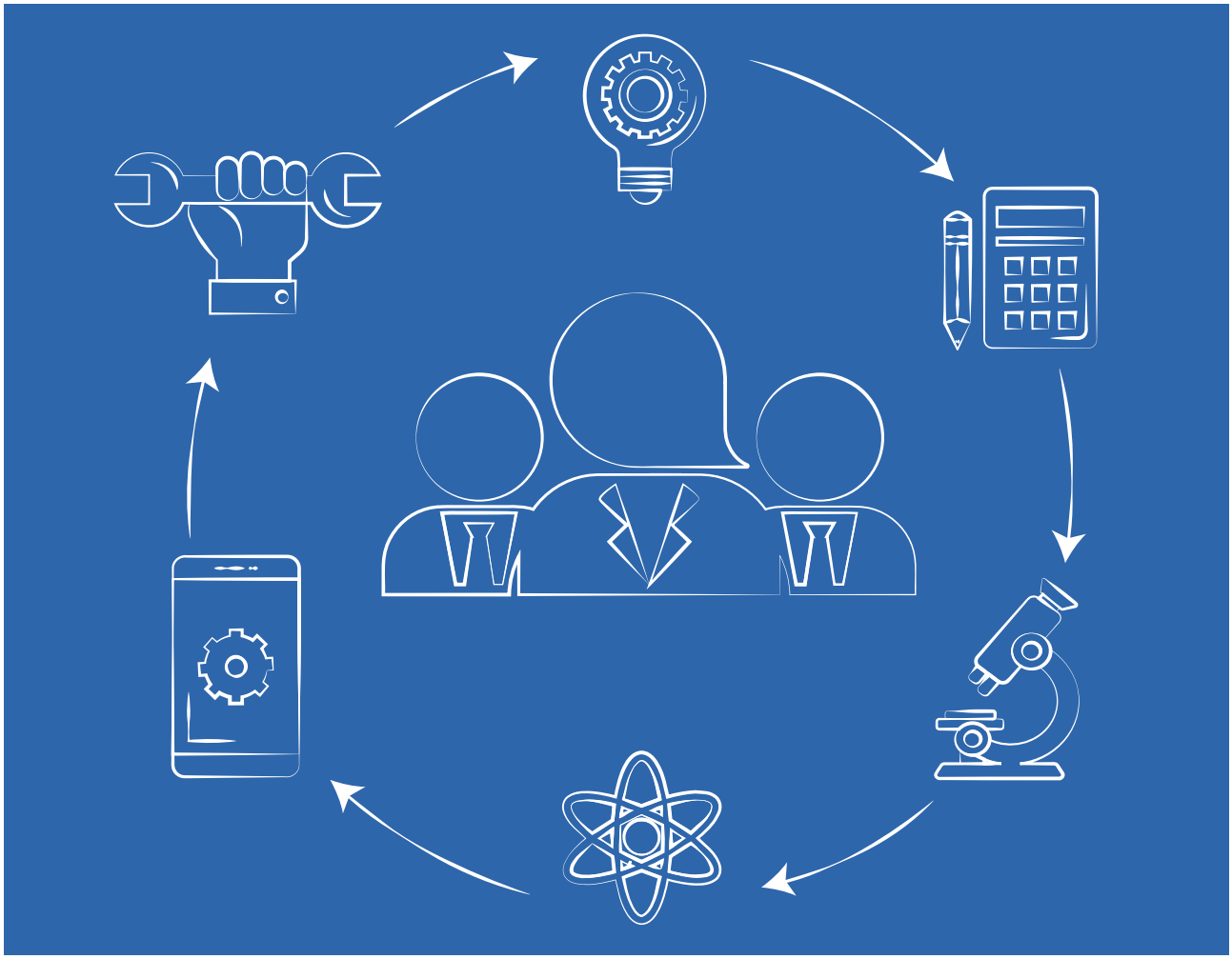
3C tecnología
Glosas de innovacion aplicadas a la pyme

Ed.45 | Vol.13 | N.1
April - June 2024

ISSN: 2254-3376



https://doi.org/10.17993/3ctecno.2024.v13n1e45

3C Tecnología. Glosas de innovación aplicadas a la pyme.

Quarterly periodicity.

Edition 45, Volume 13, Issue 1 (April - June 2024).

National and internacional circulation.

Articles reviewed by the double blind peer evaluation method.

ISSN: 2254 - 4143
Legal: A 268 - 2012
DOI: https://doi.org/10.17993/3ctecno.2024.v13n1e45

Edita:
Área de Innovación y Desarrollo by UP4 Institute of Sciences, S.L.
info@3ciencias.com _ www.3ciencias.com

https://doi.org/10.17993/3ctecno.2024.v13n1e45

https://doi.org/10.17993/3ctecno.2024.v13n1e45

EDITORIAL BOARD

https://doi.org/10.17993/3ctecno.2024.v13n1e45

AIMS AND SCOPE

PUBLISHING GOAL

3Ciencias wants to transmit to society innovative projects and ideas. This goal is reached thought the publication of original articles which are subdue to peer review or thorough the publication of scientific books.

TEMATIC COVERAGE

3C Empresa is a scientific - social journal, where original works are spread, written in English, for dissemination with empirical and theoretical analyzes on financial markets, leadership, human resources, market microestructure, public accounting and bussines management.

OUR TARGET

- Research staff.
- PhD students.
- Professors.
- Research Results Transfer Office.
- Companies that develop research and want to publish some of their works.

SUBMISSION GUIDELINES

3C Empresa is an arbitrated journal that uses the double-blind peer review system, where external experts in the field on which a paper deals evaluate it, always maintaining the anonymity of both the authors and of the reviewers. The journal follows the standards of publication of the APA (American Psychological Association) for indexing in the main international databases.

Each issue of the journal is published in electronic version (e-ISSN: 2254-3376), each work being identified with its respective DOI (Digital Object Identifier System) code.

STRUCTURE

The original works will tend to respect the following structure: introduction, methods, results, discussion/conclusions, notes, acknowledgments and bibliographical references.

The inclusion of references is mandatory, while notes and acknowledgments are optional. The correct citation will be assessed according to the 7th edition of the APA standards.

https://doi.org/10.17993/3ctecno.2024.v13n1e45

PRESENTATION WORK

All the information, as well as the templates to which the works must adhere, can be found at:

https://www.3ciencias.com/en/journals/infromation-for-authors/

https://www.3ciencias.com/en/regulations/templates/

ETHICAL RESPONSIBILITIES

Previously published material is not accepted (they must be unpublished works). The list of signatory authors should include only and exclusively those who have contributed intellectually (authorship), with a maximum of 4 authors per work. Articles that do not strictly comply with the standards are not accepted.

STATISTICAL INFORMATION ON ACCEPTANCE AND INTERNATIONALIZATION FEES

- Number of accepted papers published: 6.
- Level of acceptance of manuscripts in this number: 80%.
- Level of rejection of manuscripts: 20%.
- Internationalization of authors: 1 countries (China).

Guidelines for authors: https://www.3ciencias.com/en/regulations/instructions/

https://doi.org/10.17993/3ctecno.2024.v13n1e45

INDEXATIONS

https://doi.org/10.17993/3ctecno.2024.v13n1e45

INDEXATIONS

https://doi.org/10.17993/3ctecno.2024.v13n1e45

/SUMMARY/

https://doi.org/10.17993/3ctecno.2024.v13n1e45

https://doi.org/10.17993/3ctecno.2024.v13n1e45

https://doi.org/10.17993/3ctecno.2024.v13n1e45

https://doi.org/10.17993/3ctecno.2024.v13n1e45

/01/

FINITE ELEMENT ANALYSIS OF POWER CABLE CONNECTION LINE FAULTS UNDER THE TIME-DOMAIN REFLECTION METHOD

Yuning-Tao*

- College of Electrical Engineering and New Energy, China Three Gorges University, Yichang, Hubei, 443000, China
- taoyuning12@163.com

Reception: 12 April 2024 | **Acceptance**: 6 May 2024 | **Publication**: 18 June 2024

Suggested citation:

Yuning-Tao (2024). **Finite element analysis of power cable connection line faults under the time-domain reflection method.** *3C Tecnología. Glosas de innovación aplicada a la pyme. 13(1)*, 15-33.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.15-33

ABSTRACT

In this paper, finite element analysis of power cable connection line faults based on the time-domain reflection (TDR) method is designed to improve the accuracy and efficiency of fault detection. In terms of methodology, a distributed parameter model is used to describe the cable, and based on the fluctuation equation and characteristic impedance theory, the impedance characteristics under different fault types are analyzed. In terms of results, the effectiveness of the method is verified by mathematical modeling and simulation of single-phase ground faults and disconnection faults through the Simulink platform in MATLAB. The simulation results show that in terms of fault localization accuracy, our method controls the error within 0.07m, which is significantly improved compared with the traditional method. In addition, the fault detection under different noise signals is analyzed, which further establishes the accuracy of the TDR method in judging different fault types. The TDR-based power cable fault detection method proposed in this study not only improves the accuracy of fault localization, but also provides a reliable theoretical basis for the identification of different types of faults, which contributes to the stable operation and maintenance of power systems.

KEYWORDS

Time domain reflectometry (TDR), power cable faults, fluctuation equations, impedance characterization

INDEX

https://doi.org/10.17993/3ctecno.2024.v13n1e45.15-33

1. INTRODUCTION

Power cable is an important channel for the transmission of information and power in the power system, the cable line in the service process will often be subject to external and internal factors such as damage, such as man-made barbaric construction, animal bites, high temperature, corrosive gases, liquids, etc., as well as with the increase in the service time resulting in the aging of the insulation layer of the surface of the cable and other situations, the above phenomena will be on the transmission of power and even the entire power grid system security power supply caused by the important impact [1-3]. All these phenomena will have an important impact on power transmission and even the safety of the entire power grid system [1-3]. All-weather monitoring of cables in the core areas or key transmission channels can greatly improve the safety of the power grid system and reduce the major economic losses and safety accidents caused by cable failures [4-5].

With the increasing proportion of high-voltage power cables in transmission and distribution lines, the safety, stability and reliability of their operation have received widespread attention [6-8]. At present, high-voltage power cables usually use a single-core structure, this structure of the cable failure, because of its insulation thickness, the use of metal sheath cross-interconnection wiring mode and the use of cable T-joints and many other factors, resulting in high-voltage single-core cables than the 10kV three-core cable fault localization is more difficult [9-10].

Cable is an important material and equipment for power system to transmit information and electric energy. With the rapid development of cities and the increasing scale of industrial electricity consumption, the safety and reliability of cables, as an important carrier of information and energy, have been increasingly emphasized. Literature [11] proposed a novel shape memory alloy (SMA) cable constrained high damping rubber (SMA-HDR) bearing. It can effectively simplify bridge design and mitigate cable failures due to potential damage to cables from strong NF earthquakes. Literature [12] studied the submarine cable detection method in shallow areas (sea area within 200 meters), and pointed out that Brillouin RF can be used to detect submarine cables, and the method has no negative impact on cables. Some discussions on the development and trend of submarine cable detection technology were also carried out. Literature [13] designed a metal device around the feeder line, aiming to reduce and minimize the ground fault current generated in the event of a ground fault in a substation powered by a cable line. The reliability of this metal device was confirmed by simulation tests. Literature [14] envisions a new type of zero sequence current filter and ground fault protection relay based on the principle of cable shielding against ground currents. Numerical tests using PowerFactory simulation software corroborate the high sensitivity and low number of false trips of this device. Literature [15] conceived a coordinated protection system designed to ensure stable operation of the power system even when high temperature superconducting cables are used, and the feasibility and superiority of the scheme were corroborated by simulation verification using the PSCAD/EMTDC program. Literature [16] designed a modified K-means algorithm for feeder selection for full

https://doi.org/10.17993/3ctecno.2024.v13n1e45.15-33

cable network grounding faults to increase the accuracy of feeder selection for full cable network grounding under various fault conditions. A full cable network model with five feeders is used for testing and it is proved that this method has high feeder selection accuracy. Literature [17] discusses the function of distributed temperature sensor, which is a spatially continuous temperature sensor that measures the temperature along the cable through a sensing fiber pair. Based on simulation tests and formula calculations, it is feasible to detect and locate power cable heating faults using this tool. Literature [18] presents a fast method for detecting and handling faults inside and on the DC side of a full-bridge bipolar MMC-HVDC line. Electromagnetic transient simulations corroborate that the above method can exert a positive influence on transient voltages and current stresses, among others. Literature [19] conceived a new method for cross-linked polyethylene cable insulation fault detection based on fiber optic temperature sensors. The distributed fiber optic temperature measurement technique was operated for insulation degradation test of XLPE pairs of cables, and it was pointed out that the above method can accurately locate the insulation faults. Literature [20] proposed a method based on input impedance spectrum to solve the problem of strong coupling between conductors of 10kv three-core armored cables, where defects and fault types are difficult to identify. After PSCAD circuit simulation tests, the results show that the proposed method can effectively identify faults and local defects in three-core cables. Literature [21] proposes an original algorithm based on known arc combustion theory to simulate the maximum overvoltage value during arc combustion. The effect of arc suppression coils on the nature and magnitude of transient overvoltages and arc excitation overvoltages due to cable insulation breakdown is investigated. According to the simulation tests on the operation simulation of different power grids, it is found that the proposed model can clarify the minimum insulation margin and overvoltage protection trigger threshold of power grid equipment to ensure the reliable operation of the power grid.

The research idea of this paper is firstly based on the time-domain reflection (TDR) method, which utilizes the distribution parameter model to characterize the cable connection line. The impedance characteristics under different fault types are analyzed by constructing fluctuation equations and characteristic impedance theory. Subsequently, the Simulink platform in MATLAB is used to conduct simulation experiments, including the mathematical modeling of single-phase ground faults and disconnection faults. The effectiveness of the theoretical method is verified by observing the simulation results. At the same time, the changes in impedance characteristics of different types of faults are explored, as well as the influence of noise signals on fault detection. Finally, through data analysis and theoretical comparison, conclusions are drawn to demonstrate the advantages of this research method in improving the accuracy and efficiency of fault detection.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.15-33

2. METHODS

2.1. FAULT LOCALIZATION TECHNIQUE BASED ON TIME-DOMAIN REFLECTION METHOD

Power cable connection line belongs to a kind of transmission line, in the ideal case, the cable connection line can be regarded as a uniform transmission line, so it is feasible to use distribution parameters to describe the cable model. Its equivalent distribution parameter model in R, L, C, G represent the distribution resistance, inductance, capacitance and conductance per unit length of the transmission line, respectively.

Its fluctuation equation can be obtained based on the above parameters, which are obtained:

$$\begin{cases} -\frac{\partial u}{\partial x} = Ri + L\frac{\partial i}{\partial t} = (R + j\omega L)i \\ -\frac{\partial i}{\partial x} = Gu + C\frac{\partial u}{\partial t} = (G + j\omega C)u \end{cases} \quad (1)$$

The characteristic impedance Z_0 is the ratio of the incident wave voltage u to the incident wave current i, which can be obtained as:

$$Z_0 = \frac{u}{i} = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \quad (2)$$

Among them, the capacitance C and inductance L are related to the dielectric constant of the cable, the cross-sectional area of the material core, etc., which indicates that the wave impedance is different for different kinds of cables.For the transmission line with small losses, the characteristic impedance Z_0 can be simplified to $Z_0 = \sqrt{L/C}$ because of $\omega L \gg > R, \omega C \gg > G$,.The reflection coefficient can be expressed as follows.

$$\rho = \frac{Z_{\mathrm{L}} - Z_0}{Z_{\mathrm{L}} + Z_0} \quad (3)$$

Where, Z_L is the load at the fault point.According to the time interval between the injection pulse time and the pulse reflection time, the location of the fault point is calculated as:

$$L = \frac{1}{2} v_\rho \left(t_1 - t_2\right) \quad (4)$$

Where, v_p for the pulse propagation speed in the cable medium, and the dielectric constant of the medium related to t_1, t_2, respectively, for the pulse sent to the moment

https://doi.org/10.17993/3ctecno.2024.v13n1e45.15-33

and the pulse returned to the launch point of the moment. in the frequency is very high, the cable in the electromagnetic wave propagation speed tends to a constant constant.

According to the theory based on time-domain reflection method, we can obtain the basic characteristics of the impedance of the cable connection line fault as follows.

When the cable is normal, the load impedance and characteristic impedance is matched, that is, $Z_L - Z_0$, then according to the formula (3) reflection coefficient $\rho = 0$ no reflection echo. at this time the whole section of the cable impedance uniformity without mutation points, the impedance waveform curve is specifically expressed as a straight line section of basically unchanged amplitude.

When the cable connection line breakage fault occurs, there is $Z_L = \infty$, then the reflection coefficient $\rho = 1$, the incident wave and the reflection wave has the same polarity, the incident and reflected voltage waveform shown in Figure 1. at this time, the cable impedance in the breakage point of a sudden change in the impedance waveform curve is specifically expressed by the normal impedance waveform, a sudden increase in the amplitude of a certain location, and tends to be close to positive infinity. the distance between the point of the mutation and the normal impedance waveform starting point. The distance between the mutation point and the starting point of the normal impedance waveform, i.e., the incident/reflected pulse time difference, and the location of the actual cable connection line short-circuit fault point to meet the formula

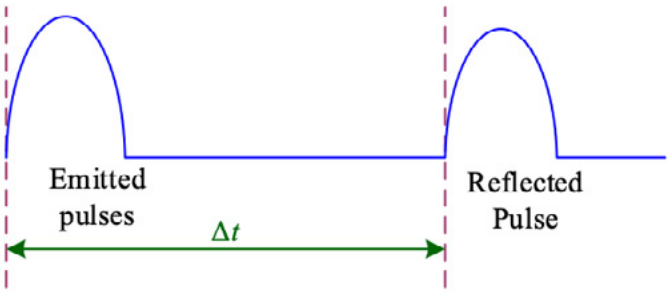


Figure 1. Schematic diagram of incident and reffected waves, when the cable open circuit occurs

When the cable short-circuit fault, there is $Z_L = 0$, then the reflection coefficient $\rho = -1$. incident and reflected waves of opposite polarity, the incident and reflected voltage waveforms shown in Figure 2. at this time, the cable impedance in the short-circuit point of a sudden change in the impedance waveform curve is specifically manifested by the normal impedance waveform suddenly in a position of amplitude changes to 0. The mutation point and the normal impedance waveform starting point of the distance between the distance between that is, the incident/reflected pulse time difference with the actual cable short-circuit fault point location to meet the equation (4).

https://doi.org/10.17993/3ctecno.2024.v13n1e45.15-33

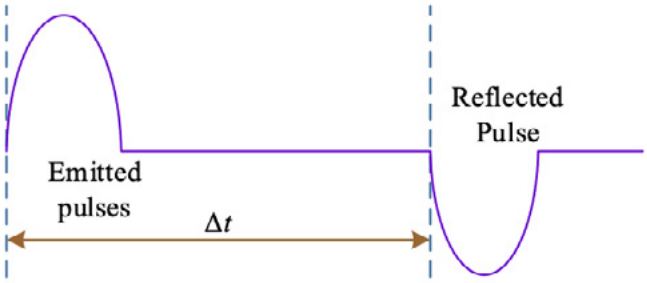


Figure 2. Schematic diagram of incident and reffected waves, when the cable short circuit occurs

Cable connection lines occur insulation skin folding, wear and shielding damage, furthermore, taking into account the on-board cables are usually bundled through the power center connector connection, where each cable is usually connected through the contact coupling, the introduction of the connection point will inevitably have an impact on the test line impedance. this time, there are $Z_L \neq 0, -1 < \rho < -1$, equivalent to the fault / anomaly for the inductor, capacitor, resistance, series or parallel connection. At this time, the cable impedance will also undergo a sudden change, but these states of the electric pulse signal can continue to propagate along the cable, only part of the signal in the fault/anomaly at the point of reflection, so the cable impedance after a sudden change can still be restored to a smooth state. the impedance waveform curve is specifically manifested as a section of the amplitude of a stable straight line in a certain position amplitude suddenly changed, and then return to normal, amplitude re The distance between the sudden change point and the normal impedance waveform start point, i.e., the incident/reflected pulse time difference, and the actual cable connection line fault/anomaly location also meets the formula (4).

According to the above analysis, it can be seen that when the cable connection line short-circuit and disconnection faults occur, the reflected waveform and cable fault impedance characteristics are obvious, and the impedance test can be used to directly determine the type of fault, while for other types of faults or anomalies, although it can be obtained from the characteristics of the change, but for the judgment of the type of faults or anomalies, the need for more in-depth theoretical support.

2.2. HIGH-PRECISION SINGLE CABLE FAULT DIAGNOSIS METHOD

The following analysis of the influencing factors of the positioning accuracy, from the TDR cable connection line fault location principle considerations, according to the law of linear superposition of error transfer, can be measured cable fault length *l* of the error can be expressed as follows.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.15-33

$$\Delta l = \frac{1}{2}(t\Delta v + v\Delta t) \quad (5)$$

v is the propagation speed of the TDR detection signal in the cable, t is the time interval between the peak of the incident wave and the peak of the reflected wave, Δv is the wave speed error, and Δt is the time interval measurement error.

The main reason for the existence of the error is mainly due to the measurement error of the wave speed, as well as the time interval measurement error. positioning accuracy is the largest positioning error, so in order to improve the positioning accuracy, the following by reducing the wave speed error and time measurement error to improve fault location accuracy.

2.2.1. REDUCED TIME MEASUREMENT ERROR

When analyzing the time measurement error of TDR peak point acquisition, when the sampling rate f is certain, the sampling time T is determined, and the waveform is finally recorded as a discrete point. Due to the principle of TDR fault detection, it is necessary to collect the incident and reflected peak points.When the peak points are collected by the direct counting method as shown in Fig. 3, the actual collection point may not be the real peak point, so that there exists a time measurement error, and the maximum time measurement error meets the formula.

$$\left|\Delta t'\right| \le \frac{T}{2} \quad (6)$$

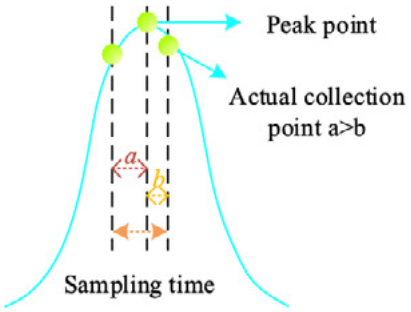


Figure 3. Schematic diagram of peak extraction by TDR through sampling

Since the system needs to collect the correlated incident wavehead and the correlated reflected wavehead, there are two samples, so the measurement error of the whole system satisfies the following equation.

$$\Delta t = \left|2\Delta t'\right| \le T = \frac{1}{f} \quad (7)$$

https://doi.org/10.17993/3ctecno.2024.v13n1e45.15-33

At present, the sampling rate of the prototype of the localizer is *500MHz*, and if it is assumed that there is no error in the propagation speed of the detected signals calculated by the theory before, it can be calculated that the positioning error is.

$$-0.207m \le \Delta l = \frac{1}{2}v\Delta t \le 0.207m \tag{8}$$

It can be seen that in the center frequency of *60.5MHz*, the signal propagation speed is *500MHz*, in the sampling rate of $2.15 * 10^8 \mathrm{m/s}$, the positioning error is within $\pm 0.215\mathrm{m}$, if you want to continue to reduce the positioning error, you need to improve the hardware sampling rate, but due to the increase in the hardware sampling rate caused by the substantial increase in the cost of the hardware, so taking into account the savings in hardware costs, through the peak extraction algorithm is studied to reduce the time measurement error and reduce the fault location error. Therefore, in consideration of saving hardware cost, the peak point extraction algorithm is used to reduce the time measurement error and fault localization error.

For the cable fault diagnosis signal obtained by the relevant operation in this paper, the signal sequence is expressed as follows:

$$(x_1, y_1), (x_2, y_2), \ldots, (x_i, y_i), \ldots, (x_N, y_N) \tag{9}$$

Where N is the length of the signal sequence, x_{i} is the time coordinate of the i th data point, and y_i is the amplitude of the i th data point.The principle of this fitting problem is that for a given signal sequence(x_i, y_i), Φ is the class of functions consisting of polynomials of all times not exceeding $m(m \le N)$, and $P_m(x) = \sum_{k=0}^{m} a_k x^k \in \Phi$ is found such that:

$$I = \sum_{i=1}^{N} \left[P_m(x_i) - y_i\right]^2 = \sum_{i=1}^{N} \left(\sum_{i=1}^{m} a_k x^k - y_i \right)^2 \tag{10}$$

The $P_m(x)$ that satisfies the minimum value of I in Eq. (10) is called the least squares fitting polynomial.

The TDR signal waveform can be fitted by a quadratic polynomial, and since the delay time needs to be obtained by analyzing the incident and reflected waveforms together, it is necessary to do two quadratic polynomial fits to the incident and reflected waveforms in the TDR signal. From this, we set the quadratic curve equation to be.

$$p(x) = a_0 + a_1 x + a_2 x^2(\tag{11}$$

https://doi.org/10.17993/3ctecno.2024.v13n1e45.15-33

2.2.2. REDUCED CALIBRATION SPEED ERROR

In practice, when fault localization is carried out by TDR on a certain cable connection, it is necessary to calibrate the speed, and the steps of calibration are as follows.

1) First take a section of cable of known length *l*.

2) The calibration method is not known in the case of signal propagation speed, the first given a signal propagation speed.

3) According to the signal transmission time is the same as the establishment of the equation, the solution to obtain the true signal propagation speed, the formula is.

$$\frac{l}{v} = \frac{l'}{v'} \tag{12}$$

Where l is the real cable length, l' is the measured cable length, v is the real signal propagation speed, v' is the given signal propagation speed.

According to Equation (12), the calibrated signal propagation speed can be calculated as.

$$v_{\text{Calibration}} = l * \frac{v'}{l'}(\tag{13}$$

However, in practice, due to errors in time measurements, it is necessary to add the time , so that the actual relationship is:

$$\frac{l}{v} = \frac{l'}{v'} + \Delta t \tag{14}$$

The exact true speed, then, is.

$$v = \frac{l}{\frac{l'}{v'} + \Delta t} \tag{15}$$

Therefore, the velocity error can be calculated according to Eq. (14) and Eq. (15), obtaining.

$$\Delta v = v_{\text{Calibration}} - v = l * \frac{v'}{l'} - \frac{l}{\frac{l'}{v'} + \Delta t} \tag{16}$$

Set $\frac{l'}{v'} = t$. Then the velocity error equation can be expressed as.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.15-33

$$\Delta v = \frac{l}{t} - \frac{l}{t+\Delta t} = \frac{l*\Delta t}{t^*(t+\Delta t)} = \frac{l}{t+\Delta t} * \frac{\Delta t}{t} = v^*\frac{\Delta t}{t} \quad (17)$$

Since the time measurement error Δt satisfies Eq.

$$0 \le |\Delta t| \le \frac{1}{f} \quad (18)$$

Therefore the range of speed measurement error is.

$$0 \le |\Delta v| \le \frac{v}{tf} \quad (19)$$

According to the above formula, the following conclusions can be drawn, because f is the real speed for a certain value, for the device sampling frequency, ignoring its error, is also a constant value. t is related to the length of the calibration cable, it can be seen that when the longer the length of the calibration, the smaller the range of speed measurement error. by increasing the length of the calibration cable, the calibration error can be confined to a smaller range to reduce the positioning error.

3. RESULTS AND DISCUSSION

3.1. POWER CABLE CONNECTION LINE FAULT SIMULATION

In this paper, the Simulink simulation platform in MATLAB is used for mathematical modeling of single-phase ground fault and single-phase disconnection fault of power cables respectively, and TDR is applied to analyze and process the reflected signals of the fault point to realize fault ranging of cables, and the feasibility and accuracy of the TDR on-line detection method and the localization algorithm designed in this paper are further verified by observing the detection effect under different signal-to-noise ratios. Accuracy of the TDR on-line detection method and the localization algorithm designed in this paper.

Due to the complexity of the power system line connection, so in the cable fault modeling process, to simplify the processing. according to the cable fault type, in Simulink to find the corresponding module for the connection, when the cable ground fault occurs, the specific simulation circuit model as shown in Figure 4, which includes the ranging signal generator module, power supply module, as well as the signal transmission and fault module, etc. where R indicates the beginning of the matching resistance, resistance value of 50Ω, C_1 for the 3.5 *nF* insulating capacitance. set the length of the first cable for 100m, the second cable section length of 500m, the total length of the line is 600m. in the two sections of the cable to join between the three-phase fault generator, this time, the cable fault occurs at the location of 100m. the three-phase power supply module in the figure for the ideal power supply system, the beginning of the phase is 0 degrees, the operating frequency is 0 degrees, the

https://doi.org/10.17993/3ctecno.2024.v13n1e45.15-33

operating frequency is 0 degrees, the operating frequency is 0 degrees. The three-phase power supply module in the figure is an ideal power system, the starting phase is 0 degrees, the operating frequency is 20Hz, taking into account the voltage amplitude in the experimental environment has nothing to do with the detection method, low-voltage value is more conducive to the graphical observation, so set the operating voltage of 220V.

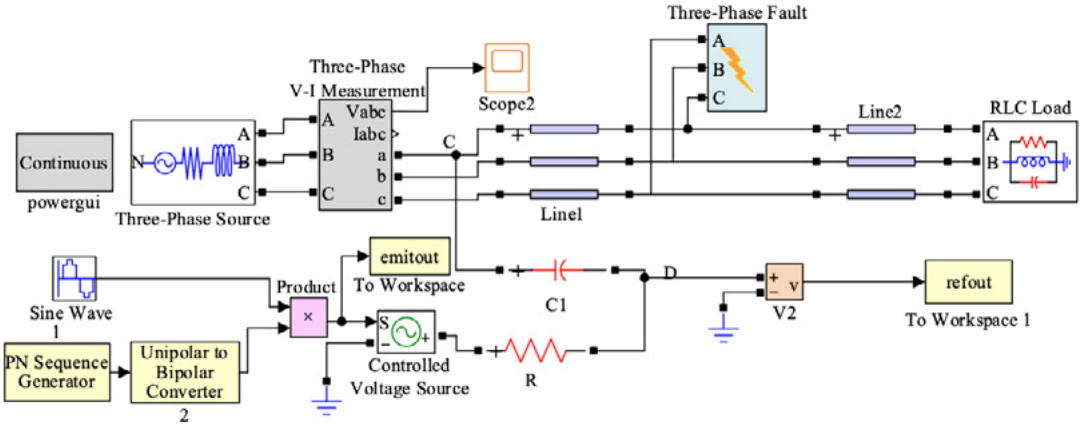


Figure 4. Cable grounding fault circuit model

Characteristic parameters of cable connection line include positive sequence and zero sequence two parts, cable characteristic parameter settings as shown in Table 1, which are set for the unit resistance, reactance and capacitive reactance of the two sections of the cable. positive sequence and zero sequence unit length reactance (Ω/km) are 0.2016 and 0.3774, respectively. through the three-phase fault generator to simulate single-phase low-resistance ground faults in cables, just select the phase A faults in the Fault Generator Parameter Dialog, and set the grounding resistance. The grounding resistance $R_g = 10\Omega$ is set.

Table 1. Cable feature parameter Settings

Parameter type	Positive sequence	Zero sequence
Reactance per unit length (Ω/km)	0.2016	0.3774
Reactance per unit length (H/km)	0.2085×10^{-2}	1.2131×10^{-2}
Unit length tolerance (F/km)	0.112×10^{-5}	0.212 x 10^{-5}

In order to further verify that the TDR method can complete the detection of intermittent cable link faults, the fault duration is set to 0.05s. For the test signals, the frequency of the local cosine carrier signal is $f = 10\mathrm{MHz}$, the corresponding duration of the *m* sequence is $2 \times 10^{-5}s$, and the period of the sequence is 128. The starting time of the simulation is 5s, respectively, through the To Workspace module, the reference signal and the fault signal are outputted to the MATLAB workspace, and analyzed by the signal processing module. in the process of the simulation, the fault location can be changed by setting the length of the first section of the cable, and at the same time, the total length of the cable can be adjusted. in accordance with the

https://doi.org/10.17993/3ctecno.2024.v13n1e45.15-33

above requirements to complete the parameter settings of the modules, the test signals are made of m-sequence and cosine signals in accordance with the 1:1 BPSK modulation, and their waveforms are shown in Fig. 5.

The upper and lower parts of the figure are the waveforms of the *m* sequence and the TDR test signal, respectively. it can be seen that in the same $2 \times 10^{-5} s$ time, the *m* sequence completed 15 reflections, while the TDR test signal reflection amplitude and number of times than the sequence. The test signal generated by the above is injected into the cable to be tested through the detection point (D), the test signal and the normal operation of the cable in the AC signal superposition of the AC signal. because of the amplitude of the test signal is very small compared to the effective signal transmission, almost does not have any impact on the normal operation of the signal transmission, so it is possible to use the TDR to achieve the on-line detection of cable faults.

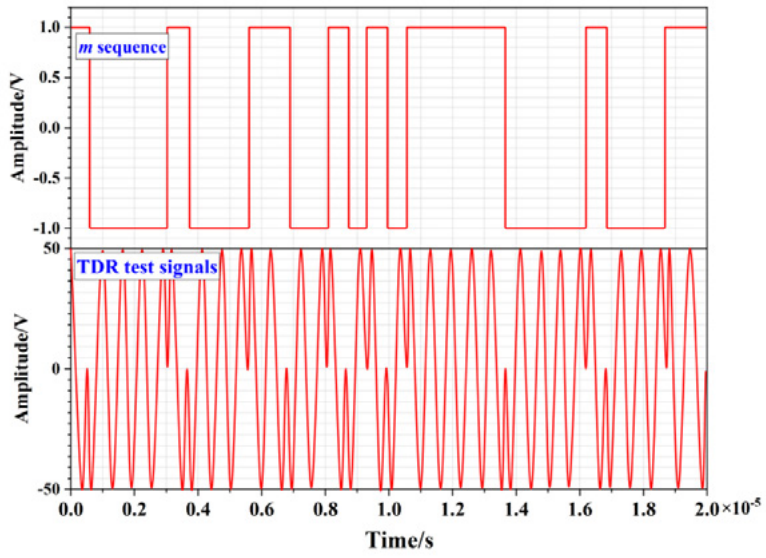


Figure 5. The sequence and corresponding TDR test signals

When the test signal in the cable transmission process encounters 100m at the single-phase ground fault point, will be reflected back, the reflection coefficient and the impedance value of the fault point of the size of the impedance value is related to the ground short-circuit fault, the value of the coefficient of -1, the reflected signal and the test signal is reversed. D point of the original signal collected by the high-pass filtering, to get the ideal zero-noise environment under the fault signal shown in Fig. 6. it can be seen that the fault signal in the cable transmission process amplitude frequency attenuation, from -50~50 (V) amplitude frequency attenuation of -9.63~9.85 (V). It can be seen that the fault signal in the cable transmission process has amplitude and frequency attenuation, compared with the test signal has a time delay, the time delay is about 0.2×10^{5}. The amount of time delay is related to the location of the fault point and the propagation speed of traveling wave in the cable.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.15-33

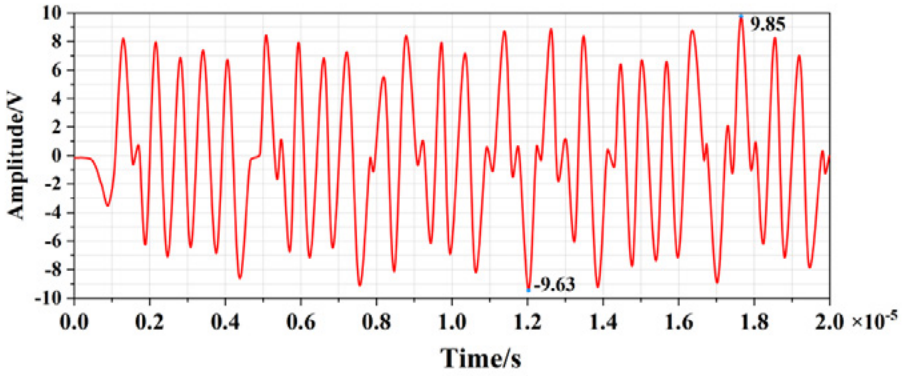


Figure 6. Reflected signal at the fault point under ideal conditions

Application of 2.2 chapter described in the high-precision single cable fault diagnosis method for isolation of fault signals and point C incident signal processing, the output of the system as shown in Figure 7. in order to facilitate the observation of the results of the figure has been transformed through the corresponding formula of the transverse coordinate by the sampling point into the amount of time, the transformation of the signal with the sampling frequency of the f related to the peak detection, you can get the relevant waveform peak point, the value of -0.1447, the corresponding horizontal coordinate is 1.8199×10^{-5}, that is, the delay time of the reflected signal is $1.8199 \times 10^{-5}\mathrm{s}$. From the characteristic parameters of the power cable. The velocity value of the traveling wave in the cable is estimated $v = 1/\sqrt{LC} = 1.6350 \times 10^{8}\mathrm{m/s}$ as by the characteristic parameter *L* and *C* of the power cable, and according to the formula of the cable fault ranging, it can be seen that the location of the cable fault is about 99.93m away from the detecting point D, which has an error of 100m from the actual fault. The error is only 0.07 m. Since the correlation peak is negative, the type of cable fault can be judged accordingly. in the case of high signal-to-noise ratio, the correlation peak is more obvious, which indicates that the TDR method can achieve better fault localization with high-precision single cable fault diagnosis algorithm.

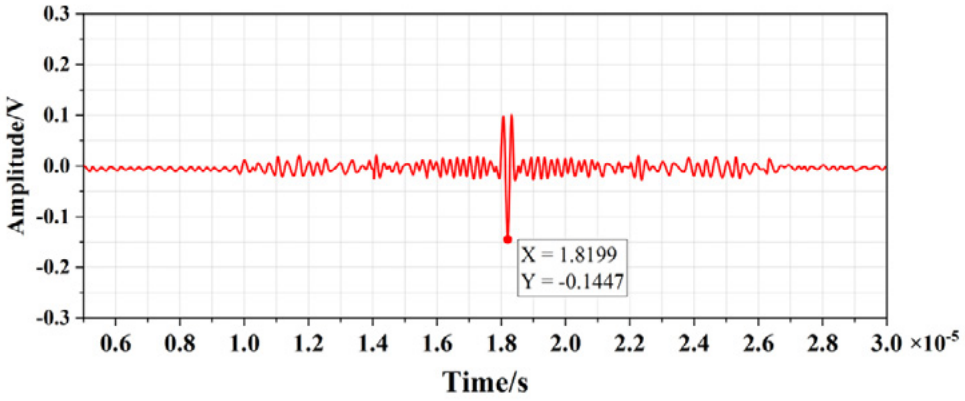


Figure 7. Waveform related to cable grounding faults at a high SNR

https://doi.org/10.17993/3ctecno.2024.v13n1e45.15-33

3.2. SIMULATION RESULTS ANALYSIS

The noise signals with different center wavelengths of TDR are used to detect the faults of different branches at the same time, and the simulation results are shown in Fig. 8, in which Fig. 8(a) shows the experimental results of detecting the faults of 63.60 m with the noise signals with the center wavelength of 1472.47 nm. The correlation peak in the 0 position is used as the reference peak of the measurement, and indicates the zero point of the measurement. The distance of the second highest correlation peak relative to the reference peak indicates the location of the break fault, and it is accurately located at 63.60 m. In addition, the two small peaks on the right side are caused by the reflected harmonics of the second (2nd refl) and the third (3nd refl) of the break fault.Fig. 8(b) sets up a short-circuit fault in the cable to be measured at 63.60 m, and the noise signal with the center wavelength of 1473.22 nm is used as the detected signal of the branch. It can be seen that there is a correlation peak at 63.60m, and it is negative (-158.45), indicating that a short-circuit fault has occurred there.

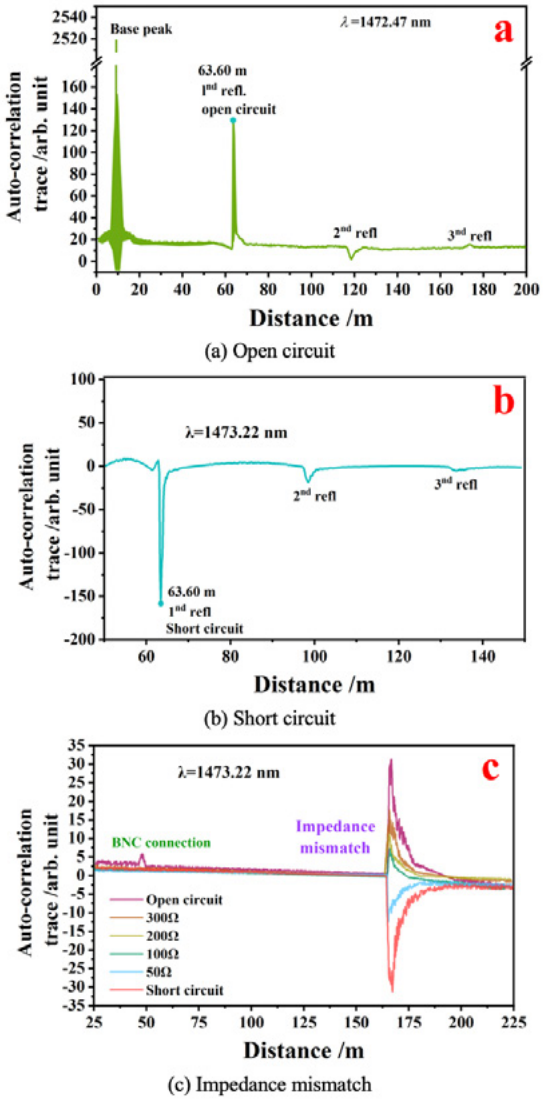


Figure 8. Detection results of cable faults on different cable branches

https://doi.org/10.17993/3ctecno.2024.v13n1e45.15-33

Fig. 8(c) shows the results of impedance mismatch fault detection using a noise signal with a center wavelength of 1474.75 nm. The impedance mismatch fault on the cable to be tested consists of a coaxial cable connector (BNC) connection point and an impedance tunable termination load, and the impedance values of the termination load are set to be 300, 200, 150, 50, and 0 Ω (short-circuit), respectively. Since the correlation peak in the disconnected state indicates total reflection, normalizing the peak there, the correlation peaks of the rest of the curve at the impedance mismatch indicate the actual reflection coefficients, which are 0.534, 0.338, 0.212, and -0.357, respectively, and therefore, according to the curves, the impedances at the impedance mismatches are 173±8, 157±6, 83±11, and 27±3 Ω. Similarly, the reflection coefficients and impedance values of the BNC connections are measured as Similarly, the reflection coefficient and impedance value of the BNC connection point can be measured as 0.0022 and 59±4Ω, respectively.

In the simulation experiment, the relative error is used to measure the measurement accuracy, and the relative error E_r is expressed as follows.

$$E_r = \frac{|x - a|}{a} \times 100\%$$

Where, *x* represents the actual measurement value of the length of power cable, represents the standard length of cable, which is agreed to be the nominal value of the cable under test. different lengths of cable were measured several times, and the polynomial fitting of the relative error results of different lengths were obtained, as shown in Fig. 9. it can be seen that in the measurement range of 100m, the relative error is around 1%, with the increase of the length of measurement, the relative error gradually increases, and in the measurement range of 700m, the relative error is controlled within the range of 3.73%. It can be seen that in the measurement range of 100 m, the relative error is around 1%, and with the increase of the measurement length, the relative error increases gradually, and in the measurement range of 700 m, the relative error is controlled in the range of 3.73%, which proves that the TDR method can be used for the online detection of cable faults, and it has high efficiency and low error.

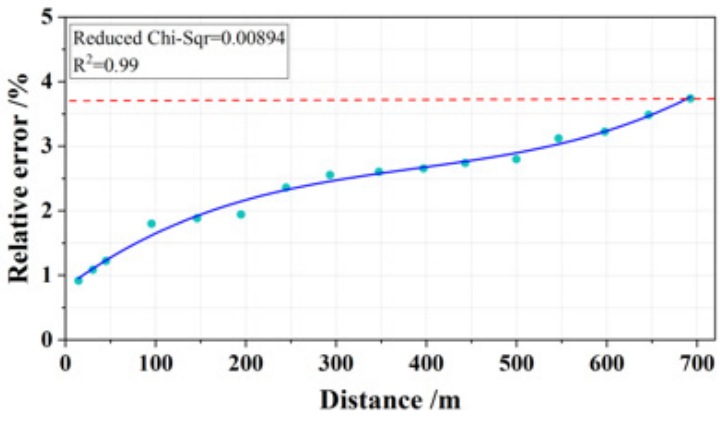


Figure 9 Relative error changes with measurement distance

https://doi.org/10.17993/3ctecno.2024.v13n1e45.15-33

4. CONCLUSION

In this study, we have achieved remarkable research results for the finite element analysis of power cable connection line faults by the time-domain reflection (TDR) method. Through simulation experiments, we found that the TDR method can effectively reflect the change of cable impedance in the case of single-phase ground faults and disconnected faults, so as to accurately locate the fault location. The simulation results show that the error of fault localization is controlled within 0.07m, which is much better than the traditional method.

In addition, this study also explores the fault detection effect under different noise signals, and finds that the TDR method is still effective in identifying and localizing faults even under different signal-to-noise ratio environments. For example, the location of the break fault detected under the noise signal with a center wavelength of 1472.47 nm is accurate to 63.60 m, which proves the stability and reliability of the method under different conditions.

The analyzed data show that the relative error is controlled within 1% in a measurement range of 100m, while in a longer measurement range (e.g. 700m), the error is still controlled within 3.73%. This result not only confirms the high efficiency of the TDR method for cable fault detection, but also demonstrates its feasibility for long distance detection.

The finite element analysis method of TDR based on power cable connection line faults proposed in this study proves its effectiveness both theoretically and practically. The method not only improves the accuracy and efficiency of fault localization, but also provides strong technical support for the safe operation of power systems. This result has important guiding significance for fault detection and prevention in the power industry, as well as for future cable design and maintenance.

REFERENCES

1. Tian, Y., Zhao, Q., Zhang, Z., Li, L., & Crossley, P. (2018). Current-phase-comparison-based pilot protection for normally closed-loop distribution network with underground cable. *International Transactions on Electrical Energy Systems, 28*(9), e2733.1-e2733.19.
2. Akbal, B. (2020). High voltage underground cable bonding optimization to prevent cable termination faults in mixed high voltage lines. *IET Generation Transmission & Distribution, 14*(20).
3. Li, B., He, J., Li, Y., Li, B., & Wen, W. (2020). High-speed directional pilot protection for mvdc distribution systems. *International Journal of Electrical Power & Energy Systems, 121*(6), 106141.
4. Cozza, Andrea. (2019). Never trust a cable bearing echoes: understanding ambiguities in time-domain reflectometry applied to soft faults in cables. *IEEE Transactions on Electromagnetic Compatibility.*

https://doi.org/10.17993/3ctecno.2024.v13n1e45.15-33

5. Samet, H., Khaleghian, S., Tajdinian, M., Ghanbari, T., & Terzija, V. (2021). A similarity-based framework for incipient fault detection in underground power cables. *International Journal of Electrical Power & Energy Systems, 133*(3), 107309.
6. Zhang, Zhihua, Xu, Bingyin, Crossley, Peter, et al. (2018). Positive-sequence-fault-component-based blocking pilot protection for closed-loop distribution network with underground cable. *International Journal of Electrical Power & Energy Systems.*
7. Cataldo, A., Benedetto, E. D., Masciullo, A., & Cannazza, G. (2021). A new measurement algorithm for tdr-based localization of large dielectric permittivity variations in long-distance cable systems. *Measurement, 174*(5), 109066.
8. Cerretti, A., D'Orazio, L., Gatta, F. M., Geri, A., Lauria, S., & Maccioni, M. (2023). Countermeasures for reduction of screen currents due to cross country faults in mv cable distribution networks. *Electric Power Systems Research.*
9. Akbal, B. (2019). Mssb to prevent cable termination faults for long high voltage underground cable lines. *Elektronika ir Elektrotechnika, 25*(6).
10. Bragatto, T., Cerretti, A., D'Orazio, L., Gatta, F. M., Geri, A., & Maccioni, M. (2019). Thermal effects of ground faults on mv joints and cables. *Energies, 12.*
11. Fang, C., Liang, D., Zheng, Y., & Lu, S. (2022). Seismic performance of bridges with novel sma cable-restrained high damping rubber bearings against near-fault ground motions. *Earthquake Engineering & Structural Dynamics, 51*(1).
12. Chen, Y., Li, X., Cai, C., Wu, C., Zhang, W., & Huang, X., et al. (2021). Submarine cable detection method based on multisensor communication. *Journal of Sensors(Pt.11), 2021.*
13. Popovi, Ljubivoje, M. (2018). Reduction of the fault current passing through the grounding system of an h v substation supplied by cable line. *International Journal of Electrical Power & Energy Systems, 99*, 493-499.
14. Lowczowski, K., Lorenc, J., Andruszkiewicz, J., Nadolny, Z., & Zawodniak, J. (2019). Novel earth fault protection algorithm based on mv cable screen zero sequence current filter. *Energies, 12*(16), 3190-.
15. Nguyen, T. T., Lee, W. G., Kim, H. M., Yang, H. S., & Sciubba, E. (2020). Fault analysis and design of a protection system for a mesh power system with a co-axial hts power cable. *Energies, 13*(1), 220-.
16. Wan, Q., Zheng, S., & Shi, C. (2022). Feeder selection method for full cable networks earth faults based on improved k-means. *IET generation, transmission & distribution.*
17. Chen, K., Yue, Y., & Tang, Y. (2021). Research on temperature monitoring method of cable on 10 kv railway power transmission lines based on distributed temperature sensor. *Energies, 14.*
18. Wenig, S., Goertz, M., Heinisch, M., Beckler, S., Kahl, M., & Suriyah, M., et al. (2018). Internal converter- and dc-fault handling for a single point grounded bipolar mmc-hvdc system. *Electric Power Systems Research, 161*(aug.), 177-187.
19. Liu, Y., Xiong, H., & Xiao, H. (2022). Detecting xlpe cable insulation damage based on distributed optical fiber temperature sensing. *Optical Fiber Technology, 68*, 102806-.
20. Wang, Y., Zhang, J., Yao, C., & Zhao, H. (2022). A mathematical method for local defects and faults identification of 10 kv three-core cable based on input impedance spectrum. *IET science, measurement & technology.*

https://doi.org/10.17993/3ctecno.2024.v13n1e45.15-33

21. Varetsky, Y., Gajdzica, M., & Kushka, B. (2022). Study of transient overvoltages on csi adjustable speed drives under arcing slgf in the industrial cable grid. *Electric Power Systems Research(Aug.), 209.*

https://doi.org/10.17993/3ctecno.2024.v13n1e45.15-33

/02/

Three-dimensional simulation and localization of power cable fault point by combining wavelet transform and neural network

Yuning-Tao*

- College of Electrical Engineering and New Energy, China Three Gorges University, Yichang, Hubei, 443000, China
- taoyuning12@163.com

Reception: 20 April 2024 | **Acceptance**: 6 May 2024 | **Publication**: 14 June 2024

Suggested citation:

Tao Yuning (2024). **Three-dimensional simulation and localization of power cable fault point by combining wavelet transform and neural network.** 3C Tecnología. Glosas de innovación aplicada a la pyme. 13 (1), 35-55. https://doi.org/10.17993/3ctecno.2024.v13n1e45.35-55

ABSTRACT

This study centers on the 3D simulation localization technology of power cable fault points, using the combination of wavelet transform and neural network, aiming to improve the accuracy and efficiency of cable fault localization. This paper adopts the method of combining wavelet transform and genetic algorithm optimization of back propagation (GA-BP) neural network. First, by constructing a three-dimensional simulation model of power cables, the wavelet transform is applied to extract fault features and the GA-BP neural network is utilized for fault point localization. The experimental results show that the average localization errors of this method in single-phase ground fault and two-phase grounded short-circuit fault are 0.112km and 0.126km, respectively, which are significantly better than the traditional method. Under different fault initial phase angle conditions, the proposed method shows strong adaptability and the error is controlled within 1%. Meanwhile, the present algorithm exhibits strong noise suppression ability under white and colored noise backgrounds, especially in low signal-to-noise ratio environments. In summary, this study demonstrates the effectiveness of the 3D simulation localization technique for power cable fault points combining wavelet transform and GA-BP neural network in improving the localization accuracy and noise resistance.

KEYWORDS

Power cable, three-dimensional simulation, wavelet transform, GA-BP neural network

https://doi.org/10.17993/3ctecno.2024.v13n1e45.35-55

INDEX

https://doi.org/10.17993/3ctecno.2024.v13n1e45.35-55

1. INTRODUCTION

China's urban power distribution network distribution form can be mainly divided into cable lines and overhead lines, the former has higher requirements for the installation environment and installation technology, but has the advantages of small power loss and strong stability, so power cables are widely used in undersea tunnels, transformer stations and other complex environments [1-3]. However, the installation environment of power cables determines the difficulty index of its fault detection increases linearly, compared with overhead lines, it is very difficult to diagnose and repair power cable faults in complex hidden environments such as undersea, underground, etc. [4-6].

With the rapid development of electric power science and technology, power cables have gradually become the "main artery" of power transmission in major cities. The overloaded operation of the power grid or long time use, will bring great burden to the power cable, and then appear low resistance failure, high resistance failure and other conditions. After the cable failure, rapid and accurate fault location, rapid restoration of reliable power supply, not only effectively reduce the fault outage time, but also reduce economic losses [7-8]. Due to the construction cost, cable channel conditions, power supply, accuracy and reliability and other factors, cable online monitoring and fault location technology has not yet been fully implemented, is still mainly offline testing for cable fault location [9-10].

Power cable fault diagnosis and localization techniques have been improved to reduce the outage time and maintenance cycle, as well as to enable fast power supply to the faulty local grid, to detect the operation of transmission lines and to reduce the economic losses.Li, G. et al. explored how to diagnose and locate the aging or deterioration of power cables by using impedance spectroscopy, and pointed out that the optimal frequency range of impedance spectrum analysis and a set of criteria for assessing the condition of cables can be determined by the reference frequency and the characteristic frequency, respectively. The reference frequency and characteristic frequency can be used to determine the optimal frequency range for impedance spectral analysis and a set of criteria for assessing the condition of cables, which is easier to diagnose and locate the aging of cables than previous methods, and the measurement process is simpler [11]. Li, M. et al. proposed a fault location method based on traveling waves, and improved it with an autonomous learning algorithm. Simulation experiments of the improved localization method on typical power cable circuits were carried out using PSCAD, and the results confirmed that the improved algorithm is fully capable of identifying short-circuit faults in power cable systems [12].Sommervogel, L. pointed out several types of faults that may be encountered during the detection of a cable's entire lifecycle and used a new methodology for modeling to conduct the study by combining measurement and simulation to improve the fault identification process [11]. measurement and simulation to improve the accuracy of fault localization, and the conclusions of the study provide an important reference for the practical application of cables [13].Tian, Y et al. conceptualized a method to conductive frequency protection of underground cables in closed-loop

https://doi.org/10.17993/3ctecno.2024.v13n1e45.35-55

distribution networks by using the current phase comparison method, which is executed through the remote unit (rtu) in the DAS, and the PSCAD simulation model is used, which verifies that the method can be used to locate and isolate the faults in time, and the application scenarios. locate and isolate faults with a wide range of application scenarios [14].Li, Z. et al. envisioned a deep-learning based cable infrared image state assessment method to measure the temperature of power cables, and based on the results of the statistical analysis, it was noted that this method can provide an effective and accurate state identification of the infrared maps of the cables [15].Siodla, K. et al. elaborated on the management of the medium-voltage cable network and the fault rate analysis of the current state of the art to manage MV cable networks by evaluating and diagnosing the operational status of individual cable lines obtained, these studies and conclusions are important references for cable line diagnostic studies [16].Samet, H. et al. conceptualized a method for the detection of early faults and differentiated such faults from other kinds of similar faults in the power system. The feasibility of the method was confirmed by simulation detection of four different arc models and analysis of actual fault data [17].Li, C. et al. investigated for the first time the effect of arc size on fire and flame propagation of 110 kV cross-linked polyethylene cables under the setting of constant power in arc discharge. After analog simulation testing, it was found that the larger the arc size, the more rapid the ignition rate of the cable fire, and the flame propagation in both vertical and horizontal azimuths increased significantly, which is of great significance for the prevention of and education about cable fires [18].A new distance-protection algorithm was proposed by Vinícius A. et al. to improve the protection of multi-terminal high-voltage direct current (MTDC) systems, and the feasibility of this algorithm was demonstrated by four-segment MTDC system detection and real-time simulation, which can provide more program choices for multi-terminal DC protection [19].

In this paper, the cable discharge data are collected and processed to build a model to simulate the actual fault scenario. Then, the wavelet transform technique is used to extract fault features, and the weights and thresholds of the backpropagation (BP) neural network are optimized by genetic algorithm (GA) to improve the training efficiency and accuracy of the network. Ultimately, the optimized GA-BP neural network is used to accurately locate the fault point. This process not only focuses on improving the accuracy of localization, but also considers the robustness of the algorithm in different noise environments.

2. THREE-DIMENSIONAL SIMULATION OF POWER CABLE FAULT POINT LOCALIZATION METHODS

2.1. CONSTRUCTING THREE-DIMENSIONAL SIMULATION MODELS OF POWER CABLES

The design of the three-dimensional simulation model of power cable needs to be based on data, that is, the collection of three-dimensional data, the establishment of

https://doi.org/10.17993/3ctecno.2024.v13n1e45.35-55

visualization information collection program, and the summary of the collected data to build a database. Power cable discharge data acquisition process, inevitably contains part of the noise information, resulting in poor authenticity of the three-dimensional model, to which this paper adopts the use of the Internet of Things and ZigBee networking real-time adjustment of the discharge signal data, to generate a three-dimensional visualization model of the power cable. After the partition block processing of the power cable, the fuzzy pixel value of the three-dimensional visual information acquisition is expressed as:

$$g(x,y) = h(x,y) * \eta(x,y) \tag{1}$$

Where (x,y) denotes the pixel value, *g* denotes the visual information sampling result, *h* denotes the set of association rules for 3D information sampling of power cables, * denotes the convolution, and η denotes the interference information.

Based on the above calculation results, fusing the collected data of each block, the template matching method is applied to establish a statistical analysis model that can be used for fault point localization. The vector of three-dimensional features of the power cable model is represented as:

$$t = A_s + ns \tag{2}$$

Where t denotes the 3D feature vector, n denotes the dimension of the 3D simulation model of power cables, s denotes the partition block, and A_s denotes the partition block vector fusion value. Figure 1 shows the three-dimensional simulation design architecture, according to which the purpose of human-computer interaction is achieved in the process of cable fault point localization. Using the results of the above vector value calculation, the three-dimensional simulation model of power cables is constructed in the embedded control platform. In order to strengthen the authenticity of the three-dimensional simulation model of power cables, it is necessary to enhance the efficiency of the screen rendering, and the comprehensive consideration of a variety of modeling software such as 3DStudio, Lightware3D, Multigen Creator and other modeling software is analyzed to analyze the rendering efficiency of the software modeling, and then the Multigen Creator software is selected to meet the three-dimensional simulation design requirements. After analyzing the rendering efficiency of the software modeling. The application of Multigen Creator software combined with the three-dimensional feature vectors of the power cable model, the output vector modeling results, to meet the virtual real-time and efficient requirements of virtual three-dimensional model building.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.35-55

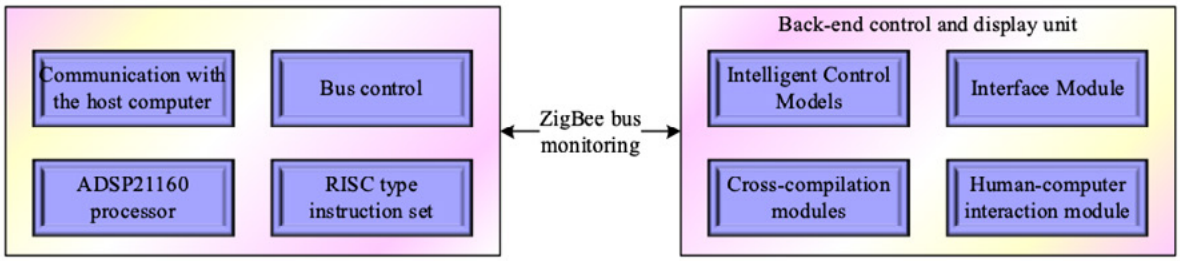


Figure 1 3D simulation design architecture

2.2. EXTRACTION OF THREE-DIMENSIONAL FEATURES OF POWER CABLES

Aiming at the three-dimensional simulation model of power cables, the analysis yields a three-dimensional feature distribution model for cable fault location:

$$\mu(n)=\begin{cases}\beta_1\left[1-\exp\left(-\alpha_1\left|e^n\right|^2\right)\right], & E_{[}\left|e^n\right|^2\geq K\\ \beta_2\left[1-\exp\left(-\alpha_2\left|e^n\right|^2\right)\right], & \text{Other}\end{cases} \tag{3}$$

Where K denotes the three-dimensional feature distribution information of power cable, μ denotes the three-dimensional feature distribution model, denotes the exponential function, e denotes the natural constant, E denotes the three-dimensional information reconstruction error, α_1 and $,\alpha_2,\beta_1$ β_2 ,denote the statistical eigenvalues. The specific constraints are:

$$\begin{cases}\alpha_1\geq 0,\alpha_2\geq 0\\ 0\leq\beta_1\leq\frac{1}{\lambda_{\max}},0\leq\beta_2\leq\frac{1}{\lambda_{\max}}\end{cases} \tag{4}$$

Where $\lambda_{\max}$ denotes the maximum feature screening amount. Combined with the principal component analysis algorithm, the database centered on 3D visual reconstruction data is constructed to ensure that the reconstruction results have a small error. The high-precision reconstruction of the 3D simulation model is completed by the above calculation results, and the feature information within the reconstructed model is collected using the information interaction mode, and the feature component output results are:

$$f=\sum_{i=1}^{n}\partial_i K+b \tag{5}$$

Where f denotes the feature component of the information interaction output, i denotes the spatial region, ∂_i denotes the modal value of the pixel point distribution

https://doi.org/10.17993/3ctecno.2024.v13n1e45.35-55

within the spatial *b* region, and denotes the dimensionality of the 3D visual reconstruction. The blurred center of mass of the view image is determined using 3D imaging technology, and the probability density function calculation formula for feature extraction is derived through 3D visual simulation:

$$\sigma(x,y \mid \mu_K) = \prod \theta_K \frac{1}{\sqrt{2}} \exp\left\{ -\frac{(x-\mu_K)^2}{2} \right\} \tag{6}$$

Where σ denotes the probability density function, μ denotes the skew angle of 3D feature extraction, and θ denotes the feature point edge roundness function. Using the continuous difference reconstruction algorithm, the power cable fault region is made to exhibit smoothness, and the three-dimensional smooth region is denoted as:

$$P = \sum \left(f_p\right) + \sum V_{p,q}\left(f_p, f_q\right) \tag{7}$$

Where *P* denotes the 3D smoothing region, *V* denotes the continuous difference function, and *(p,q)* denotes the coordinates of the fault localization region after 3D reconstruction of the power cable.

In order to ensure the accuracy of the three-dimensional feature extraction results, the image is divided into multiple sub-pixel blocks, and the feature extraction optimization formula is derived by superposition calculation:

$$f = \sum_{i=1}^{n} \left(\partial_i - \bar{\partial}_i\right) K + b \tag{8}$$

Formula, $\bar{\partial}_i$ denotes the modulus of the 3D eigenvolume at sub-pixel points.

2.3. CABLE FAULT POINT LOCALIZATION BASED ON WAVELET TRANSFORM AND GA-BP NEURAL NETWORK

The proposed wavelet transforms and GA-BP based power cable fault location model is shown in Fig. 2, which mainly contains wavelet transform and mode maxima search module, GA optimization module, BP neural training and prediction module.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.35-55

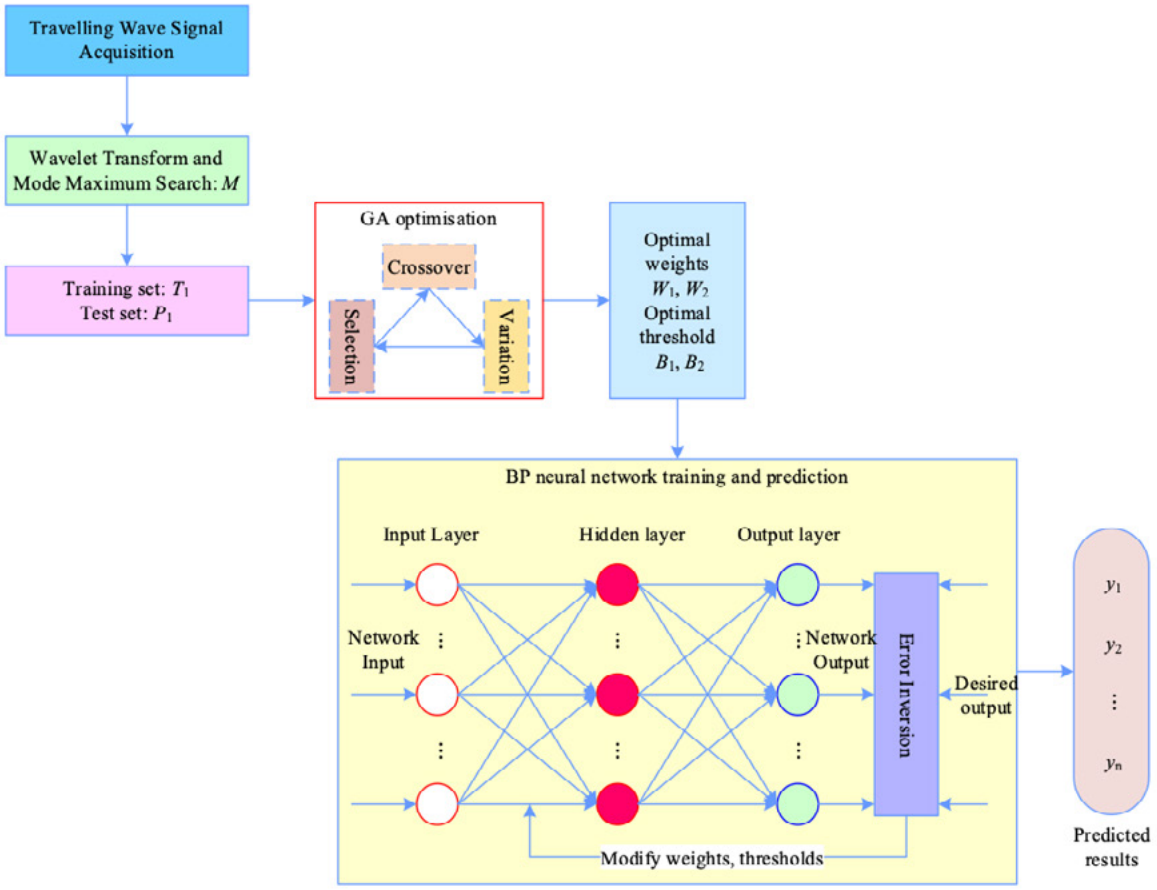


Figure 2 Fault location model of power cable based on wavelet transform and GA-BP

The flowchart of the wavelet transforms and GA-BP based power cable fault point localization algorithm is shown in Fig. 3, and the detailed steps are as follows.

The fault traveling wave mode maximum value data $M = \{(M_{11}, M_{12}), (M_{21}, M_{22}), \cdots (M_{n1}, M_{n2})\}$ collected on both sides of the cable is used as the feature value, and the set fault distance $X = \{X_1, X_2, \cdots X_3\}$ is used as the label value, which is normalized, and then the processed data is divided into the training data T_1 and the test data P_1, which are inputted into the GA-BP model.

Determine the topology of the BP neural network, the number of input layers inputnum, the number of hidden layers hiddennum and the number of output layers outputnum and initialize the BP neural network weights W1,W2 , threshold length B1, B2 .

GA optimizes the BP neural network module, determines the chromosome length and calculates the fitness by encoding the initial weight threshold of the BP neural network in real numbers, and constantly updates the fitness by selecting the function, crossover function and variation function.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.35-55

Determine whether the end condition is satisfied, the number of termination iterations or a set threshold is reached. If it is not satisfied revert to the GA optimization step, if it is satisfied, the optimal weights and thresholds obtained are given to the BP neural network to complete the optimization purpose.

Retrain the BP neural network, calculate the error, and determine whether the end conditions are satisfied, if not, update the weights and thresholds. If the set error range is satisfied, simulation prediction is performed.

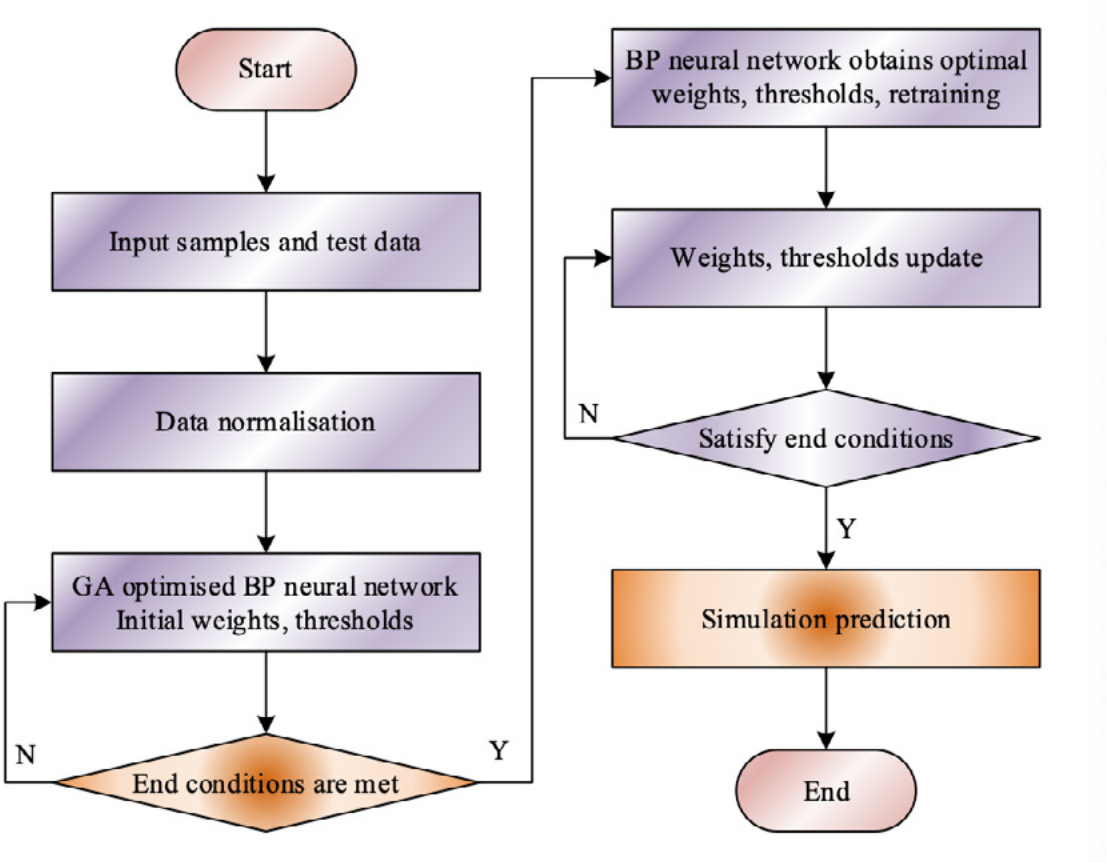


Figure 3 Power cable fault location flow chart

2.3.1. WAVELET TRANSFORM AND MODE MAXIMA SEARCH MODULE

When a power cable fault occurs, the transient traveling wave containing fault information will propagate in the cable, but because of its non-smooth and high frequency characteristics lead to the traveling wave head is difficult to identify, so the wavelet transform is used to analyze the good time-frequency domain characteristics and noise canceling ability to analyze the transient traveling wave, and the mode maxima theory is used to analyze the time of arrival of the initial head of faulty traveling wave.

(1) wavelet transform

https://doi.org/10.17993/3ctecno.2024.v13n1e45.35-55

Assuming that the Fourier transform $\psi(\omega)$ of $\psi(t)$ satisfies the admissibility condition, $\psi(t)$ can be called a mother wavelet:

$$C_\psi = \int_R \frac{|\psi(\omega)|^2}{\omega} d\omega < \infty \tag{9}$$

And for a continuous signal *g(t)* , its corresponding continuous wavelet transform (CWT) is:

$$CWT(a,b) = \frac{1}{\sqrt{a}} \int_R g(t) \overline{\psi\left(\frac{t-b}{a}\right)} dt \tag{10}$$

Where $\overline{\psi(t)}$ denotes the complex conjugate function of $\psi(t) \cdot a$ denotes the scale factor of the frequency-dependent wavelet function. *b* denotes the time-dependent displacement factor. However, in the process of practical application, it is necessary to discretize the continuous wavelet, assuming that $a = a_0^j$, $b = k a_0^j x b_0$, j, and $k \in Z$ can be obtained as the discrete wavelet transform (DWT):

$$DWT(j,k) = a_0^{-j/2} \psi\left(a_0^{-j} t - k b_0\right) \tag{11}$$

(2) Db6 Wavelet basis functions with mode maxima

Suppose there exists a simple step signal with the following functional expression:

$$S(t) = \begin{cases} 0 & 1 \le t \le 300 \\ 10 & 301 \le t \le 600 \end{cases} \tag{12}$$

This is an obvious signal with a mutation point, and according to the singularity principle, a singularity point is generated at the point.

The waveforms are more concentrated at the d1 scale and the moment of the step signal change can be clearly distinguished at 300s, so the modal maxima at the d1 scale will be collected as the characteristic vectors of the dataset. The method of determining the mode maxima is shown below.

Under a certain decomposition scale a_0, if there exists a point (a_0, b_0) such that $\frac{\partial\left(W_\psi f\right)(a_0,b_0)}{\partial b} = 0$, then ($a_0$, b_0) is called the local mode maxima. If $\left|\left(W_\psi f\right)(a_0,b)\right| \le \left|\left(W_\psi f\right)(a_0,b_0)\right|$ exists in any field of B0, then (a_0, b_0) is the mode maxima point of the wavelet transform coefficients. Therefore, in this paper, we use GA-BP neural network to train the mapping relationship between mode maxima

https://doi.org/10.17993/3ctecno.2024.v13n1e45.35-55

and fault location by taking the mode maxima data under the acquisition scale as the feature quantity and the distance to the fault is set as the label.

2.3.2. GA OPTIMIZATION MODULE

The use of traditional BP neural network for fault location in cables may cause the algorithm to fall into the local optimum or converge slowly because of the random initialization parameters of the network, in order to improve the convergence speed of the BP algorithm and the ability of global optimization searching, therefore, the introduction of GA to optimize the weights and thresholds of the BP neural network, the specific steps are as follows.

(1) Real number coding. The initial weights and thresholds of the BP neural network are encoded in real numbers with faster operation speed, utilizing the following linear transformation:

$$x(j) = p(j) + a(j)(q(j) - p(j)(j = 1, 2, \cdots, n) \qquad (13)$$

The initial change interval for the $[p(j), q(j)]$ interval of the *j* optimization variable *x(j)* corresponds to the $[0, 1]$ interval on the real number $a(j), a(j)$ in the genetic algorithm is expressed as genes, all the variables corresponding to the genes sequentially linked together to form a coded form of the solution to the problem $\left(a_1, a_2, \cdots a_k\right)$ is then called the individual or chromosome.

(2) Calculate the individual fitness. The initial weights and thresholds of the BP neural network are obtained according to the individual, and the output of the system is predicted after training the BP neural network with the training data, and the absolute value of the error between the predicted output and the desired output of the individual is taken as the value of the individual fitness *F*:

$$F = K\left(\sum_{i=1}^{n} abs\left(y_i - o_i\right)\right) \qquad (14)$$

n is the number of output nodes of the network. O_i is the desired output of the *i* th node of the *BP* neural network and y_i is the actual output of the *i* th node. *K* is the coefficient.

(3) Selection operation. Genetic algorithm selection operation betting roulette selection method, tournament selection method and geometric planning sorting selection and other methods, when selecting the roulette method, that is, based on the fitness ratio selection strategy, the selection probability P_i of each individual *i* is:

$$f_i = \frac{K}{F_i}, p_i = \frac{f_i}{\sum_{j=1}^{N} f_j} \qquad (15)$$

https://doi.org/10.17993/3ctecno.2024.v13n1e45.35-55

F_i is the fitness value of individual *i.* Since smaller fitness values are better, the inverse of fitness is taken before individual selection. *K* is the coefficient. *N* is the number of individuals in the population.

(4) Crossover operation. If the real number crossover method is used, the crossover operation of the *k* th chromosome a_k and the *l* th chromosome a_1 in the *j* dimension is as follows:

$$\begin{aligned} a_{kj} &= a_{kj}(1-b) + a_{lj}b \\ a_{lj} &= a_{lj}(1-b) + a_{kj}b \end{aligned} \tag{16}$$

Where *b* is a random number between .

(5) Mutation operation. The *j* th gene a_{ij} of the *i* th individual was selected for mutation:

$$a_{ij} = \begin{cases} a_{ij} + \left(a_{ij} - a_{\max}\right) * f(g) & r > 0.5 \\ a_{ij} + \left(a_{\min} - a_{ij}\right) * f(g) & r \le 0.5 \end{cases} \tag{17}$$

$$f(g) = r_2\left(1 - \frac{g}{G\max}\right)^2 \tag{18}$$

a_{max} is the upper bound of the gene a_{ij}. A_{min} is the lower bound of the gene. r_2 is a random number. *g* is the current iteration number. G_{max} is the maximum number of evolutions. *r* is a random number between $[0, 1]$.

(6) Calculate the individual fitness and determine whether the set end condition of minimum error is satisfied, if not, the operation of steps 3 to 5 is repeated. If the end condition is satisfied, the optimized weights and thresholds of the BP neural network are given.

2.3.3. BP NEURAL NETWORK TRAINING AND PREDICTION MODULE

The optimal weights and thresholds obtained through the GA optimization module are given to the BP neural network, and then the BP neural network is re-trained and predicted in the steps shown below.

(1) Select the transfer function and training function. Set the implicit layer transfer function, the output layer transfer function as and select the training function.

(2) Set other parameters of the BP neural network. The number of iterations epochs, learning rate learningrate, training target minimum error goal.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.35-55

(3) Start training the BP neural network, and determine whether to meet the end conditions of setting the minimum error, if not, the weights and thresholds need to be re-updated. If satisfied, simulation prediction is carried out.

3. EXPERIMENTAL ANALYSIS OF FAULT POINT LOCALIZATION ALGORITHMS FOR POWER CABLES

3.1. SIMULATION ANALYSIS UNDER DIFFERENT FAULT INITIAL PHASE ANGLE CONDITIONS

This paper builds a power cable line with a total length of 70km, and the simulation model diagram of the power cable line is shown in Figure 4. Its voltage level is 110kV, 50Hz dual-source power supply system, single-phase ground fault occurs in phase A at 0.016s, the transition resistance is 100Ω, and the total simulation time is 0.05s. The total length of the hybrid line is 70km, of which the length of the overhead line is 50km and the length of the cable line is 20km.

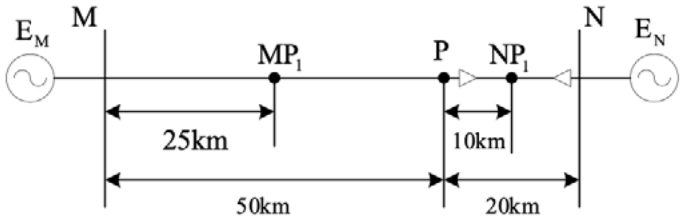
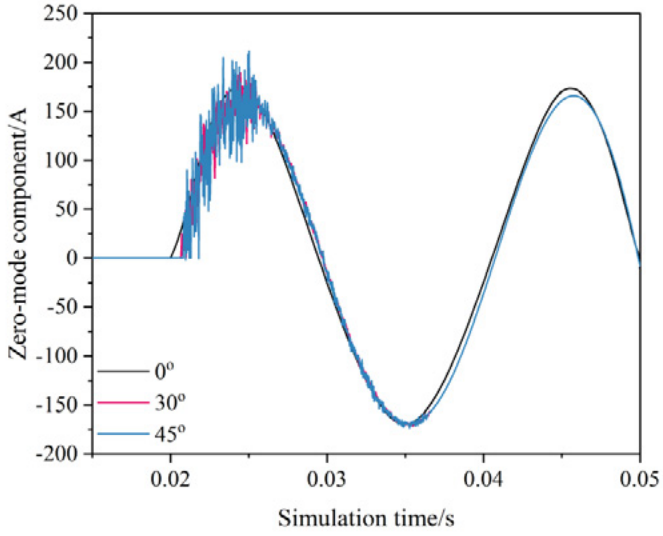


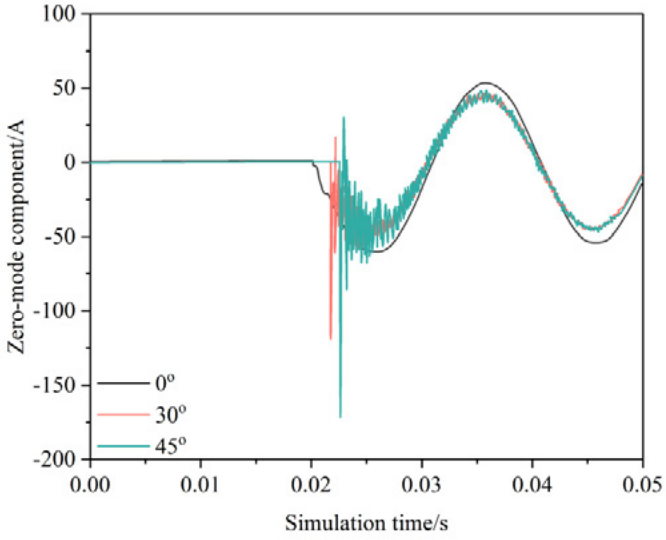
Figure 4 Schematic diagram of power cable line model

Faults may occur at any time in the power system, so the size of the initial phase angle of the fault has randomness. In order to verify that under different fault initial phase angle conditions, the power cable fault point localization algorithm in this paper has strong adaptability. In this subsection, different initial phase angles of faults at the same location are simulated and analyzed, and the values of initial phase angles of faults are set to be 0°, 30°, 45°, and the fault occurs at the 6km overhead line section as an example. Figure 5 shows the zero-mode component waveform curves of different faults, where (a)~(c) are the zero-mode component waveforms of the traveling wave of the faults on the M side of the bus, the cable junction point P and the N side of the bus, respectively. Although the zero-mode components at the same point may be slightly different for different initial phase angles of the fault, the general trend is the same, and the overall error does not exceed 5%. Therefore, for the case of short circuits with different initial phase angles of faults, the three-dimensional simulation and localization method based on wavelet transform and GA-BP power cable fault point proposed in this paper is still applicable to fault point localization.

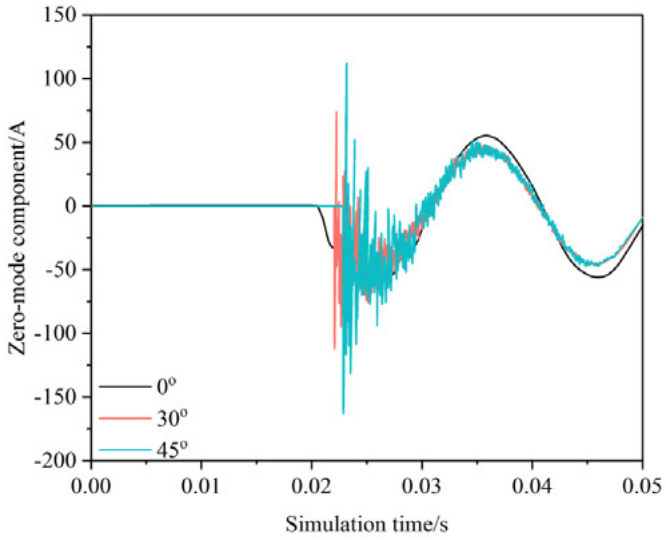
https://doi.org/10.17993/3ctecno.2024.v13n1e45.35-55



(a) Waveform diagram of zero mode component at point M



(b) Waveform diagram of zero mode component at point P



(c) Waveform diagram of zero mode component at point N

Figure 5. Waveform diagram of zero mode component of different faults

https://doi.org/10.17993/3ctecno.2024.v13n1e45.35-55

Different fault location fault initial phase angle of 0 °, 30 °, 45 °, power cable fault point localization simulation results shown in Table 1, in different fault initial phase angle, the algorithm proposed in this paper can accurately determine the fault section, and determine the results of the fault section with the occurrence of faults are consistent with the positioning results of the relative error in the majority of the 1% or less, to meet the demand for three-dimensional simulation of fault point localization accuracy of power cables.

Table 1 Short-circuit fault location results for different initial phase angles

Fault distance /km	Initial fault phase Angle/°	Fault location/km	Ranging error/%
6	0	5.961	0.65
	30	6.047	0.78
	45	5.957	0.72
11	0	11.102	0.93
	30	11.105	0.95
	45	11.096	0.87
34	0	34.021	0.06
	30	34.024	0.07
	45	33.948	0.15
54	0	54.014	0.03
	30	54.021	0.04
	45	53.984	0.03
63	0	63.038	0.06
	30	62.896	0.17
	45	63.026	0.04

3.2. COMPARATIVE ERROR ANALYSIS BETWEEN DIFFERENT ALGORITHMS

Comparing this paper's algorithm with EMD, Wavelet decomposition, SKT algorithm, Table 2 shows the error analysis of four algorithms for fault localization, and the data show that the error of the fault localization method proposed in this paper is lower than that of the other two fault localization errors. In single-phase ground fault, the average error of this paper's algorithm is 0.112km, and the average errors of EMD decomposition, traditional wavelet, and synchronous squeezed SWT decomposition are 0.32km, 0.282km, and 0.241km, respectively. in two-phase grounded short-circuit fault, the average error of this paper's algorithm is 0.126km, the average error of EMD

https://doi.org/10.17993/3ctecno.2024.v13n1e45.35-55

decomposition, traditional wavelet, and synchronous squeezed SWT decomposition are 0.304km, respectively. 0.304km, 0.272km, 0.227km respectively. the proposed method based on wavelet transform and GA-BP power cable fault point 3D simulation localization is better than the traditional wavelet decomposition method, EMD decomposition method, and synchronous squeezing SWT algorithm.

Table 2 Fault location error analysis of the four algorithms

	A Ground connection is faulty	AB Two-phase ground connection fault						
Fault distance /km	EMD	Wavelet decomposition	SKT	Textual algorithm	EMD	Wavelet decomposition	SKT	Textual algorithm
10	486	469	395	250	427	402	337	250
20	415	375	320	199	389	351	297	190
30	424	380	328	192	410	367	314	222
40	351	309	273	109	357	317	271	161
50	318	290	249	126	310	286	222	140
60	290	254	222	96	286	252	216	113
70	240	198	183	44	253	221	195	67
80	294	229	159	78	241	230	150	83
90	205	173	153	17	193	156	140	22
100	173	142	129	4	170	139	128	10

In order to further clarify the advantages and disadvantages of the method proposed in this paper among different algorithms under different fault types, two-phase short circuit and three-phase short circuit fault simulation are set up respectively, and the fault point localization errors are analyzed as shown in Table 3, and the fault localization errors of the method proposed in this paper, which is based on the wavelet transform and the three-dimensional simulation localization of the fault point of GA-BP power cables, are 0.109km and 0.12km, and are smaller than the fault localization errors of EMD decomposition method, traditional wavelet decomposition, and synchronous squeezed SWT algorithm. 0.109km and 0.12km, which are smaller than the fault localization errors of EMD decomposition method, traditional wavelet decomposition, and synchronous squeezing SWT algorithm, so the method proposed in this paper is suitable for fault point localization of power cables with different fault types.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.35-55

Table 3 Fault location error analysis of four algorithms

	AB Two-phase short-circuit fault	ABC Three-phase short circuit fault						
Fault distance /km	EMD	Wavelet decomposition	SKT	Textual algorithm	EMD	Wavelet decomposition	SKT	Textual algorithm
10	455	444	370	238	408	381	320	239
20	396	350	305	190	373	330	285	214
30	405	359	308	180	394	352	298	213
40	336	297	288	124	342	302	260	155
50	293	280	238	122	294	279	213	134
60	284	252	222	102	284	254	217	113
70	222	194	179	46	241	214	183	68
80	223	218	168	56	227	220	148	30
90	194	168	146	21	182	152	138	20
100	168	136	122	8	162	134	123	11

3.3. ERROR ANALYSIS IN A NOISY BACKGROUND

Under two background noise environments, white noise and colored noise, the signal-to-noise ratios are taken as 0dB, -10dB, -15dB, -20dB, and 300 Monte Carlo experiments are carried out for the generalized correlation, quadratic correlation, and this paper's algorithms, respectively, and the statistical results of the simulation of time delay estimation are obtained as shown in Table 4. The results use the time delay estimation mean and root mean square error (is the estimated value and *a* is the actual time delay value, which is 300) as the performance evaluation criteria. The "-" horizontal line indicates that the value is no longer of reference value because the estimation error is too large. Analyzing the data in the table, it can be seen that at not too low signal-to-noise ratios, all three types of algorithms can give accurate estimation results regardless of whether the background noise is white or colored noise. However, in the case of low signal-to-noise ratio, when the white noise signal-to-noise ratio is -20dB, only the error of this algorithm is still within the controllable range, at this time, the mean square error is 0.057, while the white noise signal-to-noise ratio is -15dB, the generalized correlation and the quadratic correlation error is no longer a reference value, at this time, the mean square error of this algorithm is 0.076. The algorithm of this paper has a certain inhibition effect on the noise, and its effect is more desirable than that of the other two methods. The effect is more ideal than the other two methods. Under the same signal-to-noise ratio, for the three delay

https://doi.org/10.17993/3ctecno.2024.v13n1e45.35-55

estimation algorithms, the estimation performance in the case of white Gaussian background noise is significantly better than that in the case of colored Gaussian background noise.

Table 4 Simulation statistical results of delay estimation

SNR		White noise			Colored noise		
		Generalized correlation	Secondary correlation	Textual algorithm	Generalized correlation	Secondary correlation	Textual algorithm
0dB	Mean	300.04	299.96	300.02	299.95	300.05	300.02
	Mean square error	13	11	12	34	49	25
-10dB	Mean	299.89	300.07	299.96	299.87	300.116	300.05
	Mean square error	42	42	22	214	198	53
-15dB	Mean	301.394	298.83	300.28	—	—	299.18
	Mean square error	85	44	34	—	—	76
-20dB	Mean	—	—	298.18	—	—	298.14
	Mean square error	—	—	57	—	—	122

4. CONCLUSION

In this study, a new 3D simulation localization method for power cable fault points is developed by combining wavelet transform and neural network techniques. Experimental results show that the method shows high accuracy and stability in different types of fault localization. Specifically, in the case of single-phase ground fault and two-phase grounded short-circuit fault, the average localization errors of the proposed method are 0.112 km and 0.126 km, respectively, which are much lower than those of the traditional EMD decomposition, traditional wavelet, and synchronous squeezed SWT decomposition methods. In addition, the proposed method shows good adaptability under different fault initial phase angles, and the error is controlled within 1%, which proves its effectiveness under different fault scenarios.

The method also shows excellent performance in noisy environments. Especially in the case of low signal-to-noise ratio, the method proposed in this paper shows better noise immunity and maintains lower localization error compared with traditional methods. This feature makes the method more valuable in practical power system operation and maintenance, especially in complex noise environments.

The proposed three-dimensional simulation localization technique for power cable faults based on wavelet transform and GA-BP neural network not only achieves significant results in improving the localization accuracy, but also significantly

https://doi.org/10.17993/3ctecno.2024.v13n1e45.35-55

improves the ability of anti-noise interference. The development of this technology will bring important practical application value to the fault detection and maintenance of the power system, and help to improve the reliability and stability of the power system.

REFERENCES

(1) Nguyen, T. H., & Lee, D. C. (2017). Protection of the MMCs of HVDC transmission systems against DC short-circuit faults. Journal of Power Electronics.

(2) Lee, S. J., Kang, S. Y., Park, M., Won, D. Y., & Yang, H. S. (2020). Performance analysis of real-scale 23 kV/60 MVA class tri-axial HTS power cable for real-grid application in Korea. Energies, 13(8), 2053.

(3) Chen, Z., Kleijn, R., & Lin, H. X. (2022). Metal requirements for building electrical grid systems of global wind power and utility-scale solar photovoltaic until 2050. Environmental Science & Technology.

(4) Khalil, A. R., Howard, I. M., Forbes, G. L., Sultan, I. A., & McKee, K. K. (2017). Design and installation of subsea cable, pipeline and umbilical crossing interfaces. Engineering Failure Analysis.

(5) Reda, A., Abu-Siada, A., Howard, I. M., & McKee, K. K. (2018). A testing platform for subsea power cable deployment. Engineering Failure Analysis, 96, 142-157.

(6) Ulku, I., & Alabas-Uslu, C. (2020). Optimization of cable layout designs for large offshore wind farms. International Journal of Energy Research, 44(8).

(7) Kwon, G.-Y., Lee, C.-K., Shin, & Yong-June. (2019). Diagnosis of shielded cable faults via regression-based reflectometry. IEEE Transactions on Industrial Electronics.

(8) Brothers, S. (2018). Finding fault with cabling connections. Electrical Engineering (OCT.), 8-8.

(9) Suyang, Wang, Jianmei, Chuang, Zhang, & Yang, et al. (2019). Synchronous online diagnosis of multiple cable intermittent faults based on chaotic spread spectrum sequence. IEEE Transactions on Industrial Electronics, 66(4).

(10) Torwelle, P., Bertinato, A., Raison, B., Le, T. D., & Petit, M. (2022). Fault current calculation in MTDC grids considering MMC blocking. Electric Power Systems Research (Jun.), 207.

(11) Li, G., Chen, J., Li, H., Hu, L., Zhou, W., & Zhou, C., et al. (2022). Diagnosis and location of power cable faults based on characteristic frequencies of impedance spectroscopy. Energies, 15.

(12) Li, M., Bu, J., Song, Y., Pu, Z., & Xie, C. (2021). A novel fault location method for power cables based on an unsupervised learning algorithm. Energies, 14(4), 1164.

(13) Sommervogel, L. (2020). Various models for faults in transmission lines and their detection using time domain reflectometry. Progress in Electromagnetics Research C, 103, 123-135.

(14) Tian, Y., Zhao, Q., Zhang, Z., Li, L., & Crossley, P. (2018). Current-phase-comparison-based pilot protection for normally closed-loop distribution network with underground cable. International Transactions on Electrical Energy Systems, 28(9), e2733.1-e2733.19.

(15) Li, Z., Yang, H., Yang, F., Tan, T., Lu, X., & Tian, J. (2022). An infrared image based state evaluation method for cable incipient faults. Electric Power Systems Research.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.35-55

(16) Siodla, K., Rakowska, A., & Noske, S. (2021). The proposal of a new tool for condition assessment of medium voltage power cable lines. Energies, 14.

(17) Samet, H., Tajdinian, M., Khaleghian, S., & Ghanbari, T. (2021). A statistical-based criterion for incipient fault detection in underground power cables established on voltage waveform characteristics. Electric Power Systems Research, 197, 107303.

(18) Li, C., Chen, J., Zhang, W., Hu, L., Cao, J., & Liu, J., et al. (2021). Influence of arc size on the ignition and flame propagation of cable fire. Energies, 14.

(19) Lacerda, V. A., Monaro, R. M., & Coury, D. V. (2021). Fault distance estimation in multiterminal HVDC systems using the Lomb-Scargle periodogram. Electric Power Systems Research, 196, 107251.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.35-55

/03/

APPLICATIONS AND PROSPECTS OF ARTIFICIAL INTELLIGENCE IN LINGUISTIC RESEARCH

Shaohua Jiang

- School of Humanities, Fujian University of Technology, Fuzhou, Fujian, 350118, China
- Krirk University, Bangkok, 10220, Thailand
- sophia_FP@126.com

Zheng Chen*

- Concord University College, Fujian Normal University, Fuzhou, Fujian, 350000, China

Reception: 2 January 2024 | **Acceptance**: 22 January 2024 | **Publication**: 19 February 2024

Suggested citation:

Jiang, S. and Chen, Z. (2024). **Applications and Prospects of Artificial Intelligence in Linguistic Research.** *3C Tecnología. Glosas de innovación aplicada a la pyme 13(1)*, 57-76.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

ABSTRACT

In modern linguistic research, the application of Artificial Intelligence has led the field and provided powerful tools and prospects for linguists. LSTM is used for extracting character features, joint vector representation and constructing text generation models and generating natural language text. LSTM is involved in the design of speech recognition network to process the input speech signals for generators and discriminators to improve the accuracy of speech recognition. By continuously optimizing the training objectives, the translation system will more accurately translate text from one language to another, thus facilitating cross-cultural communication. Through the application of artificial intelligence, the F1 value has been improved by 3.9% compared with the previous value, and the cumulative variance contribution rate of the five factors is more than 60%, with all subloadings reaching 0.4 or more. Artificial intelligence will promote the development of the field of linguistics, improve research efficiency and accuracy, and promote the innovation of language technology.

KEYWORDS

Artificial intelligence; LSTM; joint vector; speech recognition; F1 value

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

INDEX

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

1. INTRODUCTION

Perhaps because of the wider range of application scenarios, or because of the input from established technology companies such as Google, Baidu, and Tech Data, speech recognition technology is often the concept of artificial intelligence technology that comes to mind [1]. Speech recognition technology is indeed used in a large number of specific scenarios in the language learning process [2]. However, if the role and value of speech recognition technology is not understood accurately enough, it would be biased to even expect that relying on speech recognition technology can solve the challenges of language learning intelligence [3]. The key to speech recognition is recognition. No matter how high the recognition degree and accuracy, the ultimate goal is to recognize what the learner has said and display the specific text. This function and process, however, is not strictly pedagogical [4]. That is to say, the result of the recognition is simply a textual result, and it does not yet address the really important matter of how to improve the quality of what is being said. For language learners, the required results and value are much greater than for conventional translation tools [5]. This means that even if 100% recognition accuracy can be achieved, at best it will enable fast and accurate translation or presentation, and will not provide learners with methods and suggestions for learning and improvement [6].

In this paper, LSTM network is used to extract character level features from text data to capture important information and patterns in the text. LSTM is used to create joint vector representations and the structure and functionality of LSTM units are described. LSTM network is used to design speech recognition system to recognize and understand the speech content in the speech signal. Generators and discriminators are used in speech signal processing to improve the recognition accuracy and STM network is used to achieve the training objectives to improve the performance and effectiveness of speech signal processing. The present generative model is created to be used for tasks such as natural language generation. In addition, the innovation of this paper is the use of LSTM networks to create a text generation model, which is potentially valuable for natural language generation tasks. This model can be used to generate natural language text such as articles, comments, or conversations, which is expected to have a wide range of applications in the field of automated writing and chatbots.

2. LITERATURE REVIEW

Rasulova, Z emphasizes the importance of studying the processes and mechanisms of translation, referring to the methodological and psychologist's view that the issue of translation skills and their formation has an important place in translation theory and practice. It is shown that when studying translation, it is important to focus not only on the outcome of the translation, but also to delve into the skills and strategies of the translator and how these skills are formed [7]. Braithwaite, B suggests that there is a rapidly growing scholarly interest in sign languages of the

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

Global South, especially those emerging in small sign language communities. Neutral theoretical constructs about these communities and sign languages may be too abstract and may lead to a tendency to exoticize and objectify research by ignoring the actual needs and concerns of community members [8]. Bafoevna, N. D et al. point out that theological linguistics emerged partly due to the fact that religions have an important place in the social consciousness and are an integral part of any culture. Therefore, if the religious factor is ignored, the study of language will appear incomplete and may even become unfeasible in some cases [9]. Mizumoto, A et al. point out that in the field of corpus linguistics, the application of RS/MA has been very limited and confined to very few subfields. Given that corpus linguistics covers a wide range of issues, meta-analysis is considered to have great potential as a method for systematically synthesizing research results in the field [10]. Su, H et al. proposed a local grammar approach to the study of non-synchronous discourse behavior in academic texts, aiming to provide a new avenue for the study of non-synchronous academic discourse. The local grammar approach captures the realization patterns of discourse acts at both the lexico-grammatical and discourse semantic levels, which helps to understand how the realization of a particular discourse act varies across time and contexts [11]. Awad Al-Dawoody et al. selected a corpus of 60 randomly selected research articles and used them according to Hyland's classification of metadiscourse markers, using the AntConc.3.2.4 for qualitative and quantitative analysis. It was found that there is a gap between Egyptian and Saudi researchers in the use of different metadiscourse markers [12]. Chen, L et al. analyzed by binary logistic regression based on a corpus that recently published articles were more likely to express surprises triggered by a priori knowledge as compared to earlier published articles. These results can be explained by the fact that surprises are heuristic in nature and also by the pressure of academics in strategically promoting their research directions [13]. Umarova, N. R discusses conceptual terminology which is the most active and controversial terminology in modern linguistics, with a focus on the importance of concepts and their linguisticization in the way that language perceives the world, and expresses the national and cultural characteristics of the language. Cognitive approach is one of the methods of recognizing and explaining natural phenomena related to language through language. Cognitive linguistics is a discipline that studies human cognitive activity. Its main aim is to determine the involvement and share of the language system in the process of recognizing the world [14]. Hamzah, M. H et al. objective was to conduct a linguistic literature review of the aboriginal languages of Malaysia, using a systematic evaluation approach and focusing on the three main aboriginal groups of Peninsular Malaysia. The study covered linguistic subfields such as phonology, morphology, sociolinguistics, syntax, semantics, vocabulary and grammar. Further linguistic research is clearly necessary to protect and preserve these languages [15].

3. APPLICATION OF LSTM IN LINGUISTICS

Artificial Intelligence, and in particular LSTMs, are crucial for understanding and processing natural language. LSTMs are a special type of recurrent neural network

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

especially suited for processing and predicting sequential data. In linguistics, this means being able to efficiently process sequences of words, understand sentence structure, and even entire texts. Language contains complex long-term dependencies, for example the subject of a sentence may influence the verb form at the end of the sentence. LSTM is important because it can capture these long-term dependencies better than traditional RNNs [16]. This is crucial for understanding the meaning of text, for language generation and translation. Another advantage of LSTM is its ability to store and process large amounts of historical information, different languages have different grammatical structures and expression conventions, the flexibility of LSTM makes it a powerful tool for understanding and processing multiple languages.

3.1. APPLICATION OF LSTM IN TEXT ANALYSIS

3.1.1. EXTRACTING CHARACTER FEATURES

In natural language processing, CNNs are often used to extract text features, and some researchers have found that using CNNs to extract character-level features can represent the morphological features of words well [17]. Figure 1 shows the network structure for extracting character features in the model of this paper, for example, suyimen is the Latin Viennese word for I like. In this paper, the character vector dimension is set to 30 and is randomly initialized. The maximum character length of each word is 50, if the maximum length is exceeded, the first 50 letters are intercepted, and if the length is less than 50, Padding is used to make up. The character feature representation vectors of the words are extracted through the convolutional and maximum pooling layers. The size of the convolution kernel is 30 and the length of the convolution kernel is 3.

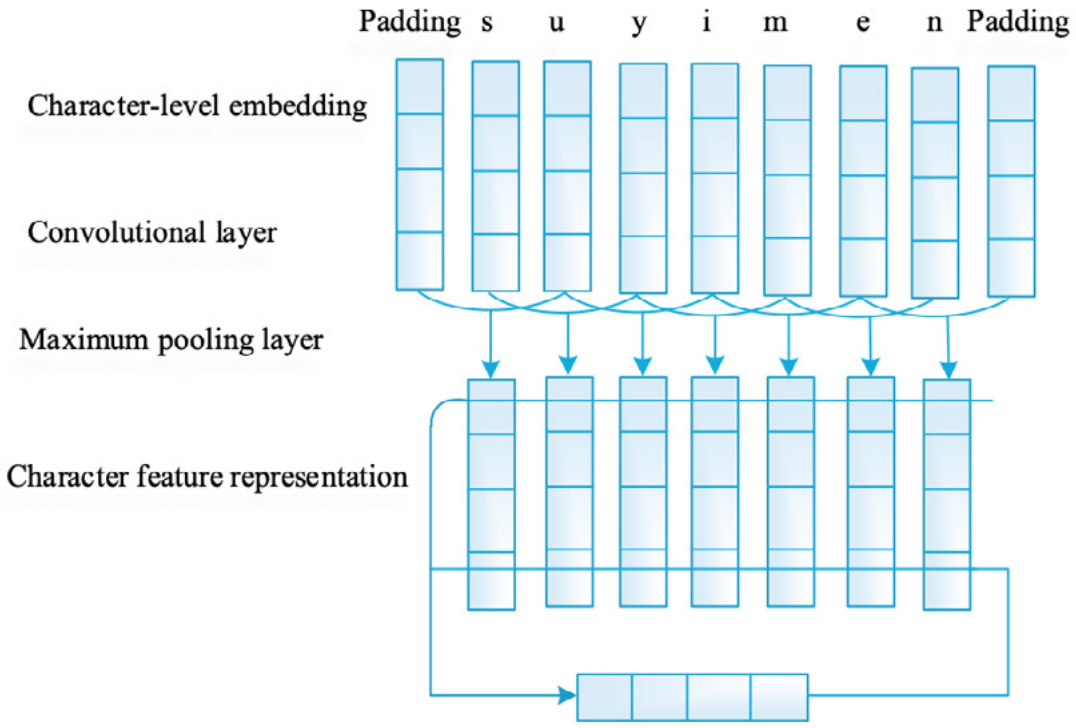


Figure 1 Character feature extraction

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

3.1.2. JOINT VECTOR REPRESENTATION

The cascade of word vectors, character feature vectors, and linguistic feature vectors is used as the input vector representation of the neural network. Assuming that V_{word} denotes the word vector, V_{char} denotes the character feature vector, and V_{fi} denotes the i th linguistic feature vector, the overall input vector can be represented as $V = \left[V_{\text{word}} : V_{\text{char}} : V_{f1} : \cdots : V_{f10}\right]$. The joint feature result is shown in Fig. 2.

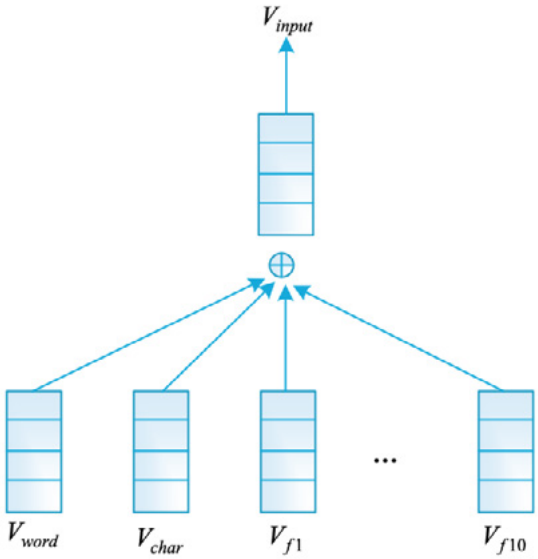


Figure 2 Joint feature representation

3.1.3. LSTM CELL STRUCTURE

Figure 3 shows the basic structure of an LSTM cell, which controls the input and output information through three special gate structures [18]

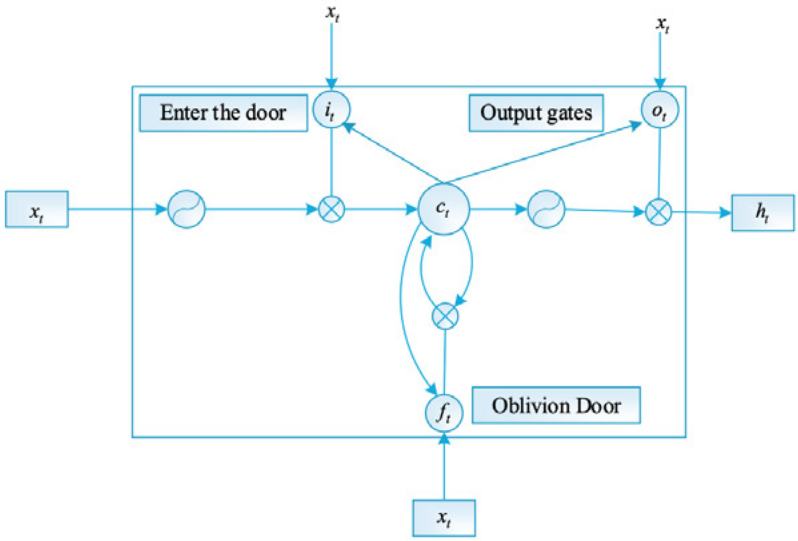


Figure 3 LSTM cell structure

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

$$i_t = \sigma\left(W_{xx}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i\right) \quad (1)$$

$$f_t = \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f\right) \quad (2)$$

$$c_t = f_t c_{t-1} + i_t \tanh\left(W_{xc}x_t + W_{hc}h_{t-1} + b_c\right) \quad (3)$$

$$o_t = \sigma\left(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o\right) \quad (4)$$

$$h_t = o_t \tanh\left(c_t\right) \quad (5)$$

where σ is the Sigmoid activation function, *i* is the input gate, f is the forgetting gate, *c* is the memory cell, *o* is the output gate, *h* is the hidden layer, *tanh* denotes the hyperbolic tangent activation function, W is the weight matrix, e.g., W_{xi} is the weight matrix between the inputs *x* and the input gate, W_{hi} is the weight matrix from the hidden layer to the input gate, and *b* is the bias vector.

3.2. ROLE OF LSTM IN PROCESSING SPEECH SIGNALS

3.2.1. SPEECH RECOGNITION NETWORK DESIGN

Under the assumption that speech and noise are independent of each other, the speech signal and the noise signal are superimposed to form a mixed speech signal Z_t, and then the mixed speech signal is transformed into a two-dimensional time-frequency signal $Y \in {}^{\sim M \times N}$ by a short-time Fourier transform, and then the spectral coefficients of the speech are deduced, where M denotes the time frame corresponding to the speech and N denotes the frequency. The spectrum $\hat{Y}^{j} \in \sim M \times N$ of the speech signal is obtained by the following equation:

$$\hat{Y}^{j} = Y \otimes M^{j} \quad (6)$$

where $\otimes$ denotes the inner product of matrix elements and $M^{j} \in {}^{\sim M \times N}$ is called the time-frequency mask, the time-frequency mask value characterizes the interrelationships between different sources in a mixed signal, such as the target and interfering speakers in speech separation, and the time-frequency mask M^{j} is estimated by using Wiener filtering of the $\alpha-$ power amplitude spectrum, with the following equation:

$$M^{j} = \frac{\left|\hat{Y}^{j}\right|^{\alpha}}{\sum_{j}\left|\hat{Y}^{j}\right|^{\alpha}} \quad (7)$$

where $|*|$ denotes the absolute value of the matrix and α is an index chosen based on the probability distribution of the hypothesized speech, which is taken as 0.5 in this paper.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

The generative adversarial network structure consists of two parts, the generator (G) and the discriminator (D). In this paper, we propose a learnable time-frequency mask generator that introduces a recursive derivation algorithm with a neural network structure and another sparse coding layer for generating the time-frequency mask M^j. In particular, the generator consists of a multilayer recurrent neural network (RNN) and a sparse coding layer, the RNN outputs to the sparse coding layer, and the output of the sparse coding layer is the corresponding time-frequency mask M. The method eliminates the need for subsequent processing such as signal filtering, and there is no need for manually defining the number of layers of the neural network.

The generative adversarial network is shown in Fig. 4, where the generator acts as an encoder ($RNN_{dec)}$ through a layer of bi-directional recurrent neural network, a layer of recurrent neural network as decoding, and a layer of feed-forward neural network as a sparse coding layer. The output of the sparse coding layer is time-frequency masked M^j, which is then multiplied by the matrix elements with the mixed signal to obtain the target speech signal. The discriminator consists of a layer of feed-forward neural network encoder and a layer of feed-forward neural network decoder FNN_{dec} and outputs as values in the interval [0,1]. The generator and the discriminator are iteratively optimized to obtain the optimal time-frequency mask M^j, which is used to estimate the amplitude spectrum of the target speech signal, and then combined with the phase spectrum of the mixed signal to reconstruct the time-domain signal with a short-time Fourier inverse transform.

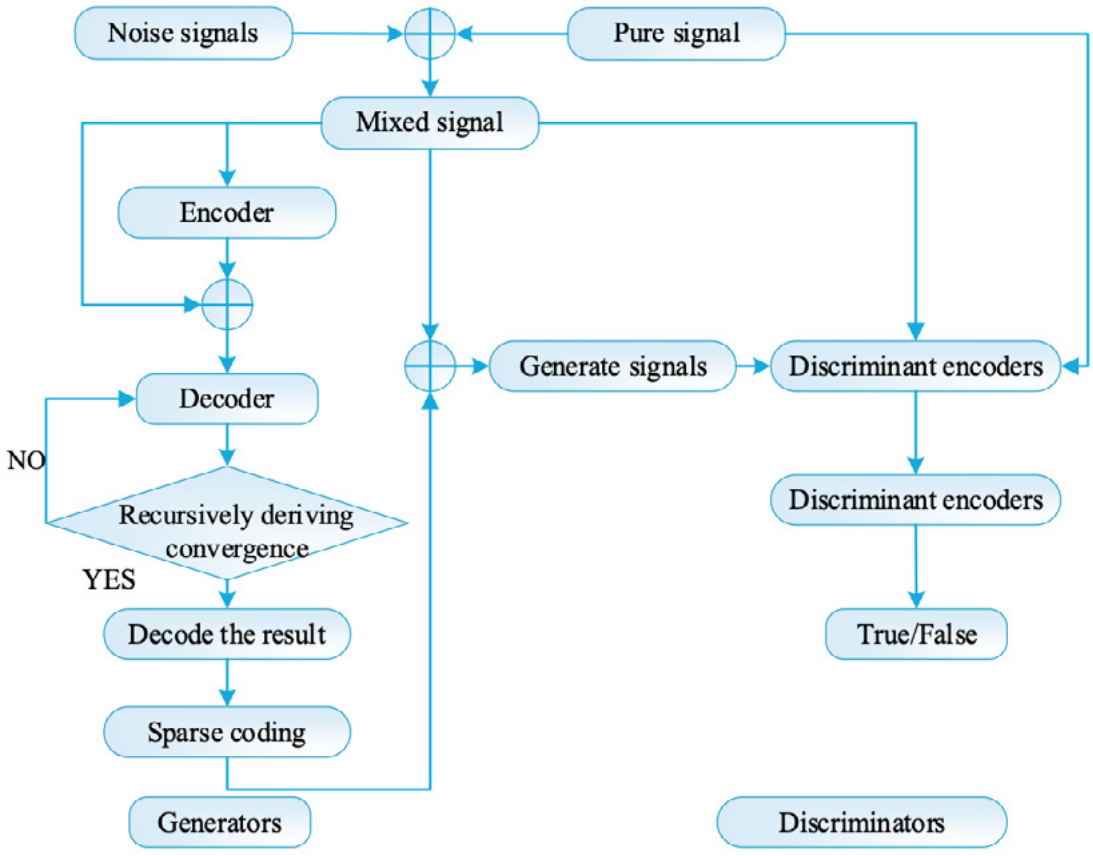


Figure 4 Generates adversarial network structure

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

3.2.2. NPUT SPEECH PROCESSING

Let Z_t be the speech time-domain signal sampled at 44.1kHz and mixed with 0dB signal-to-noise ratio, and Z_t be converted into a two-dimensional time-frequency signal YeRMxN by the short-time Fourier transform (STFT), which is a frame-adding window in accordance with the method of overlapping segmentation, and the window function adopts the Hamming window, with the length of the frame being set to 23ms, and the frame shift being set to 6ms, i.e., each frame contains N=1024 sample points, and there is an overlap of 256 sample points between neighboring time frames. After conversion, the time-frequency signal Y is partitioned into sub-band clusters B with batch data (Batch size) = M/T in a time period T. The remaining frames are padded with values of 0 so that the time dimension expands to T. In order to maintain correlation at the articulation of speech segments, the sub-bands of the latter frame overlap with the former by a time period .L x 2 The amplitude spectrum $\left|Y_{\text{in}}\right| \in^{\sim T\times N}$ of each subband b in Y is used as input to the generator, but considering that the high-frequency portion of the sound is small in energy and relatively insensitive to human hearing, the high-frequency portion of the sound larger than the frequency F is ignored during the training phase, and $\left|Y_{\text{filter}}\right| \in \sim T' \times F$ is used as the input to minimize the number of training parameters and to preserve the most important information of the speech.

3.2.3. GENERATOR

After the input speech is processed to $||Y_{\text{filter}}|$ as the input to the encoder RNN_{enc}, RNN_{enc} using a bi-directional *RNN (Bi-GRU)*, the output of each time frame h_t updated with the iteration of time frames *t* and a residual network is superimposed as:

$$\begin{cases} llh_{enc_t} = h_t + \left|y_{filter_t}\right| \\ \left|Y_{filter}\right| = \left[\left|y_{filter_T}\right|, \ldots, \left|y_{filter_t}\right|, \ldots, \left|y_{filter_1}\right|\right], \left|y_{filter_t}\right| \in^{\sim F} \end{cases} \quad (8)$$

where h_{enc_t} is denoted as the amplitude spectral vector of the output h_t superimposed on $\left|y_{\text{filter}\ t}\right|$ at each time t. The residual network facilitates faster training.

The h_{enc_1} of each time frame $t \in T$ in the merged time period T is denoted as $H_{\text{enc}} \in^{\sim T\times(2\times F)}$, and the overlapping time period *L x 2* is subtracted to obtain the loss $H_{\text{enc}} \in^{\sim T\times(2\times F)}$, where $T' = T - (L \times 2)$, specifically:

$$\tilde{H}_{enc} = \left[h_{enc_{1+L}}, h_{enc_{2+L}}, , , h_{enc_{T-L}}\right] \quad (9)$$

where L is denoted as the time period in which the sub-bands overlap and is merged according to the above equation to obtain $\tilde{H}_{enc}$.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

The introduced recursive derivation algorithm generates temporary variables H^j_{dec} continuously and recursively through the encoder RNN_{dec} until the convergence criterion is satisfied, which is the mean-square error L_{MSE} between neighboring valuations of the temporary variables H^j_{dec} and the threshold is τ_{term} Let the maximum number of iterations be iter, and $\text{func}^j_{\text{dec}}$ denotes the training function of the decoder RNN.

H^j_{dec} After the decoding converges, it is passed to the sparse coding layer that generates the time-frequency mask Mj and shares the sparse coding layer weight parameter for each time period T:

$$\tilde{M}^j = \text{ReLU}\left(H^j{}_{dec} W_{\text{mask}} + b_{\text{mask}}\right) \tag{10}$$

The modified linear unitary function is defined as follows:

$$\text{ReLU}(x) = \begin{cases} x & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x < 0 \end{cases} \tag{11}$$

where ReLU is a segmented linear function that sets all negative values to 0 while positive values remain constant, a setting known as unilateral inhibition, which gives the neurons sparse activation, and the sparsification process is done to improve interference suppression while restoring the frequency dimensions to the target speech signal frequency dimension N. $W_{\text{mask}} \epsilon^{\sim(2\times F)\times N}$ is the weight coefficients matrix for the feed-forward neural network, and $b_{\text{mask}} \in \sim N$ is the corresponding deviation.

The amplitude spectrum $\left|\hat{Y}^j_{\text{filter}}\right| \in \sim T \times N$ of the target speech signal is obtained by the encoder and decoder defined earlier with the following equation:

$$\begin{cases} \left|\hat{Y}^j_{\text{filter}}\right| = \left|Y_{\text{filter}}\right| \otimes \tilde{M}^j \\ \left|Y_{\text{filter}}\right| = \left[\left|y_{in_L}\right|, \ldots, \left|y_{in_{T-L}}\right|\right] \end{cases} \tag{12}$$

where $\left|Y_{\text{filter}}\right|$ is the real input to the generator.

3.2.4. DISCRIMINATORS

The time-frequency mask generated by the generator contains perturbations from the noise signal, and the discriminator plays a role in noise reduction by determining the true and false speech signals, so that the generated signal Й constantly approximates the target speech signal [19-20]. The discriminator consists of the codecs of feedforward neural networks FFN_{enc} and FFN_{dec}. The inputs are divided into

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

two types, one is the speech signal $\hat{Y}^{j}_{filter}$ and the mixed signal Y_{in} generated by the generator, and the other is the real speech signal Y^{j} and the mixed signal Y_{in}, and the inputs are merged into Y^{j}_{concat}. FFN_{enc} and FNN_{dec} share the weight parameters through the time period T. The output of the discriminator is:

$$\text{Re al / Fake } = \text{Re}LU\left(\text{Re}LU\left(Y^{j}_{conat} W_{enc} + b_{enc}\right)W_{dec} + b_{dec}\right) \quad (13)$$

where $W_{enc} \in \sim 2N \times (N/2)$ and $W_{dec} \epsilon^{\sim (N/2)\times 1}$ denote the weight coefficient matrices of feedforward neural networks FFN_{enc} and FFN_{dec}, respectively, with corresponding deviations of $b_{enc} \in^{\sim N/2}$ and $b_{dec} \in^{\sim 1}$.

3.2.5. TRAINING OBJECTIVES

Based on the input of the generator as well as the input of the discriminator, the objective function is adjusted to:

$$\min_{G}\max_{D} V_{CGAN}(G,D) = E\left[\log D\left(Y^{j}, Y_{in}\right)\right] + E\left[\log\left(1 - D\left(G\left(Y_{in}\right), Y_{in}\right)\right)\right] \quad (14)$$

Where Y^{j} is the real speech signal, Y_{in} is the input mixed signal, and $G(z)$ is the generated speech signal. The input to the discriminator is not only the original speech signal r_i and the corresponding signal generated by the generator Y^{j}, but also an additional mixed signal Z_t obtained by short-time Fourier transformation of the time-frequency signal Y_{in}, Y_{in} which constrains the generation direction of the generator. The GAN network enables the generated speech signal not only to approximate the probability distribution of the target speech signal, but also learns the spectral structure of the audio signals in this environment.

3.3. CREATING TEXT GENERATION MODELS USING LSTMS

The traditional machine translation model only associates the learned expression of the last word with the current word to be predicted for translation, whereas the addition of the attention mechanism associates the learned expression of each word at the source language end with the current word to be predicted for translation. Compared with the traditional machine translation, the effect of the model after adding the attention mechanism is significantly improved, two LSTM classification models, one is to use the output of the last moment of the LSTM as a higher level of representation, and the other is to average all the moments of the LSTM output as a higher level of representation. Both of these representations have certain defects, the first one is missing the previous output information, and the other averaging does not reflect the different importance of the output information at each moment. In order to solve this problem, the Attention mechanism is introduced, and the LSTM model is improved in this paper, and the LSTM-Attention model is shown in Figure 5.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

The input sequence in the figure is the vector representation of each word of a text segmentation $x_0, x_1, x_2, \cdots, x_t$, and each input is passed into the LSTM unit to get the output of the corresponding hidden layer $h_0, h_1, h_2, \cdots, h_t$. Here, Attention is introduced in the hidden layer, and the probability distribution value of the attention assigned to each input is calculated $\alpha_0, \alpha_1, \alpha_2, \cdots, \alpha_t$, and the idea is to compute the proportion of the matching score of the output of the hidden layer and the whole text representation vector to the overall score at that moment, $\alpha_i, j \in [0, t]$ the formula is as follows:

$$\alpha_i = \frac{\exp\left(\text{score}\left(\bar{h}, h_i\right)\right)}{\sum_j \exp\left(\text{score}\left(\bar{h}, h_j\right)\right)} \tag{15}$$

where h_t is the output state of the hidden layer at the i nd moment, and $\bar{h}$ can be regarded as a text representation vector one level higher than the word. As mentioned above, both text representation methods have defects, so here $\bar{h}$ is randomly initialized as a parameter to be gradually updated during the training process. $\text{score}\left(\bar{h}, h_i\right)$represents the score of the i th hidden layer output h_i in the text representation vector $\bar{h}$, the larger the score, the greater the attention of the input word in the text at this moment, the formula is as follows:

$$\text{score}\left(\bar{h}, h_i\right) = w^{\mathrm{T}} \tanh\left(W\bar{h} + Uh_i + b\right) \tag{16}$$

Where w, W, U is the weight matrix, b is the bias, and $\tanh$ is the nonlinear activation function. After obtaining the value of the probability distribution of attention at each moment, the feature vector v containing the text information is calculated as follows:

$$v = \sum_{i=0}^{t} \alpha_i h_i \tag{17}$$

Finally, the softmax function is utilized to obtain the prediction category as, which is calculated as follows:

$$y = \text{softmax}\left(W_v v + b_v\right) \tag{18}$$

In this paper, we use the gradient descent method to train the model, and gradually update the parameters of the model by calculating the gradient of the loss function, and finally reach the convergence. In order to make the objective function converge more smoothly, and also to improve the efficiency of the algorithm, only a small number of samples are taken for training each time. The model uses the cross-entropy loss function, and the calculation formula is as follows:

$$H_{y'}(y) = -\sum_i y_i' \log y_i \tag{19}$$

where y_i' is the actual category label value and y_i is the predicted category label value calculated using the softmax function

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

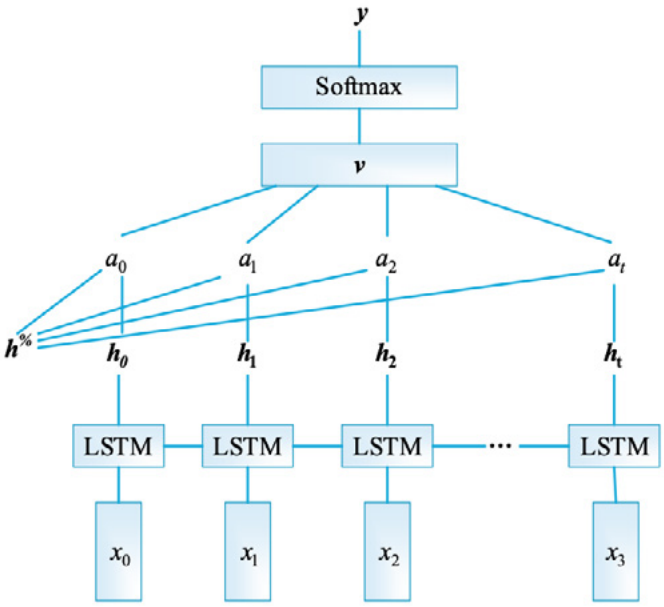


Figure 5 LSTM-Attention model

4. PROSPECTIVE ANALYSIS OF ARTIFICIAL INTELLIGENCE IN LINGUISTIC RESEARCH

4.1. QUALITY OF TRANSLATION IN DIFFERENT LANGUAGES

In order to verify the diagnostic ability of LSTM for language translation system, in the experiment, the LSTM-based artificial intelligence system is applied to different language translations to examine the ability of the diagnostic system in revealing the translation quality, strengths, weaknesses and characteristics of the translation system. The language translation systems that participated in the experiment included three statistical language translation systems, a rule-based language translation system that included diagnostic scores for each linguistic category at the lexical and phrase levels, a lexical category group that included all lexical categories, and a phrase category group that included all phrase-level category scores, system-level scores, and system-level scores computed using BLEU. Here, the small size of the test corpus resulted in a small number of sentence-level detection points with low reliability, so they were not considered for the time being. The first column in the table is the name of the diagnostic category or group of categories. The second and third columns are the diagnostic scores from System A and System B, respectively. The fourth column is the Paired t-statistic significance test score from the scores of the two systems. This score was obtained by repeating the experiment on a random subset of the test set 134). In this experiment, a Paired t-statistic value greater than 2.17 would indicate that the difference between the two scores is significant (>95%). The fifth column is the standard deviation of the diagnostic scores for Systems A and B. The sixth column is the 95% confidence interval for the diagnostic scores of System A and

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

B, accurate only to 0.01 due to space space. As can be seen by the BLEU scores. System B is 0.005 points higher than System A.

The translation system diagnostic results are shown in Table 1, where the difference in diagnostic scores between the two systems on the lexical category groups is not significant. On the diagnostic scores for each linguistic category at the lexical level, the two also have their own distinctions, and there is no obvious advantage for either one. However, on the phrase category group, the score advantage of System B, or LSTM, was more pronounced, and on the diagnostic scores for each linguistic category at the phrase level, System LSTM was higher than System A across the board, especially on the discontinuous distant phrase category. This result shows the advantage of System B in dealing with complex phrases and distant relations, an advantage that comes from recurrent neural network-based processing. Paired t-statistic statistics also show that the differences between the two systems are significant for all diagnostic scores. This comparison shows that the diagnostic system accurately captures the microscopic differences and commonalities between two systems with very similar macroscopic performance.

Table 1 Diagnostic results of translation system

	System A	System B	T Score	Score variance (A/B)	95% confidence interval (A/B)
	Lexical level				
Ambiguous word	0.59	0.59	2.88	0.00/0.00	0.58-0.61/0.58-0.61
Neologism	0.18	0.19	5.56	0.03/0.03	
Idiom	0.19	0.23	13.38	0.04/0.04	
Noun	0.59	0.59	2.68	0.00/0.00	
Verb	0.51	0.51	9.41	0.00/0.00	
Adjective	0.58	0.55	17.43	0.01/0.02	
Pronoun	0.75	0.73	13.49	0.02/0.02	
Adverb	0.53	0.54	7.11	0.01/0.01	
Preposition	0.65	0.64	6.21	0.01/0.01	
Quantifier	0.58	0.57	4.68	0.02/0.02	
Reduplicated word	0.33	0.39	9.86	0.10/0.08	
Match	0.66	0.65	8.07	0.01/0.01	
	Phrase level				
Subject-predicate collocation	0.51	0.51	7.36	0.01/0.01	
Predicate-object collocation	0.41	0.41	15.52	0.01/0.01	
Interobject collocation	0.44	0.51	9.51	0.01/0.01	
Quantifier collocation	0.51	0.51	3.56	0.01/0.01	
Azimuth collocation	0.52	0.53	2.83	0.03/0.04	
	Category group				

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

Vocabulary	0.48	0.48	8.03	0.01/0.01	
Phrase	0.47	0.49	13.97	0.01/0.01	
	System level				
Department of linguistics Class score	0.42	0.43	16.51	0.00/0.00	
BLEU series Class score	0.35	0.36	7.91	0.00/0.00	

4.2. IDENTIFICATION ACCURACY

In this paper, we use the BIO annotation specification, and the named entity category includes three categories, person name, organization name and place name. In order to determine whether this linguistic feature is useful for Uyghur named entity recognition, the four features Pos1-Pos4 are added to the LSTM intelligent model at the same time, which is used to compare whether the addition of the Pos4 feature, is helpful for the overall named entity recognition task. The affixed lexical features are shown in Table 2.It can be seen that, in terms of the F1 value, the addition of all of them improves the lexical features to some extent. There is an improvement of 0.5.

Table 2 affix characteristics /%

Trait	P	R	F1
Just_token	75.8	74.7	75.3
Pos1	76.7	74.7	75.6
Pos2	76.4	75.0	75.7
Pos3	74.4	75.8	75.4
Pos4	75.6	73.0	74.3
Pos1-Pos4	76.2	75.5	75.9(↑0.5)

After Table 2, it is found that linguistic features can improve the language named entity recognition accuracy, therefore, all the linguistic features will be added, and the comparison experiments with Pos1-Pos4 features and Suffix1-Suffix4 features will be conducted, and the comparison of linguistic features is shown in Table 3. The final F1 value is improved by 3.9%, which fully indicates that for complex morphological languages, adding linguistic features can improve named entity recognition accuracy.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

Table 3 affix characteristics /%

Trait	P	R	R
Just_token	75.8	74.7	74.7
Pos1-Pos4	76.2	75.5	75.5
Suffix1-Suffix4	78.6	75.0	75.0
All_feature	77.5	81.1	81.1

4.3. VALIDATION OF CREATIVE WRITING SKILLS

In order to facilitate statistical analysis, before conducting exploratory factor analysis, the suitability of factor analysis of questionnaire data N=727 was tested by KMO and Bartlett's test of sphericity, and the results showed that KMO=0.98 (>0.9), good level. LSTM was used to extract the common factors from the questionnaire data and the final factor loading matrix was obtained by the maximum variance method with orthogonal rotation, Table 4 shows the results of total variance interpretation of writing strategies. Five factors were extracted using the writing strategy, and the eigenvalues of each factor reached an acceptable value greater than 1. The cumulative variance contribution of the five factors was 66.5%, which is a desirable level of more than 60%. The common degree of each item, except R40, is greater than 0.5, and the factor loadings have reached 0.4 or more, indicating that the five factors extracted by the AI are all valid and can explain writing strategy ability better.

Table 4 Interprets the total variance of writing strategies

Inicial eigenvalue				Sum of squares of factor loads		
Divisor	Total	Variance %	Accumulate to %	Total	Variance %	Accumulate to %
1	25.8	52.7	52.7	8.1	16.6	16.6
2	2.1	4.4	57.1	8.1	16.5	33.1
3	1.9	4.1	61.2	6.5	13.4	46.5
4	1.4	2.9	64.1	5.9	12.2	58.7
5	1.1	2.4	66.5	3.7	7.7	66.5

5. CONCLUSION

In this paper, LSTM was used as the main tool to explore several aspects in the field of linguistics, including text analysis, speech signal processing and text generation. The suitability test (KMO=0.98) indicated that the data were at a good

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

level and suitable for factor analysis. Five common factors were extracted from the questionnaire data using the LSTM method, and the results showed that these five factors had high eigenvalues with a cumulative variance contribution rate of more than 60% of the desirable level, which indicated that these factors were able to explain the writing strategy ability better. In addition, the common degree of each item is greater than 0.5, and the factor loadings are all above 0.4, which further verifies the validity of these five factors extracted by AI. In addition, the article uses the BIO annotation specification for named entity recognition, which classifies named entities into three categories: personal names, institutional names, and place names. By adding the affix lexical features to the LSTM intelligent model, the results show some improvement in the F1 value, indicating that these features are helpful for the Uyghur named entity recognition task, which provides a strong support and innovation for the application of artificial intelligence in linguistic research.

ACKNOWLEDGMENTS

1. This research was supported by the funding of the following research project: Exploration on the Reform of College English Grammar Teaching by Educational Informationization (No.JZ180077).

2. This research was supported by the funding of the following research project: Corpus-assisted English Grammar Teaching Innovation (No.2018CG02644).

3. This research was supported by the funding of the following research project: An Innovative Model of Blended English Teaching by SPOC (No. FJJKCGZ18-793).

REFERENCES

(1) Yang, L., Fan, Z., & Zhou, J. (2022). Borderless Fusion Financial Management Innovation Based on Speech Recognition Technology. Scientific Programming.

(2) Dokuz, Y. , & Tufekci, Z. . (2020). Mini-batch sample selection strategies for deep learning based speech recognition. Applied Acoustics,171.

(3) Ho, N. H., Yang, H. J., Kim, S. H., & Lee, G. (2020). Multimodal approach of speech emotion recognition using multi-level multi-head fusion attention-based recurrent neural network. IEEE Access, 8, 61672-61686.

(4) Tsunemoto, A., Trofimovich, P., & Kennedy, S. (2023). Pre-service teachers' beliefs about second language pronunciation teaching, their experience, and speech assessments. Language Teaching Research, 7(1), 115-136.

(5) Hyland Bruno, J., Jarvis, E. D., Liberman, M., & Tchernichovski, O. (2021). Birdsong learning and culture: analogies with human spoken language. Annual review of linguistics, 7, 449-472.

(6) Bernardo, M. L. P. (2022). Localizing theory in a Spanish-language translation program. Teaching Literature in Translation: Pedagogical Contexts and Reading Practices, 262.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

(7) Rasulova, Z. (2022). TRANSLATION CONCEPTS IN THE CONTEXT OF MODERN LINGUISTIC RESEARCH. International Bulletin of Applied Science and Technology, 2(11), 161-165.

(8) Braithwaite, B. (2020). Ideologies of linguistic research on small sign languages in the global South: A Caribbean perspective. Language & Communication, 74, 182-194.

(9) Bafoevna, N. D., & Ikromdjonovna, K. N. (2023). The Main Directions of Theo linguistic Research In Modern Linguistics. Journal of Survey in Fisheries Sciences, 10(2S), 2127-2136.

(10) Mizumoto, A., Plonsky, L., & Egbert, J. (2021). Meta-analyzing corpus linguistic research. In A practical handbook of corpus linguistics (pp. 663-688). Cham: Springer International Publishing.

(11) Su, H., Zhang, Y., & Lu, X. (2021). Applying local grammars to the diachronic investigation of discourse acts in academic writing: The case of exemplification in Linguistics research articles. English for Specific Purposes, 63, 120-133.

(12) Awad Al-Dawoody Abdulaal, M. (2020). A cross-linguistic analysis of formulaic language and meta-discourse in linguistics research articles by natives and Arabs: Modeling Saudis and Egyptians. Arab World English Journal (AWEJ) Volume, 11.

(13) Chen, L., & Hu, G. (2020). Surprise markers in applied linguistics research articles: A diachronic perspective. Lingua, 248, 102992.

(14) Umarova, N. R. (2021). A linguistic approach to conceptual research. ASIAN JOURNAL OF MULTIDIMENSIONAL RESEARCH, 10(4), 62-66.

(15) Hamzah, M. H., Halim, H. A., Bakri, M. H. U. A. B., & Pillai, S. (2022). Linguistic Research on the Orang Asli Languages in Peninsular Malaysia. Journal of Language and Linguistic Studies, 18, 1270-1288.

(16) Oh, Y. R., Park, K., Jeon, H. B., & Park, J. G. (2020). Automatic proficiency assessment of Korean speech read aloud by non-natives using bidirectional LSTM-based speech recognition. Etri Journal, 42(5), 761-772.

(17) Hou, W., Wang, J., Tan, X., Qin, T., & Shinozaki, T. (2021). Cross-domain speech recognition with unsupervised character-level distribution matching. arXiv preprint arXiv:2104.07491.

(18) Santoso, J., Setiawan, E. I., Purwanto, C. N., Yuniarno, E. M., Hariadi, M., & Purnomo, M. H. (2021). Named entity recognition for extracting concept in ontology building on Indonesian language using end-to-end bidirectional long short term memory. Expert Systems with Applications, 176, 114856.

(19) Peng, L., Fang, S., Fan, Y., Wang, M., & Ma, Z. (2023). A Method of Noise Reduction for Radio Communication Signal Based on RaGAN. Sensors, 23(1), 475.

(20) Budinsky, R. , Ozmeral, E. J. , & Eddins, D. . (2023). The impact of hearing aid user's own voice on device signal processing. The Journal of the Acoustical Society of America.

ABOUT THE AUTHOR

Shaohua Jiang is working as a lecturer of School of Humanities, Fujian University of Technology. His research is focused within the fields of English Language

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

Education, Translation Education under the heading Smart Education and Artificial Intelligence.

Zheng Chen is an Associate Professor at the Department of Foreign Languages, Concord University College, Fujian Normal University. Her research is focused within the fields of English Language Education and American Literature Studies under the heading Artificial Intelligence.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

/04/

COMPUTER INFORMATION TECHNOLOGY APPLIED TO THE DESIGN OF DIGITAL LIBRARY INFORMATION INTEGRATION SERVICE SYSTEM

Ying Lin

- University of North Arizona, Flagstaff, Arizona State, USA
- 13821567380@163.com

Reception: 16 April 2024 | **Acceptance**: 6 May 2024 | **Publication**: 11 June 2024

Suggested citation:

Lin, Y. (2024). **Computer information technology applied to the design of digital library.information integration service system.** *3C Tecnología. Glosas de Innovación aplicada a la pyme 13(1)*, 78-97. https://doi.org/10.17993/3ctecno.2024.v13n1e45.78-97

ABSTRACT

*Based on computer information technology, this paper constructs a digital library information integration system. Firstly, the system is divided into information retrieval module and personalized service module according to the user's demand, and then the WebServices service is used to realize data integration and the library information obtained. And according to the principle of Fourier transform can be known to realize the conversion of information and knowledge. Finally, according to the movement state of knowledge expression, the formation of the wizard information base thus realizing information rectification. The structure shows that the digital library information integration system can provide 24h*14 hours of service, and the average response time of each operation is within 1s. It shows that the application of computer information technology in digital libraries can provide users with personalized services and perfect their interests.*

KEYWORDS

Computer information technology; digital library; personalized service; WebServices service; Fourier Transform

https://doi.org/10.17993/3ctecno.2024.v13n1e45.78-97

INDEX

https://doi.org/10.17993/3ctecno.2024.v13n1e45.78-97

1. INTRODUCTION

With the development of computer technology and network technology, digital libraries are gradually emerging [1]. However, because most of these digital libraries rely on their own strength to build, the cycle is long and costly, so that many units with such a need have encountered a lot of difficulties in the process of building digital libraries [2]. This has seriously hindered the process of building some digital libraries, so that the process of informationization construction is also constrained [3]. A more ideal solution to this problem is to implement the outsourcing of digital library construction projects, where professional system integrators provide digital library system solutions and carry out professional construction [4]. In this way, libraries can reduce investment by purchasing mature digital library integration system, and the integrator can make profits through multiple sales of integrated solutions [5-6].

Bramantoro, A. Creation of services with the ability to read all the required data from the library management system using full state technology to improve service quality. A technology acceptance model was used to measure organizational and user acceptance of the full state service system to provide some predictive basis for decision support for senior library management [7]. Sholeh, M. B. Automated Library Door Lock Control System using Arduino Uno and QR-Code. In designing the tool, Arduino Uno and QR-Code were integrated with library database application. Testing and analysis resulted in a 100% success rate of reading QR-Codes by the tested tool [8]. Shvartsman, M. pointed out new directions in e-library activities researching the storage of data, emphasizing the continuing interest in the subject of digital humanities and the fact that research in this field tends to focus on the problems of the modern society [9]. Bwalya, T. Through a descriptive survey, due to the fact that library administrators cannot effectively use all the modules of Koha due to lack of skills. Thus the main modules used are circulation module and cataloging module to enhance internet connectivity [10].

Lacuata, A identified the extent of implementation of guidelines for digitization of library resources in higher education institutions in terms of selection, technical requirements and implementation, legal aspects, budget, human resource planning, development and maintenance, preservation of digital content and project management. It will be used as a framework to provide effective and efficient digitization [11]. Afar, M. E. et al. used cultural institutions to develop new types of services under the influence of global information changes that determine new ways of knowledge mediation and management [12]. Lade, M randomly selected special libraries based on the descriptive statistics, mean scores, and standard deviation analysis of information on the attitude of professional librarians towards the attitudes of professional librarians towards digitization of professional library materials [13]. Babatope, I aimed to reveal the state of digital library education in higher education institutions and the preparation for the development of future digital librarians. The inclusion of digital library courses in order to prepare future digital librarians in this digital era, thus overcoming the often mentioned challenge of lack of ICT staff to manage digital libraries [14].

https://doi.org/10.17993/3ctecno.2024.v13n1e45.78-97

Based on computer information technology, this paper constructs an information integration service system for digital libraries. It is divided into information retrieval module and personalized service module from the user demand. And based on WebServices service to realize the data integration needs to forward the service request to the corresponding subsystem through the service bus, combined with the Brookes equation gives the framework of the transformation relationship between information and knowledge, points out the transformation relationship between information and knowledge. The effectiveness of the system in this paper is verified through performance testing, digital resource integration, integration service and update efficiency and integration system accuracy comparison, which proves that computer information technology improves the indexing efficiency of digital libraries, and the integration system is less time-consuming.

2. ANALYSIS OF DIGITAL LIBRARY INFORMATION INTEGRATION SYSTEM

2.1. SYSTEM USER REQUIREMENTS

Digital library information integration system to solve the problem of data source distribution and heterogeneity in the process of using library digital resources, it can be concluded from the requirements analysis that the roles interacting with the system can be mainly abstracted as users, administrators and system administrators [15-16]. The target user retrieval subsystem requirements are shown in Figure 1, and the user use function use cases include user login, user information, literature search, favorites, history storage, primary search, secondary search, subject search, classification search and result display customization. The specific contents are:

1. The retrieval system should have simple retrieval, natural language, phrase retrieval, etc., advanced retrieval, categorized retrieval functions, and primary and secondary retrieval.

2. The output of retrieval results should have sorting functions, such as sorting by date, title, author, relevance and so on. Users can select the source range of retrieval results, data types, etc.

3. Retrieval results collection, the user needs to personalize the collection of retrieval results and retrieval formulas and other information. Fourth, retrieval history preservation. In order to improve the retrieval efficiency, the system should provide a retrieval track saving function to provide reference for the user's next retrieval.

4. Registration and login. Users can apply for an account and use the user name and password to log in the system.

5. Users can change the personal information of the account.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.78-97

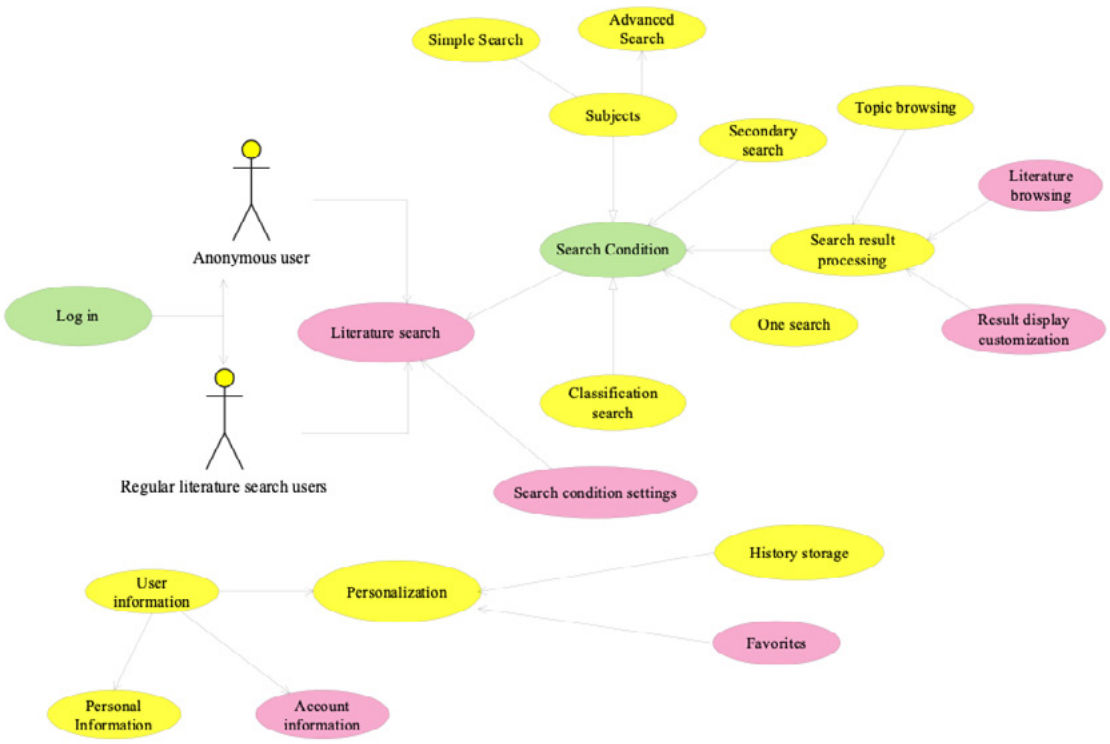


Figure 1 User example

2.2. SYSTEM ARCHITECTURE

The FDL integrated information retrieval system is divided into two main parts, the information retrieval module and the personalized service module. The main function of the user interface component is to intelligently assist the user in clarifying and refining the query request according to the user interest model. The refined query request is then submitted to the intermediary, waiting for the intermediary to return the query result and display it to the user. In addition, the UI also logs the user's basic actions while using the system. These operation logs will show the user's interest in certain information resources for the personalized service module to automatically summarize and learn the user's hobbies and improve the user's interest model. The UI resides on the user's terminal browser and provides the user and Web interaction interface.

The UI resides on the user's terminal browser and provides an interactive interface between the user and the Web. The user hobby learning machine selects training positive and negative examples according to the operation log, and updates the user interest model with user examples. Fig. 2 shows the library integrated information retrieval, where the mediator is activated upon receiving a query request from the UI, converts the query request submitted by the UI into an RDF statement to be submitted to the wrapper generation module, and then waits for the return result. When multiple wrappers respectively return the results of their respective queries in the form of RDF

https://doi.org/10.17993/3ctecno.2024.v13n1e45.78-97

documents, the intermediary summarizes the query results and combines them into a complete RDF result document in accordance with the relevance ordering.

Because each data source adopts different metadata representation standards, the intermediary needs to convert the metadata information describing the object data in the process of merging the results. The mediator queries the pre-defined metadata mapping table, and converts other metadata formats to Dublin Core-based metadata. This ensures the consistency of the RDF result document and facilitates the display and processing. As most user browsers do not support the display of RDF documents, in order to ensure that the RDF result document can be displayed in the client in a clearer and more friendly way, the intermediary also needs to be based on the pre-defined display format Schema will be converted to HTML pages, returned to the UL through the modification of the display format Schema can be changed by the display of the results of the query, which also increases the flexibility of the system. This also increases the flexibility of the system. The query of heterogeneous data sources is completed by the wrapper, FDL integrated information retrieval system adopts the mechanism of automatic generation of the wrapper, to ensure the scalability of the system [17].

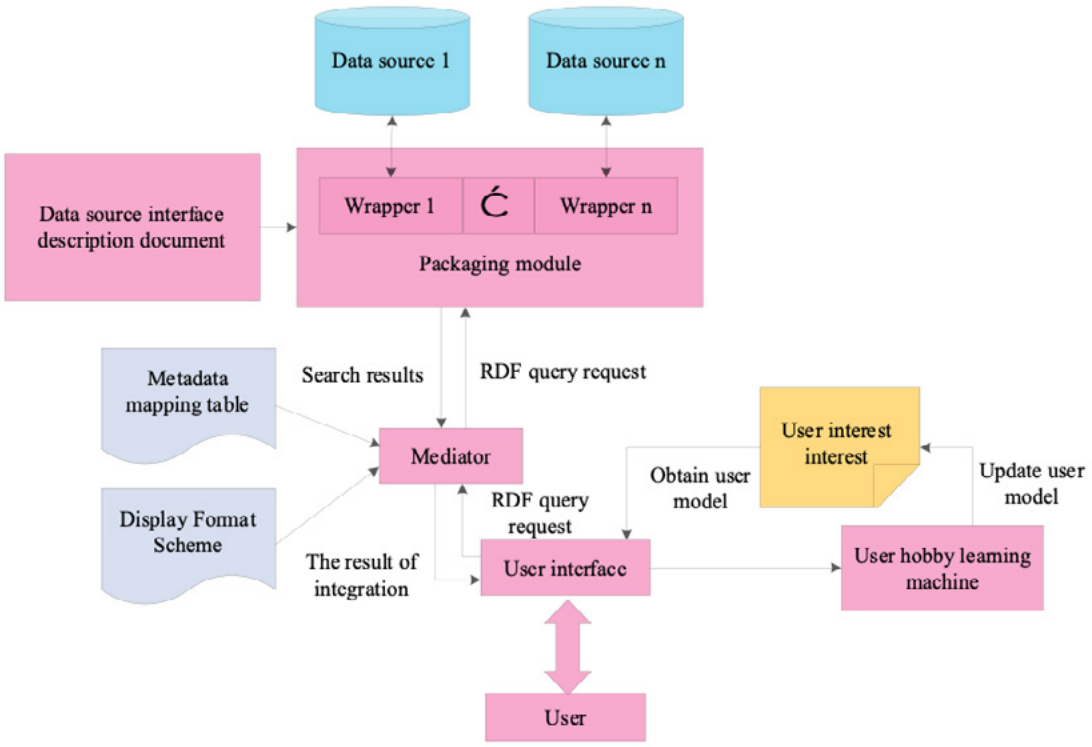


Figure 2 Library integrated information retrieval

Subsequently, based on the information obtained from the search, learning and information recommendation are based on user interests. It includes several components, including user personalized characteristic information base, user interest learning, information search, and information filtering. User interest learning and information recommendation is shown in Figure 3, including:

https://doi.org/10.17993/3ctecno.2024.v13n1e45.78-97

1. User personalized characteristics information base, or user personalized information table mainly includes information about the user's personalized characteristics. Such as the user's name, unit, address, specialty, occupation. User's previous request tasks such as classification, subject line, query range, etc. IP address, title, time of browsing, frequency of keywords and user's evaluation value, request time, URL of the requested file > HTTP version number, etc.

2. In order to provide personalized services to the user to recommend different information for different users, it is necessary to analyze the user's present or future interest preferences from the user's historical records in order to provide information services to the user actively [18]. Obtaining user needs through user-initiated descriptions. Embedded intelligent Agent, active tracking to collect the user's usual interest in information, from which to analyze the user's preference characteristics and background knowledge, the establishment of user personalized characteristics and stored in the user database.

3. Information search function through keyword matching, to the local or network resource database query, return the corresponding query results.

4. Information filtering function adopts content-based filtering method, extracts information content features from the query content, and matches them with the user personalized feature library to filter out the information related to the user's interest and recommend it to the user. Collaborative filtering technology is used to establish a user classification and recommendation mechanism to recommend information based on the same or similarity between users. Synthesize the strengths of content-based filtering and collaborative filtering to improve the accuracy of recommendation.

2.3. CHINESE WORD CLASSIFICATION MODULE

Chinese word classification module is closely related to the information query module and information indexing module in the system. The Chinese word classification model is shown in Figure 3, before indexing the crawled local web resources, the information indexing module will call the Chinese word classification module to perform word classification, and the indexed fields after word classification will be written into the Lucene library. After the user sends a query request, the information query module will also call the Chinese lexicon module to parse and slice the request submitted by the user. Then it is transformed into a query field that meets the standard of Lucene retrieval interface, and the query is compared in the index database to return the corresponding document collection. The Chinese word segmentation module mechanically slices and matches the input according to the words present in the dictionary. If a string is found in the dictionary, the match is successful, i.e., a word is recognized. Record the location of the word and return the dictionary index of the word. Dictionary management module mainly realizes the

https://doi.org/10.17993/3ctecno.2024.v13n1e45.78-97

statistics of the number of entries in the dictionary, dynamic addition and deletion of entries.

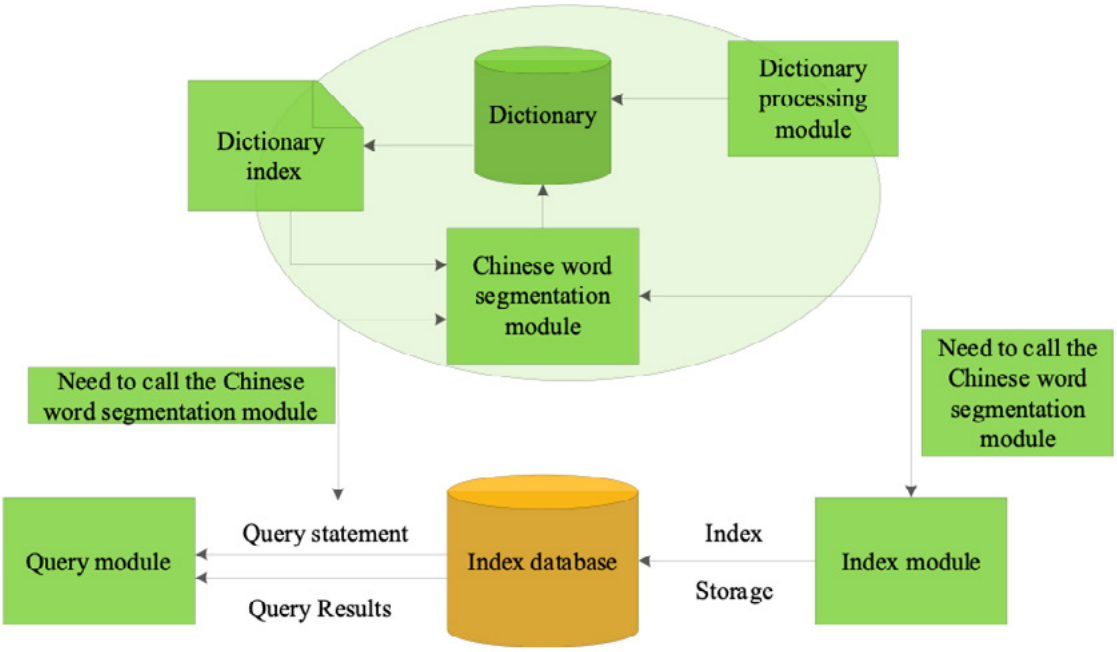


Figure 3 Chinese word classification module

2.4. IMPLEMENTATION OF DATA INTEGRATION

WebServices-based services to achieve data integration needs to be forwarded through the service bus service requests to the appropriate subsystems, based on the Web Services data integration process shown in Figure 4, the use of Web Services approach to data integration process is: first of all, the user issued a query command, analyze the query is a comprehensive query on which service or which services, and then through the unified Web Services interface to issue access orders, the Web Services interface to query the list of service buses to find the application system to access and forward services, and finally the application system will query the query to find the application system and forward services. If it is a comprehensive access to several services is decomposed into several services, and then through the unified Web Services interface to issue access commands, the Web Services interface in the service bus list query, find the application system to access and forward the service, and finally the application system will be queried by the return of the data to integrate the multi-table joint query data that is to get the desired information.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.78-97

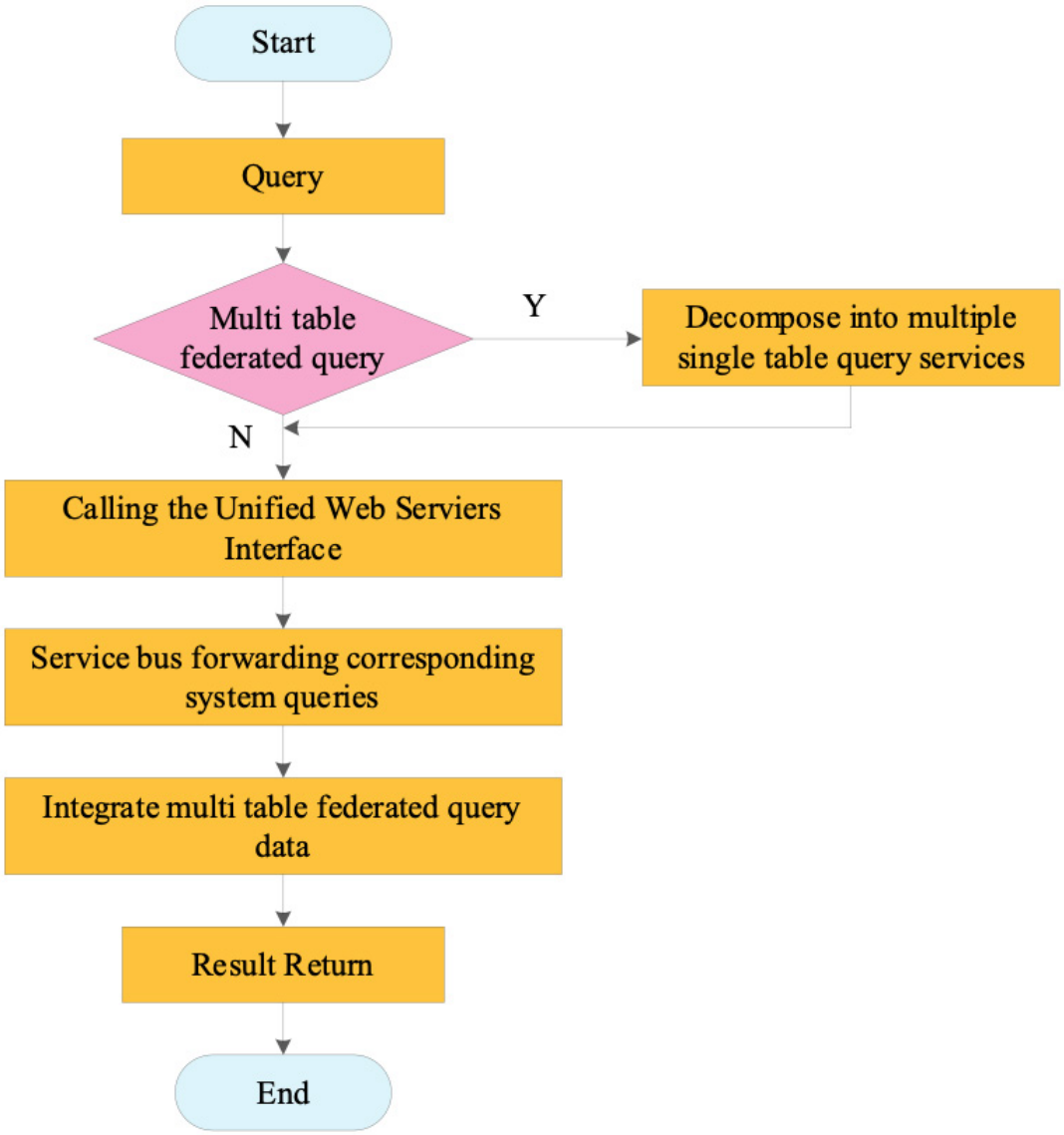


Figure 4 Web Services data integration process

3. CONDITIONS AND MECHANISMS OF COMPUTERIZED INFORMATION GENERATION

Based on the above acquired library information, this paper gives a framework of the transformation relationship between information and knowledge in conjunction with Brookes' equation, which points out the direction for research on the utilization of information. This section analyzes and discusses the equation from physical and cognitive perspectives, pointing out that there exists a transformation between information and knowledge, but the transformation from information to knowledge is the most important as far as knowledge construction is concerned.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.78-97

3.1. CONDITIONS FOR INFORMATION GENERATION

According to the principle of Fourier transform, it is known that in order to realize the transformation of two different domains of information and knowledge, the transformation process must have the following transformation functions:

$$Is(t) = I(t)S(t) = I(t)\delta_{Ts}(t) \quad (1)$$

$$I(S) * S(S) = Is(S) \quad (2)$$

$$Is(S) * P(S) = K^*(s) \quad (3)$$

Reduce it to:

$$I(S) * K(S) = K(S + \Delta S) \quad (4)$$

The above equation illustrates that the knowledge system *K (S)* is convolved with the knowledge spectrum *I (S)* of the information to produce a new knowledge system $K(S + \Delta S)$, i.e., the realization of the transformation of information and knowledge to require three different stages of conditions:

1. The first stage is the perceptual cognitive process of knowledge formation, which requires that the knowledge system must be selective, i.e., sampling the attributes of information $= I(t)$ to form a sample sequence of information $I_s(t)$.

2. The second stage is the rational cognitive process of knowledge, and then the sample sequence of $I_s(t)$ is abstractly categorized to form the transformation of the knowledge spectrum sequence Is (S).

3. The third stage is the process of utility formation of knowledge, i.e., Is (S) is convolved with the original knowledge spectrum K(S) of the knowledge system to produce a new knowledge spectrum, i.e., a new knowledge increment is introduced, thus realizing the rectification of information. ΔS represents the utility of the information to generate knowledge, there may be three cases, when $\Delta s > 0$ indicates that the information generates new useful knowledge. When $\Delta s = 0$ indicates that the information does not generate new useful knowledge, and when $\Delta s < 0$ indicates that instead of generating new useful knowledge, the information increases ambiguity.

3.2. INFORMATION GENERATION MECHANISMS

Information expresses the states of things in motion and the ways in which the states change, and knowledge expresses the states of things in motion and the laws of state change.4 For some random piece of information *X*, if there are *N* possible states of motion, i.e., if its state space is $X_1, X_2, \ldots\ldots, Xn$, and if the ways in which these states change are distributed according to some probability $P1, P2, \ldots\ldots, Pn$, then the probability space, which is a combination of the state space and probability distribution, is inscribes this information.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.78-97

According to the law of movement state and state change of knowledge expression has a certain external form, has a certain logical content, presents a certain value to the cognitive subject, and the corresponding knowledge is called morphological knowledge, content knowledge, utility knowledge, respectively [19-20]. From the point of view of the information recognition process, both morphological knowledge and utility knowledge can be directly perceived from external information. The mechanism of generating morphological knowledge is a form comparison process, which starts with sampling information to establish information samples and extracting the external formal characteristics of the motion state [21]. Such as size, height, weight form, color, etc., and formal features of the way the state changes, such as randomness, episodicity, and ambiguity. Observing multiple information samples and identifying a set of information features with common meaning constitutes a feature information base, which in turn is further formed into a wizard information base through the computation of the feature information base by the information architect. The utility of a piece of information for a certain subject not only depends on the subject's goal, but also relates to the amount of knowledge of the subject. The generation mechanism of utility knowledge is a process of comparing the distance between the state of reality reflected by the state of information and the state of the subject's goal, because from the viewpoint of the most basic method of knowledge induction, only such and such a state of things in motion as well as the way of change of state will reflect such and such information, and only such and such information can be obtained by the state of motion of the thing as well as the state of change of the law of such and such knowledge. The content knowledge, on the other hand, can only be obtained through the understanding of the way things move and change. Content knowledge can only be realized through the mapping of morphological knowledge and utility knowledge. Let the information generated by the morphological knowledge for , from the information generated by the utility knowledge for , content knowledge is the simplest mechanism to generate the information categorization division, through the wizard information base navigation links, the knowledge spectrum convolution results into the knowledge base, so as to realize the information rectification and effective conversion, performance in the morphological knowledge and utility knowledge between the establishment of a mapping relationship .

Figure 5 shows the mechanism of information generation, through the above analysis of the mechanism of information generation of morphological knowledge, utility knowledge and content knowledge, it can be concluded that the realization of information and knowledge transformation must have three conditions, the selectivity of the knowledge system, the transformability of the knowledge system, and the utility of the knowledge system.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.78-97

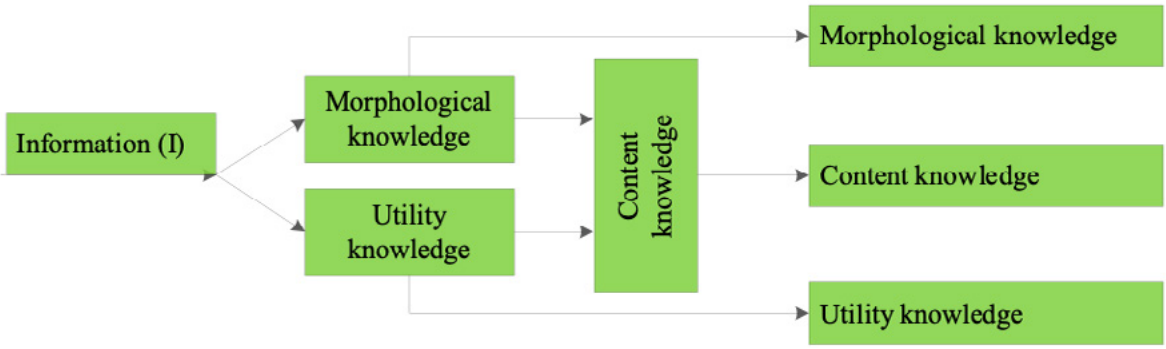


Figure 5. Information generation mechanism

4. SYSTEM PERFORMANCE TESTING

4.1. TEST METHODS

In this paper, from the user's point of view, the system function test, the test includes the performance of the system as well as the function, the performance test is from the perspective of the system's non-functional design, test the system's final results whether to meet the non-functional needs of the previous; functional test, on the other hand, is to test the system's functional modules to test the system whether it is capable of completing the various business.

4.2. ANALYSIS OF TEST RESULTS

4.2.1. PERFORMANCE TESTING

This paper samples the data of the system running for a week to test the performance of the system, and the results of the system performance test are shown in Table 1. After the system is online, it can ensure that the system can provide 24h*14 hours of service, the average corresponding time of the system function is less than 0.2s, the slowest corresponding time on the system line is less than 1.2s, and the system data analysis is done in the background. Considering the performance of the server, the time for system data analysis is less than 1h, the average trouble-free time is less than 0.3 times per month, and the system performance is excellent.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.78-97

Table 1 System performance test results

The Project	Specific Projects	Indicators
Robustness	Availability	14*24 hours normal operation
	Average Failure Frequency	<0.3 times/month
	Average Repair Time	<1h
Real-Time System	Average Response Time	<0.2s
	Slowest Response Time	<1.2s
	Statistical Time	<30min
	Average download speed	<1M/s

4.2.2. FUNCTIONAL TESTING

Functional testing is to test the various functions of the system, specifically manifested in the functional interface for data input, test the system whether it can return the expected results. Table 2 shows the navigation function test, 100 times the test results of a row to reach 100%, no error.

Table 2 Functional navigation tests

No.	Content	Number of tests	Consistency of test results/%	
			Normal test	Error test
Query_001	Navigation Functions	100	100	100
Query_100	Search by Title	100	100	100

In addition, various performance indicators of specific business operations such as system insertion, deletion and modification are also analyzed and summarized. The results of the categorized operating system test are shown in Table 3, and the average response time of each operation is within 1 s. The CPU occupancy rate is also controlled within 19%. Because now the computer memory is generally above 2G, and because of the small amount of data accumulated in 10 days, DB can be loaded into memory at one time, the number of memory exchanges and the number of hard disk reads and writes are generally low. And the computer information technology has a good control of the disk write operation, the modified data can be temporarily saved in the memory and written to the disk at one time when the CPU is idle, so the impact on the corresponding time has been controlled, and the exception rate is 0 in both entry and modification.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.78-97

Table 3 Classification operating system test results

	Corresponding time/ms	CPU usage/%	Memory usage/M	Memory swaps/session	Disk reads/writes	Abnormal/%
Search	580	18.45	274.6	0.02	0.08	0.04
Record	121	11.22	233.9	0.11	0.26	0
Modify	548	17.69	254.7	0.06	0.23	0
Delete	728	16.88	253.1	0.09	0.16	0.1

4.3. INTEGRATION OF DIGITAL RESOURCES

In order to prove the practicability of the proposed method, a general-purpose computer is set up in the experimental platform, the CPU of which is intrlCor 3110M, the memory is proposed to be 4G, and the operating system is Win-dows10. 120GB of resources are to be integrated in a library service platform, which contains domestic and foreign newspapers, drawings, scientific research results, books, and other attributes, and the integration planning of these resources is carried out in this paper. The integration planning is carried out. Figure 6 shows the results of resource storage space integration under the system, the proposed system after integration, there will be no problem of resource loss, the book category of digital resources for the 60/GB. This is because the proposed method will be converted to digital resources into associated data, relying on the existence of the correlation between the data to classify the different types of digital resources, so as to enhance the integrity of the digital resources in the integration process.

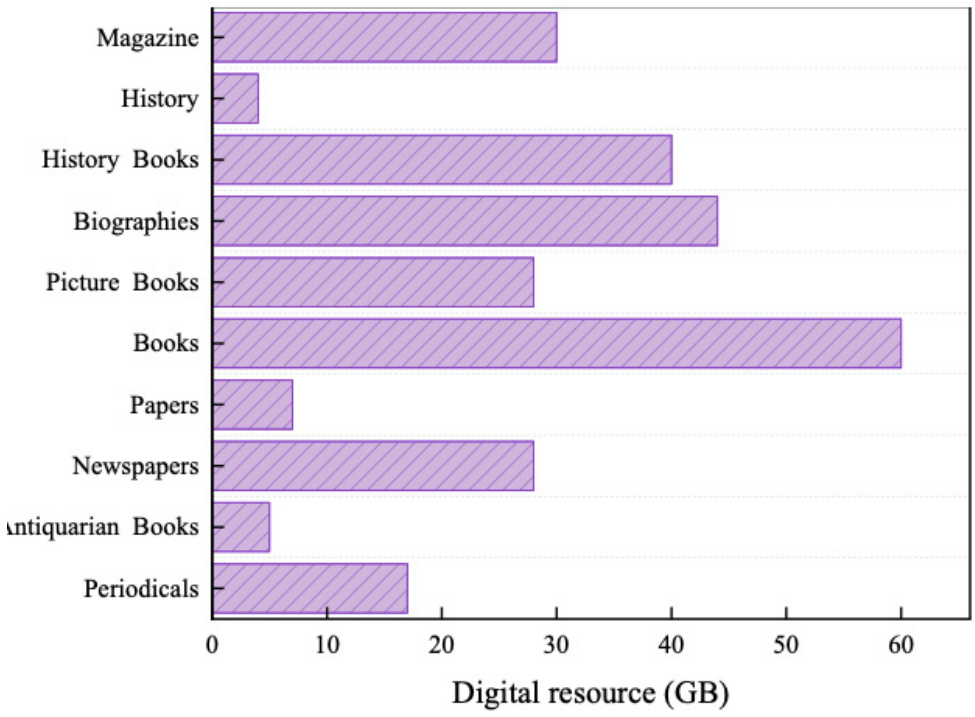


Figure 6 Results of digital resources integration

https://doi.org/10.17993/3ctecno.2024.v13n1e45.78-97

4.4. INTEGRATION SERVICES AND UPDATING EFFICIENCY

To a certain extent, the efficiency of integrated literature retrieval depends on the construction efficiency of the index. Under the computer information technology, each working node has the same working performance with the master node, which improves the construction efficiency of integrated index. The experimental system tested the traditional centralized structure, the WES retrieval method as an example with computer information technology, under the conditions of different document set size index construction efficiency, the document size of the specific 150M, 500M, 1G, 2 G, 5G, 1T, index construction efficiency results shown in Table 4. Compared with the traditional centralized structure of the retrieval method, the computer information-based retrieval method has a huge advantage in the index construction efficiency. This is mainly due to the computer information structure gives each independent node more computing power, when the size of the file to be queried is large, the advantage of the distributed structure is more obvious, in 1.25M/s-1.61M/s.

Table 4 Comparison of index building efficiency

Serial number	Literature size	Index construction efficiency M/s	
		WES search	Computer information technology
1	150M	6.68	1.25
2	500M	6.69	1.27
3	1G	13.25	1.36
4	2G	32.44	1.49
5	5G	58.57	1.52
6	1T	88.64	1.61

As the digital library resources are in a state of constant updating, the index that has been established needs to be updated in real time, the update efficiency of the document index under the same data size, the index update efficiency is shown in Table 5. When the size of the document is 150M, the index update efficiency of computer information technology is less different from the traditional method, and the computer information technology is 0.66M/s. When the size of the document continues to increase to 1T, the index update efficiency of computer information technology is much higher than the traditional method, and the computer information technology is 1.54M/s, and the WES retrieval is 2.78M/s. It shows that the computer information technology-based integrated document retrieval method is more suitable for the integrated service design of digital libraries.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.78-97

Table 5 Index update efficiency

Serial number	Literature size	Index construction efficiency M/s	
		WES search	Computer information technology
1	150M	0.78	0.66
2	500M	1.15	0.97
3	1G	2.32	1.25
4	2G	2.64	1.41
5	5G	2.78	1.54
6	1T	2.82	1.68

4.5. INTEGRATED SYSTEM ACCURACY COMPARISON

Four different fields, namely, economy, empirical evidence, sports and education, were selected and a total of 10,000 text records were collected. Under the same data size and hardware and software conditions, the proposed computer information technology-based ISS in the paper is compared and validated with the WES-based and metadata-based ISS using retrieval time and correctness of retrieval results as the comparison indexes. The results of the retrieval training time comparison are shown in Fig. 7.

When the size of literature is less than 2000 items, the training time of metadata-based integrated service system is more than 180ms, and the training time continues to improve with the expansion of literature size, and serious fluctuations occur in the middle. The smoothness of the WSE service system is better than that of the metadata method, but the efficiency of the integration continues to decline with the expansion of the size of the literature, and the time consumed is the highest. The advantage of computer information technology (CIT) in large-scale literature integration is gradually highlighted with the increase of literature dataset size, and when the literature size is more than 3000, the training time tends to stabilize, which fully explains that the integration system based on CIT is less time-consuming.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.78-97

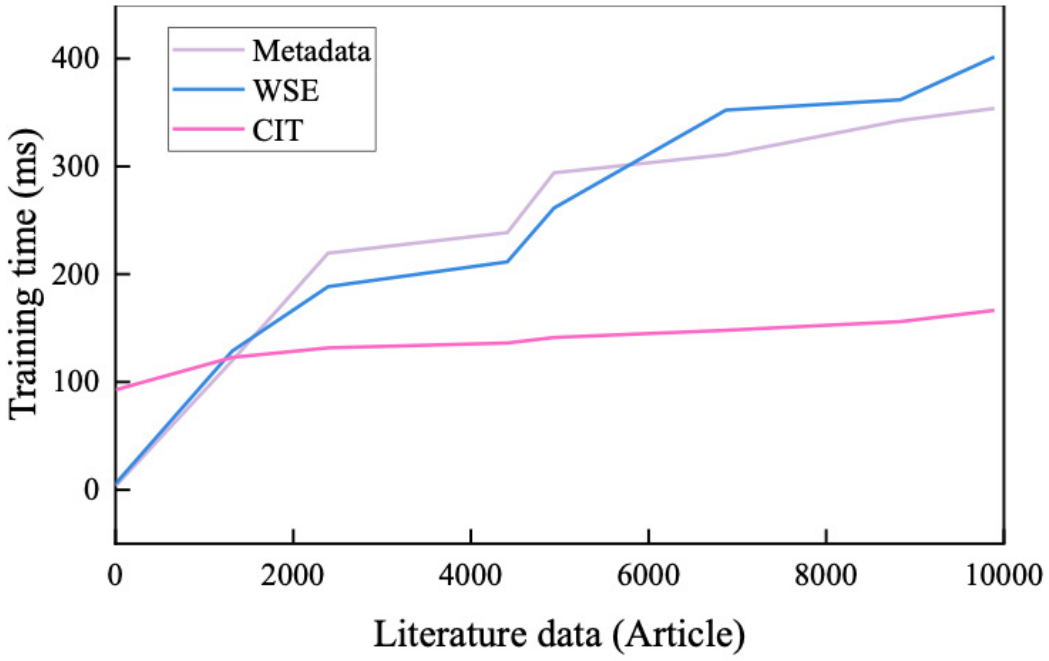


Figure 7 Integration time variation

The results of the integration correctness comparison are shown in Fig. 8, when the data size of the literature reaches 10,000 items, the correctness ratio of the integration method based on computer information technology can still exceed 85%, which is significantly better than the traditional integration method.

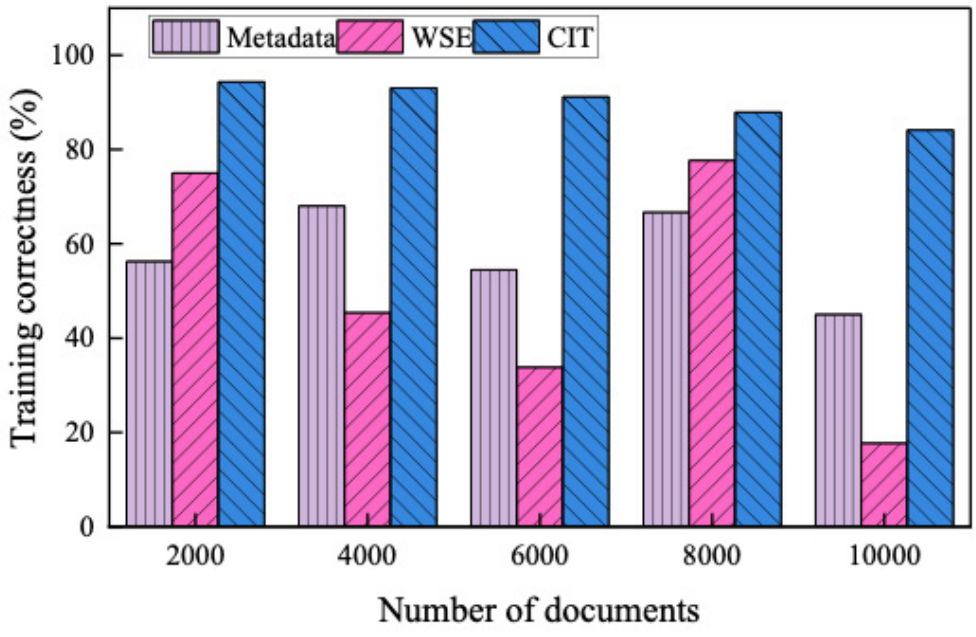


Figure 8 Comparison of correct rate of integrated system

5. CONCLUSION

This paper takes the design of digital library information integration service system as the research objective, designs it on the basis of computer information technology,

https://doi.org/10.17993/3ctecno.2024.v13n1e45.78-97

and verifies the effectiveness of the system in practical analysis, and the conclusions are as follows:

1. In the performance test, the system can provide 24h*14 hours of service, the average corresponding time is less than 0.2s. And the test results of 100 times of one line reaches 100% without error.

2. In the integrated service and update efficiency test, when the size of the documents to be queried is large, the indexing time is in the range of 1.25M/s-1.61M/s. It shows that the integrated retrieval method of documents based on computer information technology is more suitable for the design of integrated services in digital libraries.

3. By comparing the accuracy rate of the integrated system, the machine information technology when the size of the document is more than 3000, the length of the training time tends to stabilize, which fully demonstrates that the integrated system based on computer information technology is less time-consuming.

REFERENCES

(1) Kroll, H., Pirklbauer, J., & Balke, W. T. (2021). A toolbox for the nearly-unsupervised construction of digital library knowledge graphs. In *2021 ACM/IEEE Joint Conference on Digital Libraries (JCDL)* (pp. 21-30).

(2) Waterman, A. D., Wood, E. H., Ranasinghe, O. N., Lipsey, A. F., & Salas, M. A. P. (2020). A digital library for increasing awareness about living donor kidney transplant: A formative study. *JMIR Formative Research, 4*(7).

(3) Abdullah, A., Nasir, N. E., Mei, L. C., & Lee, K. M. (2021). Automating secondary school libraries: A web-based library management system. *IASL Annual Conference Proceedings*, 193-202.

(4) Shi, Y., & Zhu, Y. (2020). Research on aided reading system of digital library based on text image features and edge computing. *IEEE Access, 8*, 205980-205988.

(5) Zhang, W., Liu, B., & Tsai, S. B. (2022). Analysis and research on digital reading platform of multimedia library by big data computing in internet era. *Wireless Communications and Mobile Computing, 2022*, 1-10.

(6) Solomon, F. S., Kennedy, Z. O., Deinbofa, G., & Godwin, O. O. (2019). Design and implementation of digital library management system: A case study of the Niger Delta University, Bayelsa State. *International Journal of Scientific and Research Publications (IJSRP), 9*(12), 96112.

(7) Bramantoro, A. (2020). Stateful library service system design and implementation in Saudi Arabia. *International Journal of Electrical and Computer Engineering, 10*(3), 2690.

(8) Sholeh, M. B., Herlina, A., & Hasan, F. (2020). Design of automatic door lock control system library Nurul Jadid University based on Arduino Uno R3 and QR-code. *Buletin Ilmiah Sarjana Teknik Elektro, 2*(2), 91.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.78-97

(9) Shvartsman, M. E., & Kryzhanovskaya, O. N. (2019). Digital research methods: The new vector in library informatization. *Bibliotekovedenie [Russian Journal of Library Science], 68*(1), 103-111.

(10) Bwalya, T., & Akakandelwa, A. (2021). Challenges of using Koha as a library management system among libraries in higher education institutions in Zambia. *DESIDOC Journal of Library & Information Technology, 41*(2), 82-87.

(11) Lacuata, A. N. (2020). Digitization of library resources in higher education institutions in La Union, Philippines. *Preservation, Digital Technology & Culture*(4), 49.

(12) Afar, M. E., Lubina, T., & Mlinarević, I. (2019). Digitization of Osijek grammar school reports at the library department of the Museum of Slavonia. *Informatologia, 52*(1-2), 55-64.

(13) Lade, M., & Abubakar, S. K. (2019). Digitization of library materials in special libraries in Abuja, Nigeria. *Information Impact Journal of Information and Knowledge Management*, 14.

(14) Babatope, I., Oghuvwu, V., & Dumbiri, R. (2021). Digital library education in tertiary institutions: Roadmap to revolutionizing Nigerian university library services. *Preservation, Digital Technology and Culture, 49*(3), 99-109.

(15) Yurchenko, O. (2019). Solutions for library collection digitization: The experience of the Far Eastern Federal University Scientific Library. *Scientific and Technical Libraries*(1), 54-62.

(16) Lomakina, O., Kookueva, V., & Makarenko, A. (2021). Redistribution of economic resources in the digital society. *Business and Society Review*, 25-35.

(17) Li, G., & Jiang, G. (2019). Construction and planning of library service facilities system based on public digital culture education in international cultural metropolis. *Open House International, 44*(3), 64-67.

(18) Song, H. Y., Zhang, H., & Xing, Z. H. (2021). Research on personalized recommendation system based on association rules. *Journal of Physics: Conference Series, 1961*(1), 012027.

(19) Li, Z., Lin, H., & Zheng, W. (2020). An effective emotional expression and knowledge-enhanced method for detecting adverse drug reactions. *IEEE Access, PP*(99), 1-1.

(20) Finnegan, L. A. (2019). Supporting student knowledge using formative assessment and universal design for learning expression. *The Journal of Special Education Apprenticeship, 8*(2), 7-7.

(21) Clark, M. D. (2011). Morphological knowledge and decoding skills of deaf readers. *Psychology, 2*(2), 109-116.

ABOUT THE AUTHOR

Ying Lin was born in Tianjin, China, in 1982. She graduated from Tianjin University, China, Majoring in computer information management, and from Nankai University, China, Majoring politics of the firm. She is currently studying computer information technology at the University of North Arizona in the United States.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.78-97

/05/

INTELLIGENT CABLE FAULT DETECTION TECHNOLOGY CONSIDERING CURRENT HARMONIC CHARACTERISTICS

Yuning-Tao*

- College of Electrical Engineering and New Energy, China Three Gorges University, Yichang, Hubei, 443000, China
- taoyuning12@163.com

Reception: 8 April 2024 | **Acceptance**: 6 May 2024 | **Publication**: 8 June 2024

Suggested citation:

Yuning-Tao. (2024). **Intelligent cable fault detection technology considering current harmonic characteristics.** *3C Tecnología. Glosas de Innovación aplicadas a la pyme. 13(1)*, 99-128. https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

ABSTRACT

With the rapid development of the power system, cable faults have become an important factor affecting the stable operation of the power system. In this paper, for the problem of cable faults, we improve the overshoot, undershoot phenomenon and sieve speed of the envelope fitting in the Hilbert-Huang transform algorithm, and extract the harmonic characteristics of the current of cable faults by using the improved HHT model. Then, we utilize the information entropy and wavelet singular entropy algorithm to integrate semi-parametric support vector machine algorithm, S-SVM, and construct the wavelet singular entropy and S-SVM model. The information entropy and wavelet singular entropy algorithms are fused with semiparametric support vector machine algorithm, S-SVM, and constructed into wavelet singular entropy and S-SVM models, which are applied to the cable fault identification experiments for detecting different faults in cables. The experimental results show that, when the cables are short-circuited, the currents of different short-circuited cables are all lower than the normal currents, and the wavelet singular entropy and S-SVM models reach more than 92% of the accuracy of the identification of the degradation of the cable line and short-circuited faults. The accuracy of the wavelet singular entropy and S-SVM model for the identification of cable line deterioration and short circuit faults reaches more than 92%, and the overall cable fault detection reaches 98.04%.The maximum error value of the wavelet singular entropy and S-SVM model for detection is 0.5329, and there are only two groups of data more than 0.5.The algorithms in this paper are able to detect the various localization of the cable faults quickly and accurately, and they have a high practical value.

KEYWORDS

HHT algorithm; Wavelet singular entropy; S-SVM model; Current harmonic characteristics; Fault detection

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

INDEX

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

1. INTRODUCTION

Power resources have become one of the most important energy sources in China, the rapid increase in the consumption of electric power resources in small, medium and large cities, in order to meet the needs of the masses of electricity, widely used power cables as a transmission tool and connecting lines [1-2]. the current development of China's electric power industry projects continue to increase, in order not to take up too much land resources, cables are usually generally buried in the ground, which increases the difficulty of troubleshooting power cable faults to a certain extent. If the maintenance work is not timely, then it is easy to increase the probability of power outages, bringing difficulties to the people's lives, directly affecting people's production and life [3-4]. combined with the current social development trend in China, the most critical type of power system failure is cable failure, to ensure the stability and safe operation of China's power system, it is necessary to carry out power cable failure at the first time, the power cable failure is the most critical type of power cable failure. To ensure the stable and safe operation of China's power system, it is necessary to carry out power cable fault inspection and testing at the first time, accurately put forward the cable inspection method, and effectively put forward the measures to solve the fault, repair the power cable faults, so as to promote the stability and safety of the power project [5-6].

In order to reduce the use of land resources, to bring higher economic benefits, so the cable is usually buried in the ground, but this to a certain extent also brings certain problems, due to the cable buried deep underground, if a fault occurs, it is difficult to investigate, which increases the difficulty of investigation [7-8]. In the detection of cable faults, the principle that must be adhered to is to be green, while maximizing the economic benefits, not only to require the practicality of strong, but also to ensure that the scientific nature of the use of advanced technology to fully reduce losses, and constantly improve the efficiency of the power grid in the operation of the power industry to a certain extent will promote the progress and development of China's electric power industry, and play a huge significance of the promotion of the [9-10].

With the continuous development of the market economy, people's living standards continue to improve, the structure of the urban power grid is more complex, the number of cables in use is increasing day by day. the safety of the cable operation, directly on the power system to bring a direct impact on the cable management and maintenance of the power sector has become the focus of the attention of the electric power sector. Baranowski, J. and other researchers proposed a new method to detect different signals based on the depth distribution of the Bayesian function to diagnose cable faults, the results of the study confirm that the method has great potential for diagnosis in unknown situations [11]. Liu, X. and other researchers designed a short cable line fault location method based on the theory of transmission line and the circuit theorem, and the test was carried out on 50 coaxial cables, and the test verified that the accuracy of this method is not affected by the impedance of faults and the terminating impedance [12]. Xuebin, Q. and other researchers proposed an on-line cable fault diagnosis method to address the need for on-line diagnosis of cable faults,

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

which is more advantageous than the traditional shallow neural network-based cable fault identification method [13]. Cataldo, A. C. G. described the time-domain reflectance (TDR) based localization method along the permittivity variations (DPVs) of cable systems, and pointed out that due to the increase in the number of DPVs, the fault location is not as accurate as the fault impedance and termination impedance of the cable system. It was pointed out that increasing the pulse width to study the localization of cables over longer distances leads to focusing on different TDR reflectance maps each time, which is very time-consuming and does not guarantee optimal performance [14].Wang, F. and other researchers envisioned a fault diagnostic method based on the BP-Adaboost algorithm to solve the problem of faults on aeronautical cables, and the results of the algorithm were studied by using the Matlab software for analysis and example feedback to validate the results. Matlab software is used to analyze the algorithm results and feedback examples to verify the feasibility of the proposed fault diagnosis method [15]. Sian, H. W. and other researchers designed a hybrid diagnostic algorithm based on the Discrete Wavelet Transform (DWT) and Symmetric Dot Plot (SDP) analysis of Convolutional Probabilistic Neural Networks (CPNN) to solve the insulation faults in XLPE cables. Simulation tests show that the accuracy of this method is more than 96%, and the accuracy of this method is higher than 96%. Simulation tests have shown that the method can diagnose power cable faults with an accuracy of more than 96% and a short detection time, which makes it fully capable of detecting insulation faults in cables [16]. Marriott, N. discusses the role and advantages of Megger's new SMART THUMP ST25-30 Portable Cable Tester, which provides an automated test sequence with the ability to identify, prelocate, and pinpoint cable faults, and to automatically supplement the interpretation of the test results. Non-specialized users can obtain reliable results in a safe and easy way [17]. Lowczowski, K. and other researchers analyzed the application of cable shielding currents in the identification and location of ground faults, and in this way gave phase ground fault currents in different power system configurations, in order to test the ability to detect faults based on the above principles, and proved the reliability of this method for fault detection. Finally, a solution to improve the localization capability is proposed, and the feasibility of the improved solution is confirmed by simulation tests with PSCAD software [18]. Hu, C. and other researchers investigate the ship cable fault information acquisition model based on the automatic identification technology, and simulation tests are conducted to confirm the feasibility of the Hilbert-Huang Transform (HHT) to automatically identify cable faults, and to construct an automatic cable fault information acquisition model [19]. Lai, Q. and other researchers studied the 110kv transmission line cable terminal tail pipe breakdown fault, and give the cable installation improvement measures to reduce the probability of cable terminal breakdown faults, and after simulation and disassembly test, the feasibility analysis, for similar cable faults, is carried out. The feasibility analysis is carried out to provide important reference opinions for the investigation and solution of similar cable faults [20]. Liu, N. and other researchers conceived a deep neural network-based cable fault signal classification and identification method to accurately identify the early cable fault problems, and the reliability of the method is demonstrated through experiments [21]. Wang, Y. and other

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

researchers designed a method based on the constrained Boltzmann machine (Wang, Y. and other researchers designed a cable early fault identification method based on Restricted Boltzmann Machine (RBM) and Stacked Auto-Encoder (SAE), which demonstrated higher accuracy after simulation tests comparing with traditional methods such as convolutional neural networks [22]. A new algorithm for accurate location of early faults in underground cables using double-ended synchronized zero sequence waveforms was proposed by Qu, K. et al. A large number of simulation experiments were conducted at PSCAD/EMTDC to support the accuracy and reliability of the algorithm [23]. Kwon, G. Y. et al. discussed an improved fault localization technique, instantaneous frequency-domain reflectance and tangent-distance pattern recognition, for fault diagnosis and protection of cables, and simulation data verified that this method can improve the reliability of high-voltage DC power systems. For fault diagnosis and protection of cables, and simulated test data verified that this method can improve the reliability of HVDC power systems [24].

In this paper, the Hilbert transform is applied to the IMF components to obtain the Hilbert-Huang transform model, and then the local extreme points are densified by using the cut-contact mean points, and then the interpolated curve segments are spliced by the segmented power function to improve the insufficiency of the envelope fitting in the HHT transform model, and then the current harmonic characteristics of the cable faults are extracted and analyzed, and the wavelet singularity is obtained by combining the wavelet transform with the information entropy. The wavelet singular entropy is obtained by combining the information entropy and wavelet transform, which is integrated into the S-SVM algorithm, and the sparse greedy matrix approximation is used to select the basic element set, and the iterative reweighted least squares algorithm is introduced to iterate and optimize the S-SVM model, which finally constitutes the wavelet singular entropy and the S-SVM model. the signal is decomposed by the wavelet singular entropy and the signal components are extracted, and the extracted current harmonic features are screened and reconstructed, and the current harmonic features are extracted and analyzed. After that, the extracted current harmonic feature vectors are filtered and reconstructed, and then the cable fault samples are input to the S-SVM model for training, and finally the wavelet singular entropy and S-SVM models are realized to accurately identify different faults of cables.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

2. CONSTRUCTION OF INTELLIGENT CABLE FAULT DETECTION MODEL

2.1. CONSTRUCTION OF IMPROVED HILBERT-HUANG TRANSFORM MODELS

2.1.1. HILBERT-HUANG TRANSFORM ALGORITHM

The HHT algorithm is to sieve the various frequency components or trends contained in the signal layer by layer in order to obtain a series of IMF components containing different feature information, and then these IMF components are subjected to the Hilbert transform, which can obtain the Hilbert spectra and the marginal spectra, and then obtain the amplitude distribution pattern of the signal in the spatial or temporal scale.

1. Instantaneous frequency

In the process of analyzing nonlinear signals, the instantaneous characteristics have a very important role, in the traditional Fourier transform, less than a wavelength signal will not be able to give the definition of the frequency, that is, it cannot be used to accurately describe the instantaneous parameters of the non-smooth signal, the instantaneous frequency mentioned in the Hilbert transform has the actual physical meaning, and the basic definition of the frequency is consistent.

When analyzing a smooth signal, the frequency of the signal refers to the f in the Fourier transform.

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j\pi ft}dt \tag{1}$$

In the formula f, becomes the Fourier frequency, and the time is not related.

When analyzing non-stationary signals, the instantaneous frequency changes with time, and it is not possible to accurately describe the changing frequency, the Fourier frequency loses its significance, so a new definition of the instantaneous frequency is needed to describe this change.

Let $X(t)$ be a signal of any time series, $Y(t)$ is the Hilbert transform of the signal, then $Y(t)$ can be expressed by $X(t)$ as follows.

$$Y(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{X(t)}{t-\tau}d\tau \tag{2}$$

Meanwhile, $X(t)$ can be expressed by $Y(t)$ as follows:

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

$$X(t) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{Y(t)}{\tau - t}d\tau \tag{3}$$

From the above two equations, we can know that $X(t)$ and $Y(t)$ are complex conjugate pairs, and $X(t)$ and $Y(t)$ are correlated with the time series, so we can get the following analyzed signals.

$$Z(t) = X(t) + jY(t) = A(t)e^{j\theta(t)} \tag{4}$$

$$A(t) = \sqrt{X^2(t) + Y^2(t)} \tag{5}$$

$$\theta(t) = \arctan\left(\frac{Y(t)}{X(t)}\right) \tag{6}$$

Where $A(t)$ is the instantaneous amplitude and $\theta(t)$ is the phase. another instantaneous parameter is known from the phase and frequency.

$$f(t) = \frac{1}{2\pi}\frac{d\theta(t)}{dt} \tag{7}$$

As can be seen from the above equation, the Hilbert transform obtains a unique function, $Y(t)$ is the convolution of the Hilbert transform with $X(t)$ and $\frac{1}{t}$, which emphasizes the limitation of the characteristics of $X(t)$ The three instantaneous parameters of the signal can be found out through the Hilbert transform, i.e., the instantaneous amplitude, the instantaneous phase, and the instantaneous frequency. Each of the parameters is analytic, thus, the Hilbert transform has a real-life instantaneous characteristic.

2. Intrinsic Modal Functions

In general, a data often contains more than one oscillation mode, a simple Hilbert transform can not be decomposed into all the frequencies of a signal, so the data must first be decomposed into the intrinsic modal function. the definition of the instantaneous frequency of a signal needs to have the following necessary conditions, firstly, in the entire data segment, the function is symmetric, and secondly, the number of zeros and the number of extrema are the same, and finally, the local mean value of the signal is zero. Finally, the signal is locally zero-mean.

N.E. Huang defined that the intrinsic modal function must satisfy the following two conditions.

One is that the number of zero crossing points and the number of extreme points in the signal data are the same or differ by at most one.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

One is that the average value of the upper and lower envelopes formed by the local extreme value points and the local extreme value points of the IMF at any moment is zero, i.e., the local signal is symmetric about the time axis.

3. Empirical modal decomposition method

Empirical modal decomposition (EMD) method is the essence of the Hilbert-Huang transform, this method is a non-stationary complex signal from the separation of several IMF process, the process is known as the screening process, screened out of the various components superimposed on the actual data series. at any time, most of the signals are unable to meet the conditions of the IMF, so it is necessary to screen the signal first and then use the Hilbert transform on the signal. At any moment, most of the signals cannot meet the conditions of IMF, so the signals need to be filtered first and then processed by Hilbert transform for each IMF component, each IMF can be linear or nonlinear.

The decomposition process is based on three assumptions: one is that the time-domain characteristics are determined by the interval between the maxima and minima, two is that the original signal should have at least one maxima and one minima, three is that if the original signal contains only inflection points, the maxima and minima can be calculated by taking the first derivative or multiple derivatives of the signal, and then the signal can be reduced by integrating the signal. The specific steps of the empirical mode decomposition are as follows.

The first step is to write the original signal as $X(t)$, determine all the maximum and minimum value points of $X(t)$, and fit the maximum value points to the upper envelope with the three times spline sampling function, and the minimum value points to the lower envelope with the three times spline sampling function, and the upper and lower envelopes should contain all the data.

In the second step, the average value of the upper and lower envelopes is calculated and denoted as m_1, and the original signal X(t), is subtracted from m_1 to obtain a new sequence h_1.

$$h_1 = X(t) - m_1 \tag{8}$$

In this decomposition, the low-frequency quantities of the signal are separated out, and h_1 is the high-frequency quantity of the signal. Ideally, if h_1 satisfies the two necessary conditions of IMF, then h_1 is the first IMF component of the original signal $X(t)$, and is denoted as $c_1 = h_1$.

In the third step, if the high-frequency quantity h_1 does not satisfy the two necessary conditions of IMF, then h_1 is continued as the original signal, repeat the above two steps, first calculate the average value of h_1, denoted as m_{11}, and then calculate $h_{11} = h_1 - m_{11}$, and judge whether h_{11} satisfies the two necessary conditions of IMF, if it does, then h_{11} is denoted as c_1, if it doesn't, then continue to

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

calculate, and repeat the cycle of the first and the second steps, until the obtained h_{1k} satisfies the two necessary conditions and is denoted as $c_1 = h_{1k}$, and c_1 is the first IMF component of the original signal $X(t)$.

In the fourth step, c1, is separated from the original signal $X(t)$ to obtain a signal that removes the high-frequency quantityv c_1, denoted as r_1.

$$r_1 = X(t) - c_1 \tag{9}$$

The r_1 signal as the original signal, repeat the above steps, until we get a component that meets the two necessary conditions of IMF, which is recorded as c_2, and have been so looped n times, we can get the n IMF components of the original signal $X(t)$, as shown in Eq. (10).

$$\begin{cases} r_1 - c_2 = r_2 \\ \vdots \\ r_{n-1} - c_n = r_n \end{cases} \tag{10}$$

In the fifth step, the loop to the end, there is a residual r_n, the end of the loop termination conditions for the residual r_n is a monotonic function, that is, can no longer be extracted from it to meet the two necessary conditions of the IMF components, the end of the decomposition. the final decomposition of the form as shown in Equation (11).

$$X(t) = \sum_{i=1}^{n} c_i(t) + r_n \tag{11}$$

Where c_1 is the IMF of the original signal and r_n is a residual that converges to a constant.

2.1.2. IMPROVEMENT OF ENVELOPE FITTING ALGORITHM

When the upper and lower envelopes of the signal are approximately symmetric, it is inappropriate to use the upper and lower envelopes to find the average envelope, and for the problems of overshoot, undershoot and sieving speed in the process of EMD, this paper proposes to firstly use the tangent mean point to densify the local extreme point, and then use the segment power function to interpolate and fit the tangent mean point and local extreme point, and finally splice the interpolated curve segments together, which is the final average envelope. Then the segmented power function is used to fit the interpolation of the tangential mean and local extreme points, and finally the interpolated curve segments are spliced together, which is the final mean envelope.

1. Tangential mean points

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

Since the density of the distribution of the points to be interpolated directly determines the interpolation accuracy and fitting effect, the densification of the extreme points by the tangent mean points can optimize the fitting effect, and then inhibit or eliminate the overshooting and undershooting phenomena in the envelope fitting.

All the local extreme points are arranged in a time series, notated as $p\left(t_1, y_1\right), p\left(t_2, y_2\right), \ldots, p\left(t_n, y_n\right)$, where t_i is the moment when the extreme point appears, and y_i is the amplitude of the extreme point. Let $p\left(t_{i-1}, y_{i-1}\right), p\left(t_i, y_i\right), p\left(t_{i+1}, y_{i+1}\right)$ be the three neighboring extreme points, and the tangent touching the mean point is shown in Eq. (12).

$$P\left(t_i, y_i\right)=\frac{1}{2}\left[p\left(t_i, y_i\right)+\frac{t_i-t_{i-1}}{t_{i+1}-t_{i-1}} p\left(t_{i+1}, y_{i+1}\right)+\frac{t_{i+1}-t_i}{t_{i+1}-t_{i-1}} p\left(t_{i-1}, y_{i-1}\right)\right] \quad (12)$$

Where, $p\left(t_i, y_i\right)$ is the extreme point of the signal, $P\left(t_i, y_i\right)$ denotes the tangential mean point.

2. Segmental power function interpolation

Set the point to be interpolated as $P\left(t_1, y_1\right), P\left(t_2, y_2\right), \ldots, P\left(t_n, y_n\right), y=f(x)$ as the interpolation function, and use the segmented power function to interpolate any three neighboring points $P\left(t_{i-1}, y_{i-1}\right), P\left(t_i, y_i\right), P\left(t_{i+1}, y_{i+1}\right)$.

$$Q_i=\frac{\left(t_{i+1}-t_{i-1}\right)\left(y_{i-1}-y_i\right)-\left(t_{i-1}-t_i\right)\left(y_{i+1}-y_{i-1}\right)}{t_{i+1}-t_{i-1}}. \quad (13)$$

The interpolation function $f_i(t)$ satisfies equation (14).

$$\begin{cases} f_i(t)=\left(\frac{t-t_i}{t_{i-1}-t_i}\right)^\beta Q_i+\frac{y_{i+1}-y_{i-1}}{t_{i+1}-t_{i-1}}\left(t-t_i\right)+y_i, t \leq t_i, \\ f_i(t)=\left(\frac{t-t_i}{t_{i+1}-t_i}\right)^\beta Q_i+\frac{y_{i+1}-y_{i-1}}{t_{i+1}-t_{i-1}}\left(t-t_i\right)+y_i, t \geq t_i . \end{cases} \quad (14)$$

The value curve is shown in equation (15).

$$f_{i, i+1}(t)=\frac{t_{i+1}-t}{t_{i+1}-t_i} f_i(t)+\frac{t-t_i}{t_{i+1}-t_i} f_{i+1}(t) . \quad (15)$$

$f(t) IMF$

Finally, the interpolated line segments are spliced together to obtain the mean envelope, which is constructed by the segmented cubic function after many experiments and comparisons.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

2.1.3. IMPROVED HHT ALGORITHM FLOW

The theory of the HHT algorithm and the operation of the specific improvement measures have been introduced, the improved HHT algorithm flowchart is shown in Fig. 1. firstly, after inputting the signal $x(t)$, find out all the local maxima and minima contained in the signal $x(t)$, and then use the cubic spline interpolation method to construct the envelope of the local maxima and minima, and then find out the average value of the envelope. If the $f(t)$ meets the condition of IMF component, then find out the energy of the remaining signal and proceed to the next step, if not, then re-screen repeatedly until the condition of IMF can be determined, and repeat the above steps until there is only one extreme point, then stop iterating.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

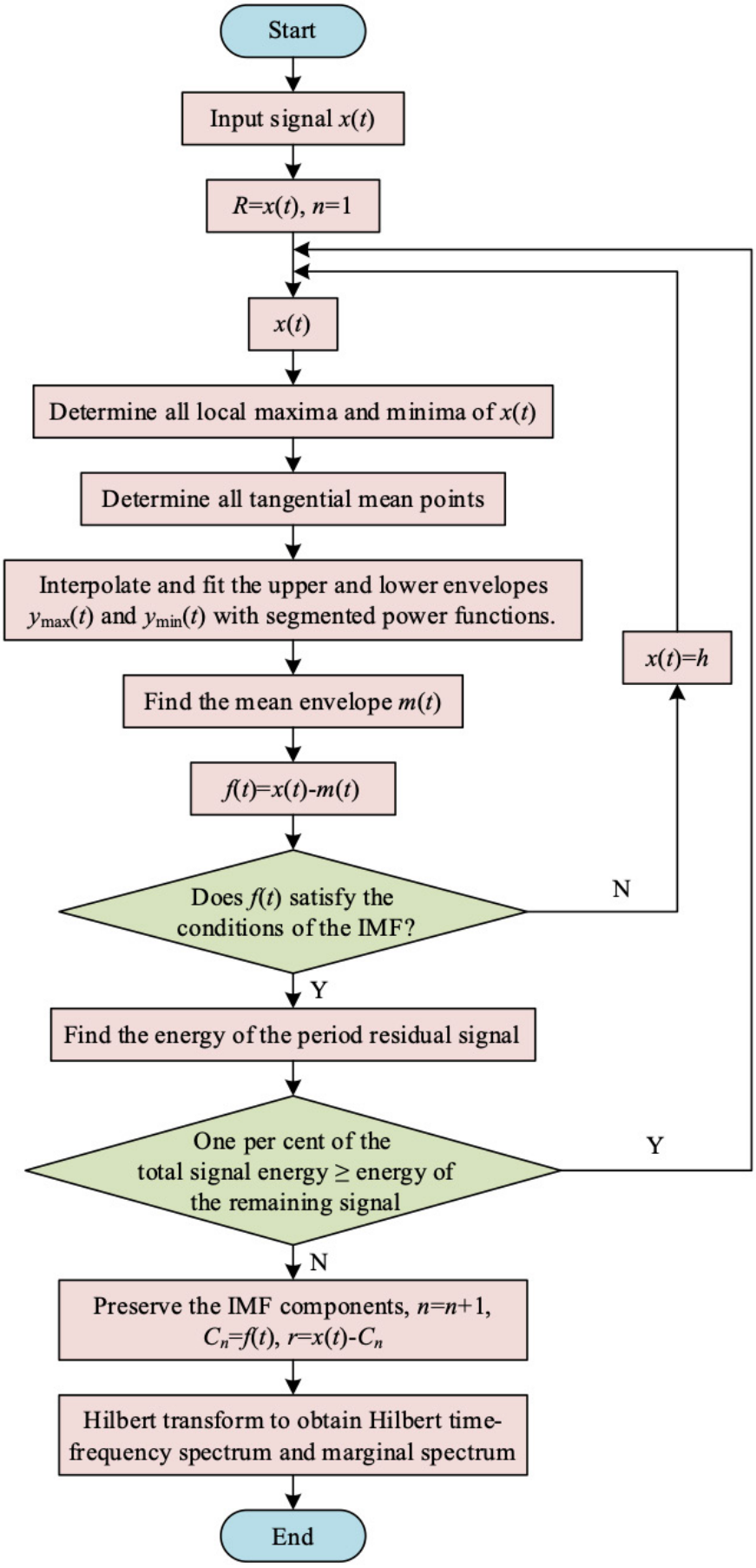


Figure 1 HHT algorithm improves the process

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

2.2. CONSTRUCTION OF WAVELET-BASED SINGULAR ENTROPY SUM AND S-SVM NETWORK MODELS

2.2.1. SEMIPARAMETRIC SUPPORT VECTOR MACHINES

Semi-parametric support vector machine, S-SVM, perfectly combines the advantages of parametric and non-parametric support vector machines, and improves the computational efficiency. Compared with linear support vector machine, S-SVM can handle non-linear sample data, and has better classification performance than non-linear support vector machine using soft intervals. S-SVM can avoid the problem of slow classifier computation caused by a large number of support vectors by using a predefined model. S-SVM avoids the slow computation problem caused by a large number of support vectors by using a predefined model.

The predefined model selects a representative set of basic elements $C = \{c_1, c_2, \ldots, c_m\}$ from the sample data to reduce the size of the sample data set, and estimates the normal vector ω, and then calculates the discriminant function to construct the classifier. The expression of the normal vector estimated according to Equation (16) is as follows.

$$\omega \simeq \sum_{i=1}^{m} \beta_i \varphi(c_i) \tag{16}$$

Where $\varphi(\cdot)$ is the mapping function, β is similar to the Lagrange multiplier, and the objective function is similar to Eq. (17), denoted as.

$$\begin{gathered} \min_{\beta} \frac{1}{2} \beta^T K_C \beta + C \sum_{i=1}^{m} \xi_i \\ \text{s.t. } y_i\left(\beta^T K_{ij} + b\right) \geq 1 - \xi_i \forall i = 1, 2, \ldots, m \\ \xi_i \geq 0, i = 1, 2, \ldots, m \end{gathered} \tag{17}$$

Where the matrix K_C denotes the $n \times n$ kernel matrix of the basic elements, i.e., $\left(K_C\right)_{i,j} = k\left(c_i, c_j\right), \forall i, j = 1, 2, \ldots, n, K_{ij} = k\left(x_i, c_j\right)$. The discriminant function is constructed as described in the previous section.

$$f(x_i) = \sum_{j=1}^{m} \beta_j k\left(x_i, c_j\right) + b \tag{18}$$

It can be seen that by limiting the size of the basic element set m, the complexity and operation speed of the classifier can be effectively limited. However, the element set C plays an important role for this kind of classifier, which has a great influence on

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

the classification performance and accuracy, and a processing tool that can accurately select the element set C is needed.

2.2.2. SEMI-PARAMETRIC SUPPORT VECTOR MACHINES BASED ON LEAST SQUARES

For the selection of the basic element set, this subsection adopts the sparse greedy matrix approximation, SGMA algorithm, and introduces the iterative reweighted least squares algorithm, IRWLS, to calculate the weights of the kernel function.

The kernel function used in this paper is Gaussian kernel function, as shown in Eq. (19).

$$k\left(x_i, x_j\right) = \exp\left(-\frac{\| x_i - x_j \|^2}{2\sigma^2}\right) \tag{19}$$

The kernel matrix K of the sample data set $T = \left\{x_1, x_2, \ldots, x_n\right\}$ is denoted.

$$K = \begin{bmatrix} k\left(x_1, x_1\right) & \cdots & k\left(x_1, x_j\right) & \cdots & k\left(x_1, x_n\right) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ k\left(x_i, x_1\right) & \cdots & k\left(x_i, x_j\right) & \cdots & k\left(x_i, x_n\right) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ k\left(x_n, x_1\right) & \cdots & k\left(x_n, x_j\right) & \cdots & k\left(x_n, x_n\right) \end{bmatrix} \tag{20}$$

In this paper, we use the hinge loss function to construct a soft interval classifier.

$$\ell_{\text{hinge}}(y) = \max(0, 1 - y) \tag{21}$$

Where y is the predicted classification output.

2.2.3. SPARSE GREEDY MATRIX APPROXIMATION

Given that the semiparametric support vector machine needs to select a suitable basic element set C, this section introduces the sparse greedy matrix approximation algorithm to find the element set C. The basic element set Cselected by using the SGMA algorithm can represent the whole sample dataset with the most representative features, which is very useful for solving the support vectors, constructing classifiers, and improving the classification performance.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

For the sample dataset $T = \{x_1, x_2, \ldots, x_n\}$, a subset $N = \{x_1, x_2, \ldots, x_k\}$ is randomly selected; for optimal performance, the range of k needs to be larger than . $log0.05/log0.95 = 59$ The optimal set of basic elements $C = \{c_1, c_2, \ldots, c_m\}$ is selected by the elements in the subset N.

Assuming that the kernel function of the column vectors in the subset N is estimated to be $k(x_i, .)$, set a value $\lambda, k(x_i, .)$ to be a linear combination of λ and the kernel function $k(c_i, .)$, of the elements in C.

$$k(x_i, .) = \sum_{j=1}^{\kappa} \lambda_{ij} k(c_i, .) \tag{22}$$

Where λ_{ij} is noted as the weight value.

The approximation error of the weights λ_{ij} can be determined by the kernel function of the kernel matrix trace of the sample dataset T and the column vectors in the subset N as $k(x_i, .)$.

$$\mathrm{Err}(\lambda) = \mathrm{tr}K - \sum_{i=1}^{n} \sum_{j=1}^{m} \lambda_{ij} k(x_i, c_j) \tag{23}$$

After the new element c_{m+1} is added to the basic element set C by the SGMA algorithm, the new weight error is shown in Eqs. (24) and (25).

$$\begin{aligned} &\mathrm{Err}(\lambda^{n,m+1}) = \mathrm{Err}(\lambda^{n,m}) - \eta^{-1} \parallel K^{n,m} z - k_{Sn} \parallel^2 \\ &(k_{NC})_i = k(x_i, c_{m+1}) \\ &z = K_c^{-1} \cdot k_{mc} \\ &\eta = 1 - z^T k_{mc} \\ &(k_{mC})_i = k(c_i, c_{m+1}) \end{aligned} \tag{24}$$

$$ED = \mathrm{Err}(\lambda^{n,m}) - \mathrm{Err}(\lambda^{n,m+1}) \tag{25}$$

K_C is the $m \times m$ kernel matrix of the basic element K, and $K^{n,m}$ is the $n \times m$ kernel matrix of the subset N and the basic element set C.

From the above analysis, the SGMA algorithm can calculate the element with the maximum ED value in the data set by the formula (25), and add this metamethod to the basic data set C, so as to find out a group of subsets that can more accurately represent the distribution of the whole feature space.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

2.2.4. ITERATIVE REWEIGHTED LEAST SQUARES

The iterative reweighted least squares, IRWLS, process is a solution method that has been widely used in the field of support vector machines. Compared with the weighted least squares algorithm, IRWLS can gradually correct for the effects of anomalous sample data and set the weights always within a relatively optimal range.

For the Lagrangian function with the penalty factor C_p added.

$$L(\omega, b, a, \xi, \mu) = \frac{1}{2} \parallel \omega \parallel^2 + C_p \sum_{i=1}^{n} \xi_i + \sum_{i=1}^{n} a_i \left(1 - \xi_i - y_i\left(\omega^T \varphi(x_i) + b\right)\right) - \sum_{i=1}^{n} \mu_i \xi_i \tag{26}$$

By taking the partial derivation of Eq. (26) and eliminating the term about ξ_i and calculating $C_p = a_i + \mu_i$, Eq. (27) is converted as shown below.

$$\begin{aligned} L(\omega, b, a, \xi, \mu) &= \frac{1}{2} \parallel \omega \parallel^2 + \sum_{i=1}^{n} a_i \left(1 - y_i\left(\omega^T \varphi(x_i) + b\right)\right) \\ &= \frac{1}{2} \parallel \omega \parallel^2 + \frac{1}{2} \sum_{i=1}^{n} \frac{2a_i}{1 - y_i\left(\omega^T \varphi(x_i) + b\right)} \left(y_i\left(\omega^T \varphi(x_i) + b\right)\right)^2 \\ &= \frac{1}{2} \parallel \omega \parallel^2 + \frac{1}{2} \sum_{i=1}^{n} a_i e_i^2 \end{aligned} \tag{28}$$

In which:

$$e_i = y_i\left(\omega^T \varphi(x_i) + b\right) \tag{29}$$

$$\alpha_i = \frac{2a_i}{1 - y_i\left(\omega^T \varphi(x_i) + b\right)} \tag{30}$$

e_i is denoted as the error of the sample data points, and α_i is denoted as the weight coefficient associated with it. Substituting $\omega \simeq \sum_{i=1}^{n} \beta_i \varphi(x_i)$ into Eq. (30) to obtain Eq. [112] with respect to α_i, the objective function is obtained.

$$\text{s.t.e } e_i = y_i - \left(\sum_{j=0}^{m} \beta_i K(x_i, c_i) + b\right) \tag{31}$$

$$\alpha_i = \begin{cases} 0, & y_i e_i \leq 0 \\ \dfrac{C_p}{e_i y_i}, & y_i e_i > 0 \end{cases} \tag{32}$$

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

Determine the weight β through equation (33).

$$\begin{bmatrix} K_C + K_{sC}^T D_a K_{sc} & K_{sc}^T D_\alpha 1 \\ 1^T D_a K_C^{-1} & 1^T D_a 1 \end{bmatrix} \begin{bmatrix} \beta \\ b \end{bmatrix} = \begin{bmatrix} K_{sc}^T D_\alpha y \\ 1 D_a y \end{bmatrix} \tag{33}$$

Where $\left(K_C\right)_{ij} = k\left(c_i, c_j\right), K_{SC} = k\left(x_i, c_j\right), \left(D_\alpha\right)_i = \alpha_i$, vectors $1 = [1, \ldots, 1]^T$, $y = \left[y_1, y_2, \ldots, y_n\right]^T$

The iterative process allows each round of training and learning to gradually correct the weight values, so that the weights α_i and β eventually converge to a fixed value, and the iterative calculation steps are as follows.

Step 1: Pre-set a weight α_i, solve the least squares problem to obtain β.

Step 2: Based on the correlation between weight BB and α_i, re-calculate the value of α_i by calculating weight β.

Step 3: Repeat the first two steps until the weights β and α_i converge to a fixed convergence value.

At this point, the discriminant function of the classifier is:

$$f\left(x_i\right) = \sum_{j=1}^{m} \beta_j k\left(x_i, c_j\right) + b \tag{34}$$

2.2.5. WAVELET TRANSFORM

The continuous wavelet transform of the function $f(t)$ is shown in equation (35).

$$W(f, a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) \psi^* \left(\frac{t-b}{a} \right) dt \tag{35}$$

Where, a is the scale factor, b is the translation factor and $\psi(t)$ is the mother wavelet.

Continuous wavelet transform can accurately extract the characteristics of the signal, but in each possible scale discrete points to calculate the wavelet coefficients, will be a huge project. if only a small part of these scales, and part of the time point, will greatly reduce the workload, and without loss of accuracy, the use of such an approximation will be obtained by the discrete wavelet transform, the definition of which is shown in equation (36).

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

$$\mathrm{DWT}(f,m,n)=\frac{1}{\sqrt{a_0^m}}\sum_k f(k)y^*\left(\frac{n-ka_0^m}{a_0^m}\right) \tag{36}$$

In Eq. (36), the parameters a and b of continuous wavelet transform are replaced by a_0^m and ka_0^m, k and MM are integers, and A02 is taken in general. The wavelet decomposition tree of the discrete wavelet transform is shown in Fig. 2. The original signal sequence of SN is shown in Fig. 2. CA_k denotes the low-frequency coefficients of the decomposition of the k layer, and CD_k denotes the high-frequency coefficients of the k decomposition. The signal components obtained from the decomposition of the decomposition coefficients of each layer by the single-branch reconstruction are denoted as a_k and D_k, and then the original signal $S(n)$ is the sum of the signal components obtained by the reconstruction, and the formula is as follows.

$$x(n)=D_1+A_1=D_1+D_2+A_2=\sum_{j=1}^{n}D_j+A_n \tag{37}$$

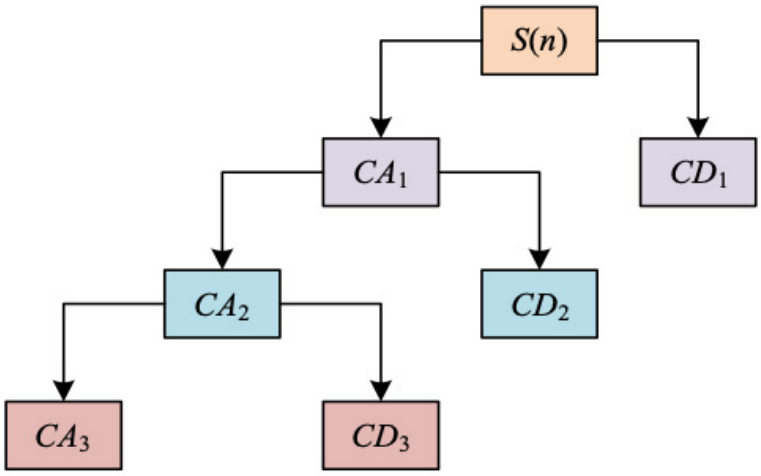


Figure 2 The wavelet decomposition tree

2.2.6. SHANNON INFORMATION ENTROPY AND WAVELET SINGULAR ENTROPY

Shannon information entropy and its state characteristics can be expressed by a variable X that takes a finite number of values, the probability that the state characteristics of the source take the value of x_j is $p_j=P\left\{X=x_j\right\}j=1,\cdots,L$ and $\sum_{j=1}^{L}p_j=1$, then $I_j=\log\left(1/p_j\right)$ represents the information obtained from the j result of X, and the information entropy is defined as shown in Equation (38).

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

$$H(X) = -\sum_{j=1}^{L} p_j \log\left(p_j\right) \tag{38}$$

Wavelet singular entropy is a new data processing method obtained by combining wavelet transform with Shannon's information entropy theory, which is defined as follows.

$$W_k = \sum_{i=1}^{k} \Delta p_i \tag{39}$$

In Eq. (40), Δp_i is the incremental wavelet singular entropy of order j.

$$\Delta p_i = -\left(\lambda_i / \sum_{j=1}^{l} \lambda_j\right) \log\left(\lambda_i / \sum_{j=1}^{l} \lambda_j\right) \tag{40}$$

The $\lambda_i (i = 1, \cdots, l)$ is calculated as follows.

After the signal $S(n)$ wavelet decomposition and reconstruction of the components at the $j(j = 1, \cdots m)$th scale is $D_j(n)$, then the m components of the signal $S(n)$ can form a $m \times n$ matrix $D_{m\times n}$, by the theory of signal singular value decomposition, for the above matrix $D_{m\times n}$, there exists a $m \times l$-dimensional matrix U and a $l \times l$ -dimensional diagonal matrix Λ and a $l \times n$-dimensional matrix V, so that the matrix Dmxn decomposition is as shown in Eq. (41).

$$D_{m\times n} = U_{m\times 1} \Lambda_{I\times 1} V_{l\times n} \tag{41}$$

Where, $\lambda_i (i = 1, \cdots, l)$ is the element on the main diagonal of the diagonal matrix Λ, i.e., the singular value of the matrix $D_{m\times n}$ formed after the signal $S(n)$ is decomposed by the wavelet decomposition of the m layer and reconstructed.

2.2.7. CABLE FAULT DETECTION STRATEGY BASED ON WAVELET SINGULAR ENTROPY AND S-SVM MODELS

In this section, wavelet singular entropy and semiparametric support vector machine are combined and applied to cable fault detection, and the cable fault detection strategy based on wavelet singular entropy and S-SVM model is shown in Fig. 2. firstly, the cable current signal is extracted, then wavelet singular entropy is used to decompose the signal, extract the meaningful sub-signal components, and then the correlation analysis is used to filter the components to reconstruct the fault signal, and then wavelet singular entropy is used to reduce the influence of non-Gaussian noise on the fault signal, extract the wavelet energy entropy of harmonic components as feature vectors, and then the fault sample data is inputted into the S-SVM model to be trained. Then the wavelet singular entropy is used to reduce the

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

effect of non-Gaussian noise on the fault signal, and the wavelet energy entropy of the harmonic components is extracted and used as the feature vectors. After that, the fault sample data are inputted into the S-SVM model for training, and the S-SVM model is finally used to recognize and diagnose the test samples.

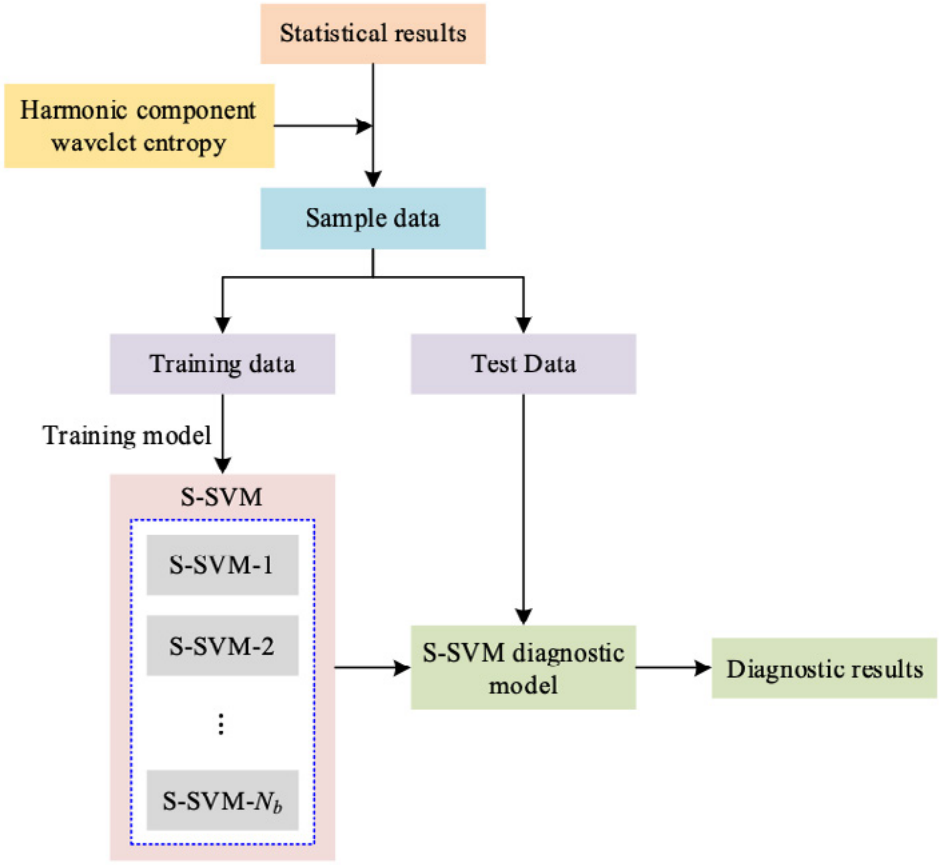


Figure 3 Cable fault detection strategy based on wavelet singular entropy and S-SVM model

3. ANALYSIS OF INTELLIGENT CABLE FAULT DETECTION RESULTS CONSIDERING CURRENT HARMONIC FEATURES

3.1. ANALYSIS OF INTELLIGENT DETECTION RESULTS OF CABLE LINE DETERIORATION FAULTS

3.1.1. HARMONIC FEATURE EXTRACTION FOR CABLE LINE DEGRADATION

In this section, according to the structural characteristics of the cable, 40,000 sets of harmonic diagnostic data of the same power cable are selected, and the harmonic content of the main part of the cable from the 2nd to the 10th harmonic is multiplied with its corresponding contribution rate, and 9 harmonic vectors are obtained as the

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

input data, and the degree of deterioration of the insulation, shielding, protective layer and cable joints are derived through the improved HHT transformation model. The energy spectrum of the harmonic vectors of different parts of the cable is shown in Fig. 4, and the relative energy of each harmonic is 1. The relative energies of the harmonic vectors are obviously different in diagnosing the operation status of different parts of the cable, the operation status of the cable insulator mainly depends on the change of the 2nd harmonic vector, with the relative energy value of 0.32, and the operation status of the shielding layer mainly depends on the change of the 2nd, 3rd, and 5th harmonic vectors, with the relative energy value of 0.24, 0.25, 0.2, 0.5, and 0.6, respectively. 0.25 and 0.2 respectively, and the operating state of the protective layer and the roving state of the cable joints mainly depends on the changes of the 2nd, 4th, 7th, 8th and 9th harmonic vectors, with the relative energy values of 0.3, 0.25 and 0.26, 0.26, 0.3 respectively. The harmonic vectors obtained based on the improved HHT transformation model completely characterize the operating state of the different parts of the cables.

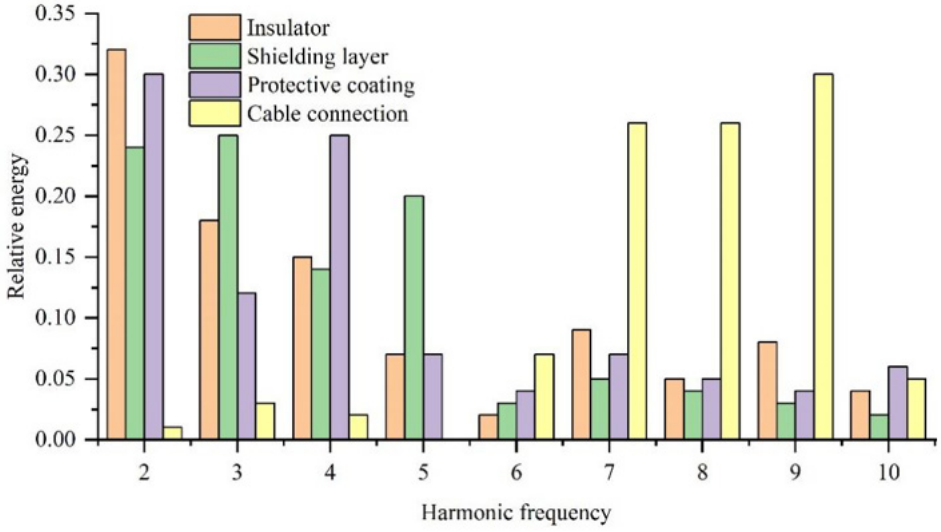


Figure 4 The shoe wave vector energy spectrum of the different parts of the cable

3.1.2. ANALYSIS OF CABLE LINE DETERIORATION IDENTIFICATION TEST RESULTS

In this section, based on the wavelet singular entropy and S-SVM model to identify and detect the degradation of cable lines, in order to verify the accuracy of the wavelet singular entropy and S-SVM model, the harmonic characteristics database of different fault loss currents of cables is formed by taking the harmonic total aberration rate, the fundamental content and the harmonic contents of each harmonic as the characteristics.Firstly, we extracted the sample data of the loss currents in the database to form the sample data set of the wavelet singular entropy and S-SVM model. Then some data in the sample data set are randomly extracted as the training sample set, and the rest of the data in the sample data set are used as the test sample set. finally, different ratios of the training set to the test set are set, and

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

different numbers of test samples are taken to identify the deterioration of cables with the wavelet singular entropy and the S-SVM model trained in the above steps. red part indicates the classification of the predicted test set, and blue part indicates the classification of the actual test set. red part indicates the classification of the predicted test set, and blue part indicates the classification of the actual test set. red part indicates the classification of the actual test set. The red part indicates the classification of the predicted test set, and the blue part indicates the classification of the actual test set, and the coincidence of the red part and the blue part indicates that the test set data matches with the prediction result, and the prediction result is accurate, and vice versa.The accuracy results of the wavelet singular entropy and S-SVM models with different ratios of the training set to the test set are shown in Fig. 5, and the results of the accuracy results of the models with the ratio of the training set to the test set 6:4, 7:3, 8:2, and 9:1 are shown in Figs. 5(a) to (d) respectively. Vertical coordinates 1, 2, 3 and 4 indicate the degradation of insulator, shield, protective layer and cable connector, respectively.When the ratio of training set to test set is 6:4, there are 12 groups of prediction failures, 6 groups of samples predicted the protective layer to be insulator, and 4 groups of samples predicted the shield to be insulator, with an identification accuracy of 93.93%.When the ratio of training set to test set is 7:3, there are 3 groups of samples predicted the protective layer sample to be insulator, with an identification accuracy of 93.93%. When the ratio of training set to test set is 7:3, there are 3 groups of protective layer samples predicted to be shielding layer and 2 groups of shielding layer samples predicted to be insulators, the recognition accuracy is 94.38%. when the ratio of training set to test set is 8:2, there are 1 group of protective layer and 1 group of shielding layer samples predicted to be insulators, the recognition accuracy is 96.58%. when the ratio of training set to test set is 9:1, there is only 1 group of protective layer sample predicted to be insulators, the recognition accuracy is 97.36%. The more training samples, the higher the accuracy of wavelet singular entropy and S-SVM model in recognizing different cable line degradation, and the recognition rate is above 93%, which indicates that this model can well identify the categories of cable line degradation.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

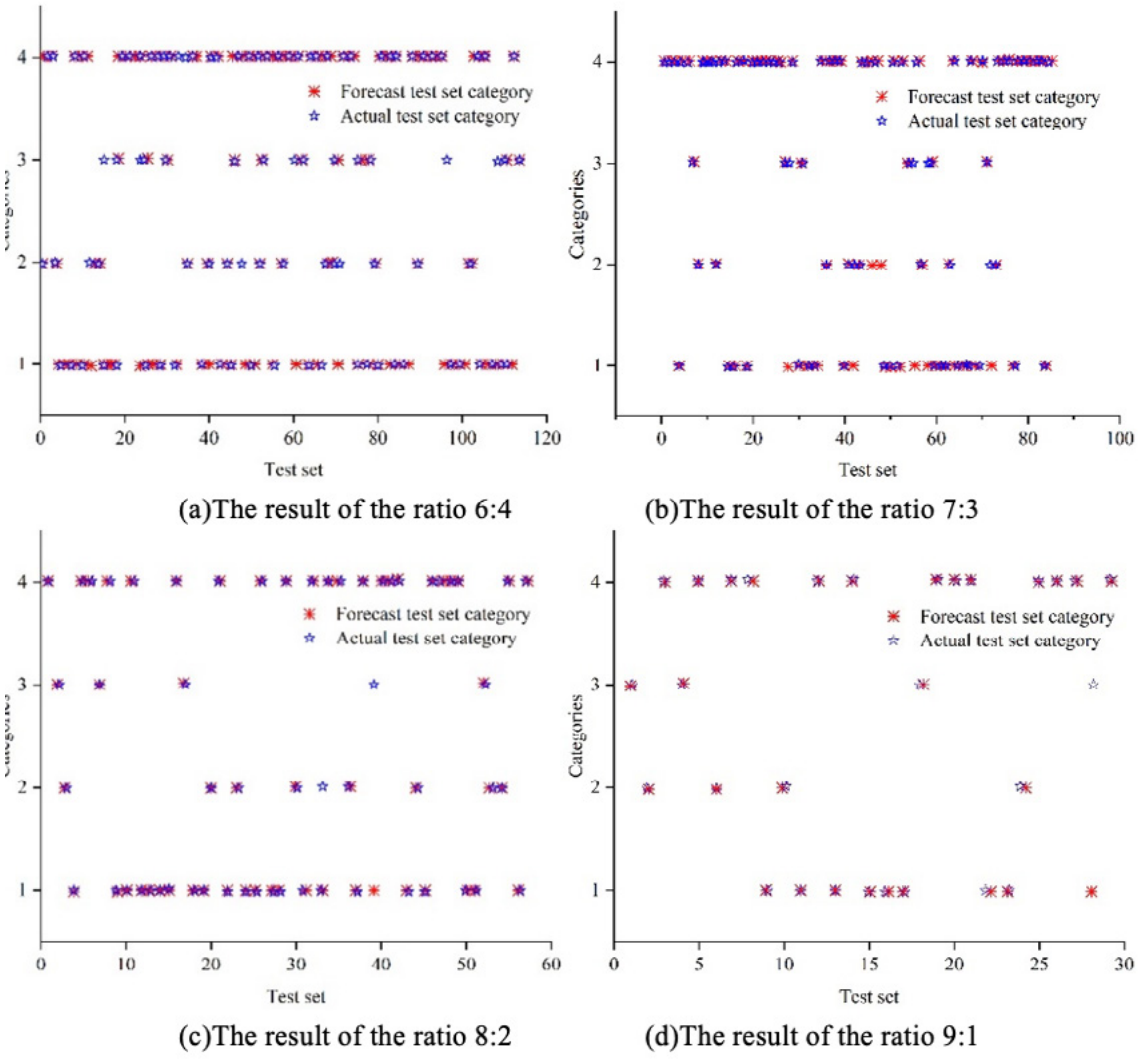


Figure 5 Results of different training sets and test sets

3.2. ANALYSIS OF INTELLIGENT DETECTION RESULTS OF CABLE SHORT-CIRCUIT FAULTS

In order to provide data support for cable fault diagnosis, this paper adopts the 10KV cable line model to construct the fault data set, and utilizes the electrical components in the tool library to build the simulation circuit.

3.2.1. CABLE SHORT CIRCUIT FAULT ANALYSIS

Cable short-circuit fault is the largest proportion of cable faults occurring, cable a phase connected to another phase or one of the phases connected to the earth is called a short-circuit fault, in the case of mixed faults do not take into account the occurrence of 10 types of faults may occur: A_G, B_G, C_G were A-phase, B-phase and C-phase ground faults, AB, AC, BC, respectively, on behalf of AB, AC two-phase and BC two-phase short-circuit faults, AB_G, AC_G, BC_G were AB-phase, AC phase and BC two-phase short-circuit ground faults, ABC is a three-phase short-circuit fault,

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

ABC is a three-phase short circuit. AB_G, AC_G, BC_G are AB phase, AC phase and BC two-phase short-circuit ground faults, and ABC is three-phase short-circuit faults.Figure 6 shows the current waveform curves of the cable in normal operation and various short-circuit faults, and Fig. 6(a)~(d) shows the current waveform curves of the cable in normal operation, the single-phase grounded short-circuit faults, and two-phase indirect short-circuit current waveforms, and Fig. 6(a)~(d) shows the current waveform curve of the cable in normal operation, single-phase ground faults, and two-phase short-circuit faults, and Fig. 6(a)~(d) shows the current waveform curves of the cable in normal operation. Indirect ground short-circuit current waveform curve, three-phase ground short-circuit waveform curve. cable normal operation current waveform is more symmetrical, the overall current between -10K-10K amperes. cable short-circuit faults, short-circuit current is lower than the normal current. single-phase ground short-circuit, fault A-phase current in the 0.07s current is smaller than the non-fault B, C-phase current, and the waveform is more chaotic. two-phase short circuit in the 0.075s, the AB phase current is lower than the non-fault C phase current, AB two short-circuit faults occur. three-phase short-circuit fault, in 0.075s, the three-phase current is lower than the normal current, and the three-phase current sum is equal to zero.

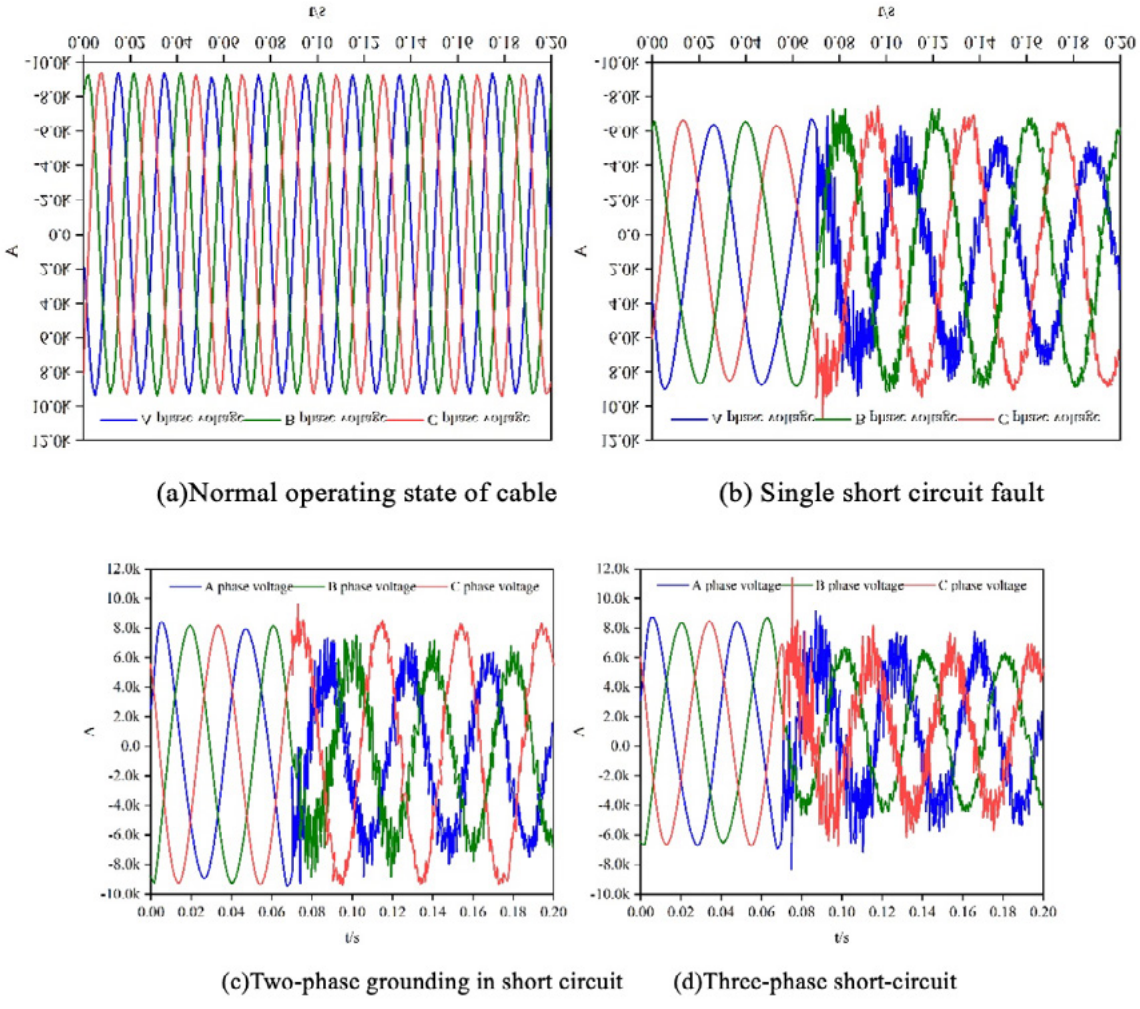


Figure 6 The current waveform of the cable's various circuited faults

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

3.2.2. HARMONIC FEATURE EXTRACTION FOR CABLE SHORT-CIRCUIT FAULTS

In this section of the experiment, the sampling frequency is set to 12.8kHz, the sampling time is 0.2s, and the fault is imposed after 0.05s of the system's normal operation state, each simulation samples 2780 points of voltage, each line has three phases, and each line collects 8670 points, i.e., the length of each sample is 8670, and then collects the operation state of 10 cables, 80% of them are used as training samples, 20% as test samples, and 18 harmonic vectors are obtained by using the improved HHT transform model. Figure 7 shows the energy spectrum of harmonic vector during various short-circuit faults of cable. 20% are used as test samples, and 18 harmonic vectors are obtained by using the improved HHT transform model. Fig. 7 shows the energy spectrum of harmonic vectors in various short-circuit faults of cables. In single-phase cable short-circuit, phase A mainly looks at the 1st and 9th harmonics, with the relative energy values of 0.18 and 0.12, respectively, and phases B and C mainly look at the 1st and 2nd harmonics, with the relative energy values of 0.19 and 0.12, 0.24 and 0.14, respectively. Two-phase cable short-circuit mainly looks at the 1st and 2nd harmonics, with the relative energy values of 0.19 and 0.12, 0.24 and 0.14, respectively. When two-phase cable is short-circuited, phase A mainly looks at the 18th harmonic with a relative energy value of 0.38, phase B mainly looks at the 13th and 16th harmonics with relative energy values of 0.17 and 0.14, and phase C mainly looks at the 15th harmonic with a relative energy value of 0.21. When three-phase cable is short-circuited, phase A mainly looks at the 2nd harmonic with a relative energy value of 0.05, phase B mainly looks at the 1st and 2nd harmonics with relative energy values of 0.19 and 0.12, and phase B mainly looks at the 1st and 2nd harmonics with relative energy values of 0.19 and 0.14, and phase C mainly looks at the 1st and 2nd harmonics with relative energy values of 0.19 and 0.14 respectively. The relative energy values are 0.19 and 0.13 for phase B, and 0.13 and 0.1 for phase C, mainly for the 2nd and 3rd harmonics.

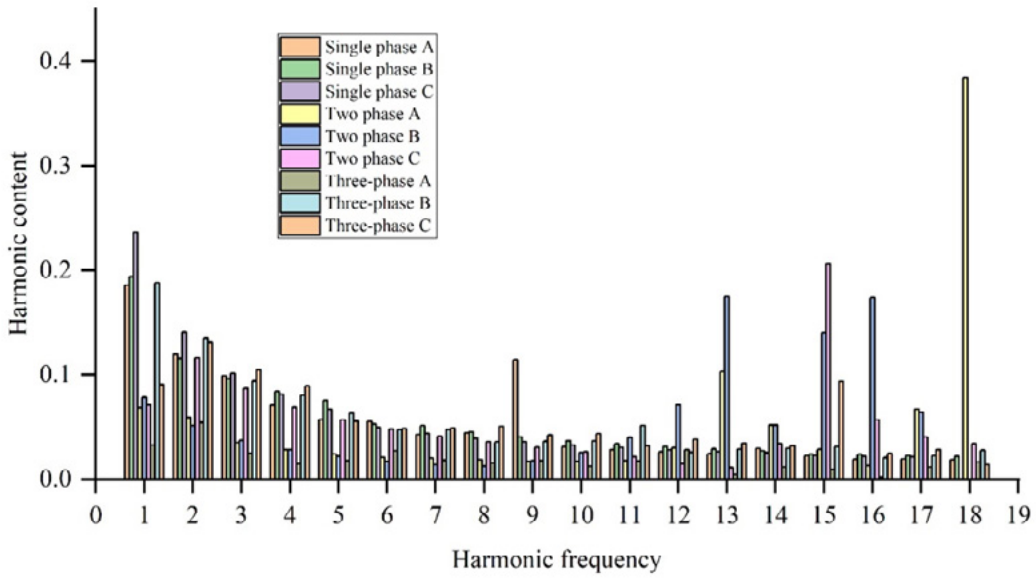


Figure 7 The harmonic energy spectrum of all kinds of short-circuit failures of the cable

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

3.2.3. IDENTIFICATION AND DETECTION RESULTS OF CABLE SHORT-CIRCUIT FAULTS

This section extracts local features of the input target based on wavelet singular entropy and S-SVM model, achieving target detection and improving diagnostic efficiency. Figure 7 shows the confusion matrix of the cable fault diagnosis model, where the vertical axis represents the true category of the cable fault signal and the horizontal axis represents the predicted category of the network for the cable fault signal, The diagonal values on the matrix represent the probability of correctly classifying cable fault signals. Wavelet singular entropy and S-SVM models can fully identify 10 types of cable short circuit faults. For A_ G, B_ G, C_ G, AB, AC, BC, AB_ G, AC_ G, BC_ The correct classification probabilities for 10 types of short-circuit faults, including G and ABC, are 0.94, 0.95, 0.96, 0.96, 0.98, 0.99, 0.94, 0.92, and 1, respectively. Although the correct classification probability for BC phase short-circuit grounding is relatively low, the overall probability is above 90%, indicating that the diagnosis of cable faults based on wavelet singular entropy and S-SVM model is completely feasible

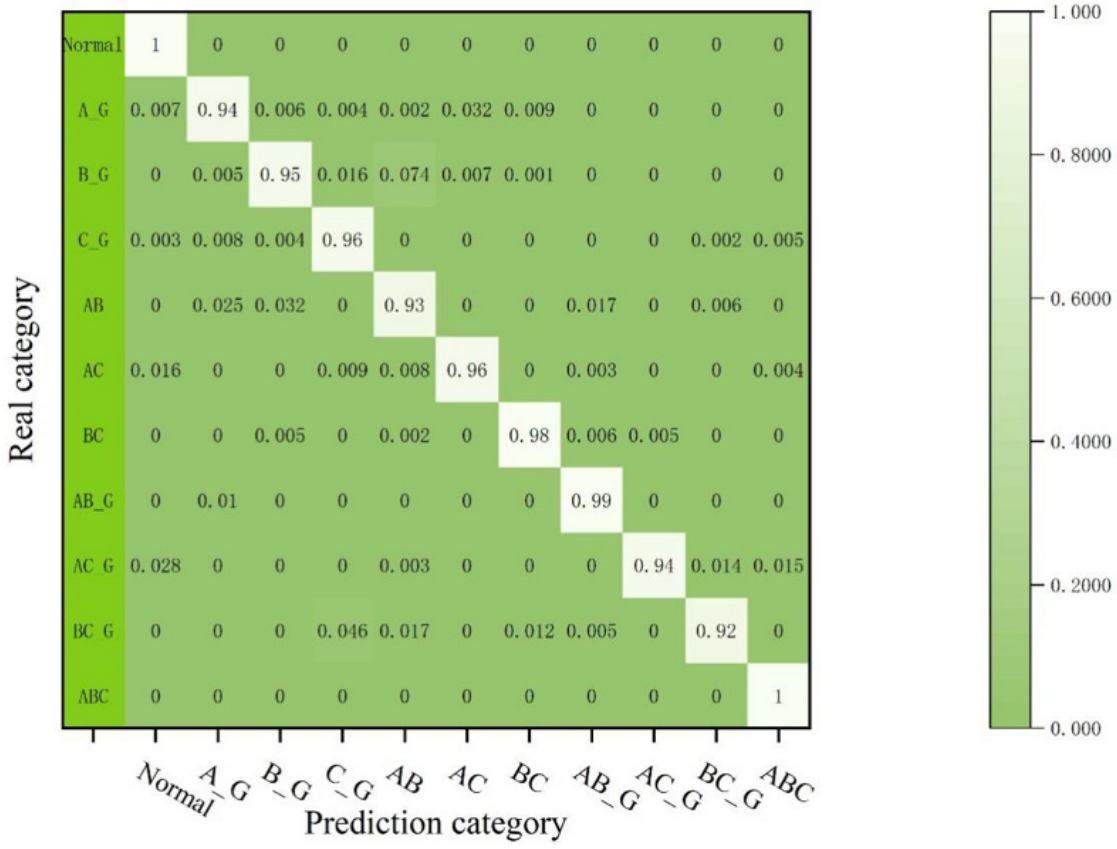


Figure 8 Confusion matrix of cable fault diagnosis model

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

3.3. ANALYSIS OF INTELLIGENT DETECTION RESULTS OF CABLE FAULTS

In order to verify the diagnostic effect of wavelet singular entropy and S-SVM model on all faults of cables, 180 samples are selected for test experiments in this section. The maximum error of wavelet singular entropy and S-SVM model on the test samples is 0.5329, only two samples have the absolute error value more than 0.5, and the other samples have the absolute error value lower than 0.5, and the average error is 0.1124.Among 180 sets of test samples, the correct rate of wavelet singular entropy and S-SVM model to diagnose various cable faults reached 98.04%, and the performance is good. In 180 sets of test samples, the correct rate of the wavelet singular entropy and S-SVM model in diagnosing various cable faults reaches 98.04%, which is good, and the wavelet singular entropy and S-SVM model can be used in the actual cable fault diagnosis.

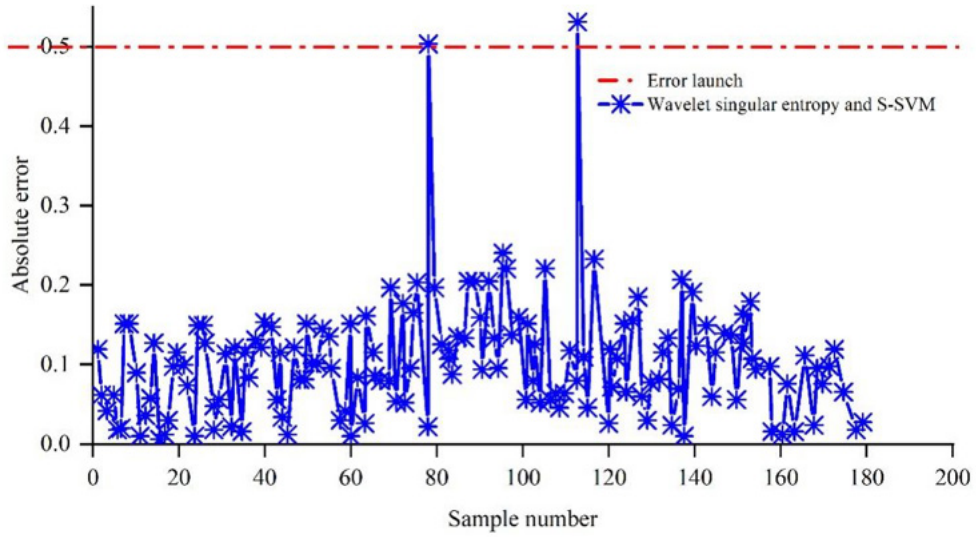


Figure 9 Wavelet singular entropy and S-SVM model diagnose cable fault curve

4. CONCLUSION

In this paper, by improving the envelope fitting algorithm of HHT algorithm, the HHT transform model is proposed to extract the current harmonic features of the cable. wavelet singularity algorithm and S-SVM algorithm are utilized to construct into the wavelet singular entropy and S-SVM model for detecting the faults of the cables. the experimental results are as follows.

In the detection of cable line deterioration, the current harmonic vectors characterize the operation status of different parts of the cable, and the operation status of cable insulator, shield, protective layer and cable joints are shown in the changes of 2nd, 2nd, 3rd, 5th, 2nd, 4th and 7th, 8th, 9th harmonic vectors, respectively, and the accuracy of the Wavelet Singularity Algorithm and S-SVM algorithm on the detection of the cable line deterioration is 93.93%, 94% and 94%, respectively. are 93.93%, 94.38%, 96.58% and 97.36%.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

In the detection of cable short circuit faults, the current of different short circuit faults in the cable is lower than the normal current, and in the case of three-phase short circuit faults, the three-phase currents add up to 0. For the extraction of the different short circuit current harmonics, in the case of a single-phase short circuit in the cable, the operating state of phases A, B, and C is shown in the first and ninth, first and second, and first and second harmonics of the current respectively, in the case of a two-phase short circuit in the cable, the operating state of phases A, B, and C is shown in the 18th, 13th, and 16th, and 15th harmonics of the current respectively. When three-phase cable is short-circuited, the operating states of phases A, B and C are characterized by 2, 1 and 2, 2 and 3 harmonic vectors, respectively. The wavelet singularity algorithm and S-SVM algorithm have an accuracy of more than 92% in identifying 10 different cable short-circuit faults, such as A_G, B_G, C_G, AB, AC, BC, AB_G, AC_G, BC_G, ABC, and so on.

Overall for cable fault detection, the wavelet singularity algorithm and S-SVM algorithm have reached 98.04% correct rate for detecting all kinds of pegged accounts of cables, and the error value is only more than 0.5 for two groups of data out of 180 groups of sample data. The improved HHT transform model and the wavelet singular entropy and S-SVM models proposed in this paper have high accuracy and practicability, and provide a a new method.

REFERENCES

(1) Otsu, H., Yone, Y., Takahashi, E., & Koike, K. (2017). Supply and demand situation of wood-chips before and after the increase in power-plant demand in the Chugoku region. Journal of Forest Economics, 63.

(2) Ros, A., & J. (2017). An econometric assessment of electricity demand in the United States using utility-specific panel data and the impact of retail competition on prices. Energy Journal.

(3) Nguyen, T. T., Lee, W. G., Kim, H. M., Yang, H. S., & Sciubba, E. (2020). Fault analysis and design of a protection system for a mesh power system with a co-axial HTS power cable. Energies, 13(1), 220.

(4) Zhang, Z.-h., Xu, B.-y., Crossley, P., & et al. (2018). Positive-sequence-fault-component-based blocking pilot protection for closed-loop distribution network with underground cable. International Journal of Electrical Power & Energy Systems.

(5) Zhang, Z., Chen, Q., Xie, R., & Ranran. (2019). The fault analysis of PV cable fault in DC microgrids. IEEE Transactions on Energy Conversion, 34(1), 486-496.

(6) Cozza, A. (2019). Never trust a cable bearing echoes: understanding ambiguities in time-domain reflectometry applied to soft faults in cables. IEEE Transactions on Electromagnetic Compatibility.

(7) Khavari, S., Dashti, R., Shaker, H. R., & Santos, A. (2020). High impedance fault detection and location in combined overhead line and underground cable distribution networks equipped with data loggers. Energies, 13.

(8) Popovi, L. M. (2018). Reduction of the fault current passing through the grounding system of an HV substation supplied by cable line. International Journal of Electrical Power & Energy Systems, 99, 493-499.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

(9) Nguyen, T. T., Lee, W. G., Kim, H. M., Yang, H. S., & Sciubba, E. (2020). Fault analysis and design of a protection system for a mesh power system with a co-axial HTS power cable. Energies, 13(1), 220.

(10) Peake, L. (2018, May). Making cable fault location even easier. Electrical Engineering.

(11) Baranowski, J., Grobler-Dbska, K., & Kucharska, E. (2021). Recognizing VSC DC cable fault types using Bayesian functional data depth. Energies, 14.

(12) Liu, X., & Wang, H. (2020). Analytical expression for estimation of the fault location in cable line. IET Science, Measurement & Technology, 14(8).

(13) Xuebin, Q., Yizhe, Z., Wang, M., Gang, D., Jun, G., & Pai, W. (2018). A cable fault recognition method based on a deep belief network. Computers & Electrical Engineering, 71, 452-464.

(14) Cataldo, A. C. G. (2021). A new measurement algorithm for TDR-based localization of large dielectric permittivity variations in long-distance cable systems. Measurement, 174(1).

(15) Wang, F., Yuan, G., Guo, C., & Li, Z. (2022). Research on fault diagnosis method of aviation cable based on improved Adaboost. Advances in Mechanical Engineering, 14.

(16) Sian, H. W., Kuo, C. C., Lu, S. D., & Wang, M. H. (2023). A novel fault diagnosis method of power cable based on convolutional probabilistic neural network with discrete wavelet transform and symmetrized dot pattern. IET Science, Measurement & Technology.

(17) Marriott, N. (2021, June). Megger launches cable fault locator. Electrical Engineering.

(18) Lowczowski, K., Lorenc, J., Zawodniak, J., & Dombek, G. (2020). Detection and location of earth fault in MV feeders using screen earthing current measurements. Energies, 13.

(19) Hu, C., Huang, Y., & You, Z. (2019). Ship cable fault info acquisition model based on auto recognition. Journal of Coastal Research, 94, 510-514.

(20) Lai, Q., Chen, J., Hu, L., Cao, J., & Zhu, N. (2020). Investigation of tail pipe breakdown incident for 110 kV cable termination and proposal of fault prevention. Engineering Failure Analysis, 108, 104353.

(21) Liu, N., Fan, B., Xiao, X., & Yang, X. (2019). Cable incipient fault identification with a sparse autoencoder and a deep belief network. Energies, 12(18), 3424.

(22) Wang, Y., Lu, H., Xiao, X., Yang, X., & Zhang, W. (2019). Cable incipient fault identification using restricted Boltzmann machine and stacked autoencoder. IET Generation Transmission & Distribution, 14(7).

(23) Qu, K., Zhang, W., Zhang, S., Xiao, X., Xi, Y., & Zhang, H. (2022). Model-free underground cable incipient fault location using two-terminal zero-sequence measurements. International Journal of Electrical Power & Energy Systems, 140, 108057.

(24) Kwon, G. Y., Lee, C. K., Lee, G. S., Lee, Y. H., Chang, S. J., & Jung, C. K. (2017). Offline fault localization technique on HVDC submarine cable via time–frequency domain reflectometry. IEEE Transactions on Power Delivery

https://doi.org/10.17993/3ctecno.2024.v13n1e45.99-128

/06/

POWER CABLE FAULT LOCATION TECHNIQUE BASED ON PARAMETER OPTIMIZED VARIATIONAL MODAL DECOMPOSITION

Yuning-Tao*

- College of Electrical Engineering and New Energy, China Three Gorges University, Yichang, Hubei, 443000, China
- taoyuning12@163.com

Reception: 11 April 2024 | **Acceptance**: 6 May 2024 | **Publication**: 10 June 2024

Suggested citation:

Yuning-Tao (2024). **Power cable fault location technique based on parameter optimized variational modal decomposition.** 3C Tecnología. Glosas de innovación aplicada a la pyme, *13(1)*, 130-155. https://doi.org/10.17993/3ctecno.2024.v13n1e45.130-155

ABSTRACT

With the complexity of power systems and the increase of loads, power cable faults occur frequently, affecting the safety of power supply. A parameter optimization-based variational modal decomposition (VMD) technique combined with wavelet transform is used. Through in-depth analysis of cable fault signals, this study first applies VMD to decompose the signals, and then further analyzes the processed signals in combination with wavelet transform. The experimental results show that the method achieves a significant reduction in the average localization error in the simulated fault test, with an average error of less than 1%, which improves the localization accuracy by about 30% compared with the traditional method. In the processing of complex fault signals, the present method shows better adaptability and robustness. Overall, this study not only improves the accuracy of fault localization, but also provides a new technical path for the diagnosis and maintenance of power system faults.

KEYWORDS

Power cable fault location, variational modal decomposition, parameter optimization, wavelet transform

https://doi.org/10.17993/3ctecno.2024.v13n1e45.130-155

INDEX

https://doi.org/10.17993/3ctecno.2024.v13n1e45.130-155

1. INTRODUCTION

Cables are widely used in power distribution networks and play a significant role in power supply reliability. Although the power supply reliability of power cables is higher than that of overhead lines, they are generally buried directly underground and the line structure is more complex, so it is not easy to locate the fault location and troubleshoot the fault in time when a fault occurs [1-3]. Therefore, fast and accurate localization of distribution network cable faults is very important for fault removal, reducing fault outage time and improving power supply reliability [4-5].

In recent years with the increasing service life of cables, the incidence of cable faults has begun to rise gradually, and the problem of detecting and diagnosing cable faults has gradually attracted people's attention [6-8]. Cable fault refers to a series of faults occurring in cables, including but not limited to insulation aging, insulation breakdown, joint failure, conductor breakage, etc. The occurrence of these faults will seriously affect the stable operation of the power system, and may even cause safety accidents. Therefore, the detection and diagnosis techniques of cable faults are of great significance in ensuring the stable operation of power systems and extending the service life of cables [9].

Different techniques have different advantages and disadvantages, and in the actual signal cable fault screening and localization, the applicability of different techniques, the reliability is different. Literature [10] introduces a method based on transient current measurement of cable shield grounding, which can identify the fault identification and localization of underground cables and hybrid lines, based on which a new transient grounding fault detection and localization algorithm is proposed, the new algorithm has the ability of autonomous learning, compared with the traditional method can be identified under the conditions of transient grounding faults on the distribution feeder branch, and the feasibility of the algorithm is examined through the test. Literature [11] aims to realize the precise location of cable faults, and proposes two hybrid schemes, MT-CT-DFT scheme and ANN-WT, and detects and compares the two schemes through multi-fault type tests, and finds that the ANN-WT method is more accurate and shows low sensitivity to parameter changes. Literature [12] describes the evaluation of a fault localization technique based on the FasTR method in an airborne portable system, and the results show that the system can achieve on-line detection of transient faults in complex networks and ensure the accuracy of the detection. Literature [13] describes Megger's new EZ-THUMP 3 cable fault location system, which works by combining TDR measurements with a surge generator to identify high resistances and thus locate faults. It can help users to locate faults easily, quickly and accurately. Literature [14] designed a hybrid fault location method for overhead lines containing underground cables, which was proved to be superior to other traditional methods by simulation techniques, showing high accuracy in both fault location and identification functions. Literature [15] discusses Megger's new SMART THUMP ST25-30 portable cable tester, which is based on automated test sequencing technology and provides fault identification and location, as well as interpretation and analysis of the test results for non-specialized users to obtain

https://doi.org/10.17993/3ctecno.2024.v13n1e45.130-155

reliable fault location results in a safe and fast manner. Literature [16] on the iterative method and non-iterative method of localization function of the simulation test method for comparison, the results show that the positioning accuracy of the PD by the algorithm error, iteration algorithm, the number of iterations and iteration of the initial value of the impact of the algorithm. Literature [17] describes a strategy to reduce the occurrence of CCFS currents in cables by using current limiting reactors to shield the cables and then connect them to the electrodes of the primary substation, which can effectively reduce the thermal burning of insulation due to CCF in cables. Literature [18] studied cable joint breakdown faults, and through the analysis of fault waveforms, simulation model calculations and discussion of finite element methods, it was pointed out that moisture in the internal joints of cables causes transient ground faults, damages the grounding wire, and reduces the sensitivity of the protective device. Literature [19] proposes a Bayesian function based signal type identification method to identify the signal. A scheme to discriminate pole-to-pole and pole-to-ground short circuits in VSC cables is also proposed, and the reliability and innovation of the described method is confirmed by the study of real cases. Literature [20] points out that in cable-related construction, cross-connection errors often occur resulting in no effective limitation of the induced voltage in the metal shield, resulting in cable accidents, in order to solve this status quo, the cross-grounding faults in the metal shield of the 5kv single-core power cables are investigated, and at the same time remedial suggestions and measures are given for the cross-grounding faults.

In this paper, we analyze the characteristics of cable faults and select a suitable signal processing method. Then, VMD is used for the initial processing of the cable fault signal, followed by the application of wavelet transform for further analysis and noise reduction. On this basis, the key parameters of VMD and wavelet transform are adjusted using parameter optimization methods to adapt to different fault signal characteristics. The effectiveness of the proposed method is evaluated through comparative analysis and compared with traditional methods to verify its superiority.

2. POWER CABLE FAULT RELATED ANALYSIS

Power cables with its small footprint, high reliability, easy maintenance and other unique advantages in the distribution network is increasingly widely used, and along with the social and economic take-off, national consumption level, the reliability and security of the distribution network operation of high-level quality requirements. Under the threat of extreme weather, line aging, natural disasters, external damage, man-made theft of cables and other factors, it is very easy to cause cable operation faults, resulting in the entire power line blackout accidents, power transmission interruptions directly lead to the production of life safety hazards and economic property losses. This chapter mainly discusses the relevant basic knowledge of power cable faults, in order to realize the precise positioning of power cable faults to provide support.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.130-155

2.1. POWER CABLE FAILURE RELATED CAUSES

2.1.1. CAUSES OF POWER CABLE FAILURES

City cables are generally buried in the ground and the pipeline by the cable length of the impact of the existence of two cables with joints, cable storage is not good there are old and new cables and cables in the laying of grafting whether in strict accordance with the requirements of these are for the occurrence of faults buried hidden dangers. In addition, the actual operation of many other reasons for failure, power cable line failure causes are mainly the following:

1. External damage. Cables are mostly laid in the ground, the rise of central China in recent years, infrastructure construction increased, urbanization accelerated, rural decentralized building into a centralized, mechanical damage caused by a lot of cable failures.

2. Cable joint failure. Cables and cables are connected by the joints, in the cable joints are prone to failure, the joints have become the entire cable part of the most "fragile" place.

3. Chemical corrosion. For the existence of acid and alkali soil laying cable will inevitably be eroded by acid and alkali, often resulting in damage to the external protective layer.

4. External environment. Whether the heat source around the cable exceeds the standard needs to be taken into account, overheating will cause damage to the insulation of the cable, reduce the insulation strength of the cable caused by insulation breakdown resulting in power outages.

5. Cable long-term overload operation. Due to the cable's own design defects, three-phase load imbalance, etc., so that the cable in the unprotected environment for a long time overloading operation will lead to the cable with the operation of the time to lengthen the temperature is too high, too high a temperature will make the "fragile" joints part of the first damage.

6. The lack of technology. Cable body in the manufacturing process is not in strict accordance with the design requirements of the environment and technology, materials, such as failure to lead to normal aging or in the normal energized lead to breakdown and so on.

2.1.2. CHARACTERISTICS OF POWER CABLE FAULTS

When power cables have various types of faults, the consequences of which can lead to regional local power outages, or paralyze the entire power supply system, thus seriously affecting the productivity of enterprises. In order to quickly determine the type of cable faults and maintenance strategies, first of all need to classify the power

https://doi.org/10.17993/3ctecno.2024.v13n1e45.130-155

cable faults, and then according to the classification results to choose the appropriate protection measures, in order to quickly restore power supply, minimize the loss of enterprises. The classification method of cable faults can be classified according to different criteria, and its specific classification is shown in Figure 1, which mainly includes three categories of series faults, parallel faults and compound faults.

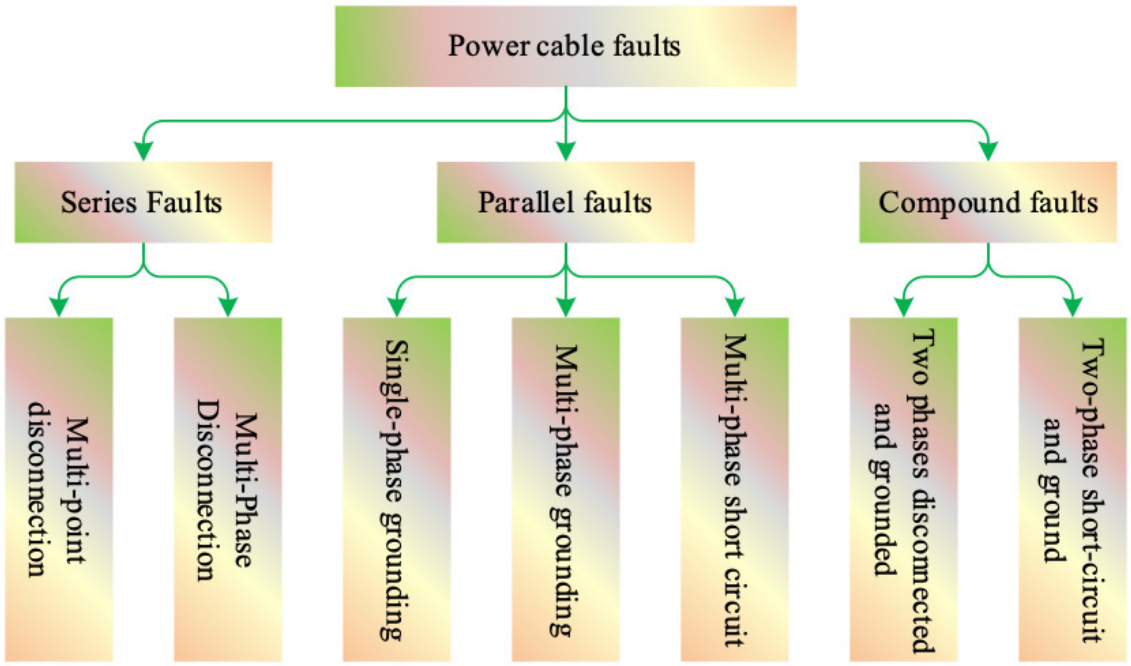


Figure 1 Power cable fault classification chart

Series fault refers to the internal conductor of the cable at a place where a break or short circuit occurs, resulting in a cable blackout accident, parallel fault is due to the loss of insulation medium between the core wires inside the cable and lead to parallelism between the core wires, resulting in faults, composite faults refers to the simultaneous existence of the above two types of faults in the cable.

2.2. POWER CABLE FAULT DETECTION PROCESS

2.2.1. FAULT DETECTION OF POWER CABLES

Power cable fault diagnosis process shown in Figure 2, the specific principle is that the cable core conductor resistance and core distance into a proportional relationship, as long as the calculation of the beginning of the fault phase of the cable to the fault point of the core conductor resistance and the fault phase of the proportionality coefficient of the conductor resistance, then in the case of the full length of the cable known to be able to calculate the distance from the beginning of the point of failure, to complete the diagnosis of the cable fault and localization.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.130-155

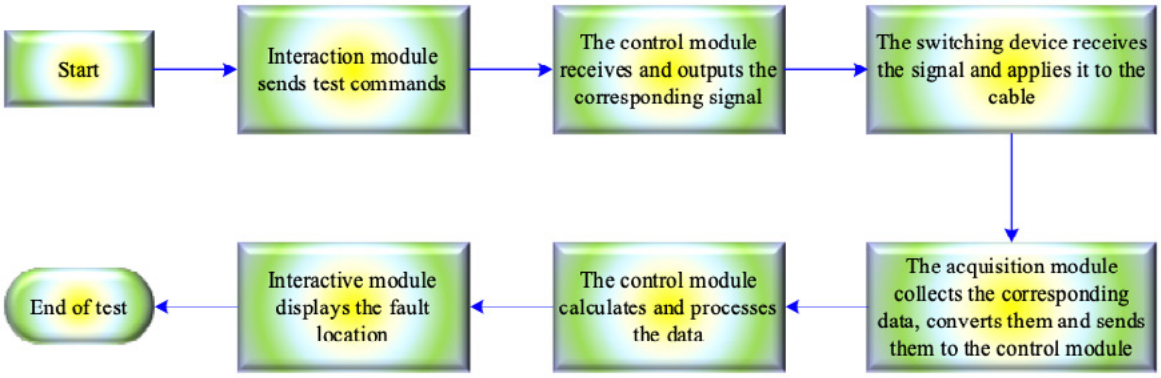


Figure 2 Fault diagnosis process of power cable

The above method is currently a relatively simple and fast way to diagnose and locate power cable faults, can effectively, accurately, quickly, conveniently and safely determine the cable fault category and locate the fault point, is to protect the normal operation of the power system is the key. Reasonable analysis of the causes of power cable faults to help solve the source of cable faults, and for the development of scientific cable fault detection technology to provide a basis.

2.2.2. EQUIVALENT MODELING OF POWER CABLES

Power cable is a power transmission line, when the power cable is considered a long line, it is no longer a simple conductor - insulation - to ground circuit, but many more equivalent resistance, conductance, inductance, capacitance composed of these parameters are uniformly distributed along the entire cable line, so called distribution parameters. Figure 3 for the cable equivalent long line distribution parameter circuit. Figure $R_0(\Omega/m)$ for the cable line unit length of the resistance, $G_0(S/m)$ for the cable line unit length of the conductance, $L_0(H/m)$ for the cable line unit length of the inductance, $C_0(F/m)$ for the cable line unit length of the capacitance.

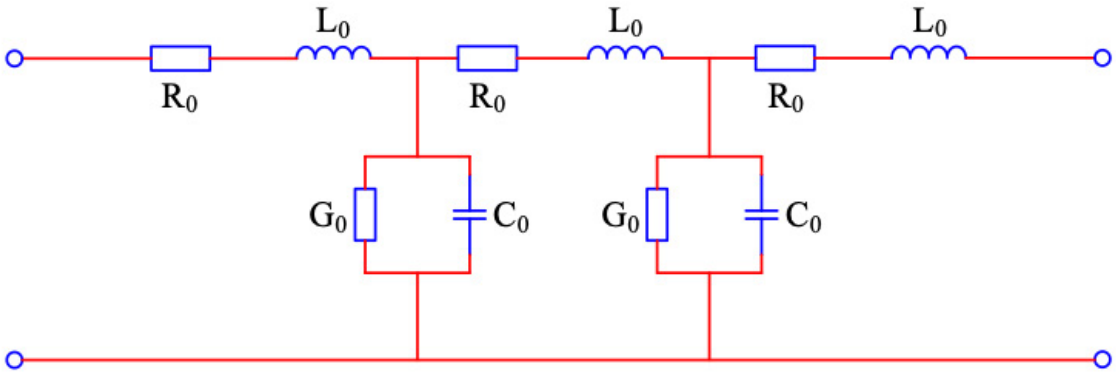


Figure 3 Equivalent long line distribution parameter circuit of cable

https://doi.org/10.17993/3ctecno.2024.v13n1e45.130-155

Signal current after resistance and inductance, through the unit length of the cable line, are generated voltage drop and through the conductance, capacitance, etc., split the flow, and in the middle of the flow back. Cable transmission of high-frequency waves, it will lose a considerable amount of energy into the attenuation, at this time, conductivity, resistance and other losses can be ignored, and such a circuit is known as lossless circuit.

2.3. POWER CABLE FAULT TRAVELING WAVE PROCESS

2.3.1. TRAVELING WAVE VELOCITY OF POWER CABLE

Power cable traveling wave speed is expressed in the traveling wave propagation process fast and slow physical quantities, when the traveling wave propagation in the cable line, from a point in the cable propagation to another point needs to go through a certain amount of time, the traveling wave propagation distance and propagation time ratio is known as the wave speed.

For the purpose of calculating the wave velocity *v,* assume that the current wave acting on the wire is an oblique angle current wave. Suppose that the initial condition is zero, and that an oblique wave current of value $i = \alpha t$ (α in units of A/s, t in units of *s*) is applied to the wire at point *A* at $t = 0$. Suppose that the wave moves along the wire with some known wave speed *v* and after an elapsed time *t*, the wave reaches point *B* and the potential at point *B* is zero. Then the voltage drop across the inductor *L*o*x* point *A* to point *B* is the potential $\boldsymbol{u}_A$ at point *A*. And since $x = vt$ and therefore $L_0 x = L_0 vt$, the following relation for u_A can be written:

$$u_A = L_0 vt \frac{di}{dt} = L_0 vt\alpha \tag{1}$$

Assuming that the charge per unit length of wire at point *A* is *q*, according to the definition of capacitance, it can be seen that the charge *qdx* stored on the capacitance $C_0 dx$ of the d_x section at point *A* and the potential u_A at point A can be expressed as $C_0 dx = u_A q dx$, i.e., the potential u_A at point *A* can be expressed in terms of capacitance:

$$u_A = \frac{qdx}{C_0 dx} = \frac{q}{C_0} \tag{2}$$

According to the definition of current, the current *i* at point *A* is the number of charges passing per unit time is:

$$i = \frac{qdx}{dt} = q\frac{dx}{dt} = qv \tag{3}$$

Thus, the charge per unit length *q* can be expressed as:

https://doi.org/10.17993/3ctecno.2024.v13n1e45.130-155

$$q = \frac{i}{v} \tag{4}$$

Bringing Eq. (4) into Eq. (2) and replacing *i* with *at* yields the relation for u_A as:

$$u_A = \frac{i}{vC_0} = \frac{\alpha t}{vC_0} \tag{5}$$

Up to this point, two expressions (1) and (5) for u_A are obtained and this two equations should be equal, i.e.:

$$L_0 vt\alpha = \frac{\alpha t}{vC_0} \tag{6}$$

Organizing the above equation, the expression for the wave velocity *v* is obtained as:

$$v = \pm \frac{1}{\sqrt{L_0 C_0}} \tag{7}$$

Substituting $L_0 = \frac{\mu_0 \mu_r}{2\pi} \ln \frac{2h}{r}$ and $C_0 = \frac{2\pi \varepsilon_0 \varepsilon_r}{\ln \frac{2h}{r}}$ into equation (7) we get:

$$v = \pm \sqrt{\frac{1}{\mu_0 \mu_r \varepsilon_0 \varepsilon_r}} \tag{8}$$

Since ε_0 and μ_0 are both constants, it can be seen according to Eq. (8) that the wave velocity *v* of the cable line is only related to the relative permittivity $\mathscr{E}$ of the medium around the cable core and the relative permeability coefficient μ_r of the medium around the cable core such as insulation and shielding, while factors such as the material of the conductor core and the cross-sectional area do not affect the wave velocity of the cable line.

2.3.2. CALCULATE REFLECTED AND REFRACTED WAVES

Reflected and refracted waves are important concepts in the calculation of traveling waves. When a voltage is applied to a power cable, a current is generated on the cable, then if a sudden change occurs in the power a A, then refracted and reflected waves will be generated at A now. The refraction and reflection of traveling waves is shown in Figure 4, according to the relationship between circuit current and voltage, then there is:

$$\left.\begin{aligned} u_{r2} &= u_{r1} + u_{f1} \\ i_{r2} &= i_{r1} + i_{f1} \end{aligned}\right\} \tag{9}$$

https://doi.org/10.17993/3ctecno.2024.v13n1e45.130-155

$$\left.\begin{aligned} u_{r1} &= i_{r1}Z_1 \\ u_{r2} &= i_{r2}Z_2 \\ u_{f1} &= -\,i_{f1}Z_1 \end{aligned}\right\} \tag{10}$$

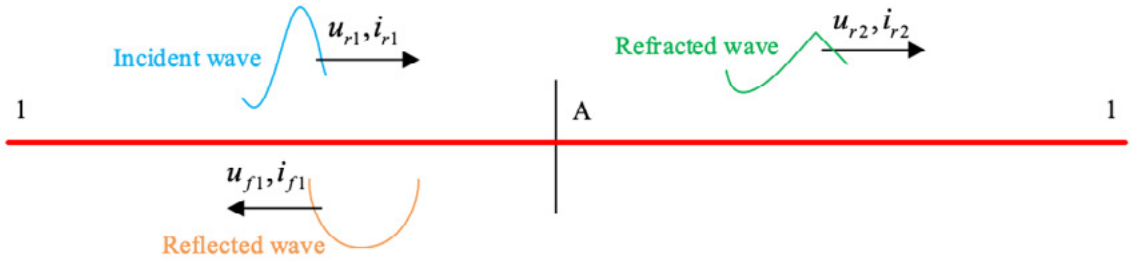


Figure 4 Refraction and reflection of the traveling waves

from Eq. (9) and Eq. (10):

$$\left.\begin{aligned} u_{r2} &= \frac{2Z_2}{Z_2+Z_1}u_{r1} = \gamma_u u_{r1} \\ u_{f1} &= \frac{Z_2-Z_1}{Z_1+Z_2}u_{r1} = \rho_u u_{r1} \end{aligned}\right\} \tag{11}$$

Where γ_u is the voltage refraction coefficient, $\gamma_u = \frac{2Z_2}{Z_2+Z_1}$, ρ_u is the voltage reflection coefficient, $\rho_u = \frac{2Z_2}{Z_2+Z_1}$. .

From Eq. (11) and Eq. (10), it can be solved as follows.

$$i_{f1} = -\,\rho_u i_{r1} \tag{12}$$

$$\rho_i = -\,\rho_u \tag{13}$$

Where,ρ_i is the current emission coefficient.

3. POWER CABLE FAULT LOCATION DIAGNOSIS

The development of urbanization needs to build more power cables as a support, urban power transmission needs to use advanced scientific power transmission methods, to provide great convenience and safety for people's electricity. nowadays, the city mainly uses underground cables for power transmission, this way not only can ensure the stable transmission of electricity, but also can reduce the probability of power cable line faults. however, underground cable power transmission can also have faults and need to be repaired. this chapter mainly explores the optimization based on the skyhawk algorithm variable modal decomposition combined with wavelet transform in power cable fault location related technology, in order to realize the power cable faults. This chapter mainly explores the optimized variational modal

https://doi.org/10.17993/3ctecno.2024.v13n1e45.130-155

decomposition based on the eagle algorithm combined with the wavelet transform in the power cable fault localization technology, to provide technical support for the realization of the accurate positioning of power cable faults.

3.1. PARAMETER OPTIMIZED VARIATIONAL MODAL DECOMPOSITION

3.1.1. VARIATIONAL MODAL DECOMPOSITION (VMD)

Variational modal decomposition (VMD) algorithm is a non-recursive decomposition of a set of signals into *K* quasi-orthogonal and specific sparsity intrinsic modal functions (IMFs) to achieve effective separation of signals. The VMD algorithm is based on the concepts of Wiener filtering, the Hilbert transform, and frequency mixing, etc., and the overall idea is to construct a variational problem. The constraints on the variational components need to be satisfied that all the sums of the components are consistent with the original signal, and the constraints on the variational model are as follows.

$$\begin{cases} \min_{\{u_k\}w_k\}} \left\{ \sum_{k=1}^{K} \| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-jw_k t} \|_2^2 \right\} \\ \text{s.t. } \sum_k u_k = f(t) \end{cases} \tag{14}$$

Where, $\{u_k\}$ is the *K* eigenmode components obtained after VMD decomposition, $\{u_1, u_2, \cdots, u_k\}$, $\{w_k\}$, are the center frequencies of each of the *K* eigenmode components, $\{w_1, w_2, \cdots, w_k\}$, $f(t)$ are the original signals, ∂ is the sign of gradient computation, $\delta(t)$ is the Dirac Lay function,* is the sign of convolution operation, and s.t. is the constraint term.

In order to solve the variational constrained problem and complete the transformation from constrained to unconstrained problem, the Lagrange operator and the quadratic penalty factor α are introduced, and the transformations result in the augmented and generalized Lagrange formulas, namely:

$$\begin{aligned} L\left(\{u_k\}, \{w_k\}, \lambda \right) = \alpha \sum_{k=1}^{K} \| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-jw_k t} \|_2^2 \\ + \| f(t) - \sum_{k=1}^{K} u_k(t) \|_2^2 + \left\langle \lambda(t), f(t) - \sum_{k=1}^{K} u_k(t) \right\rangle \end{aligned} \tag{15}$$

In the formula, the value of α affects the denoising effect of the algorithm, and the appropriate value of α can reduce the noise interference.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.130-155

Finally, the alternating direction multiplier algorithm is used to iteratively update $\{u_k\}, \{w_k\}$, and λ until the termination condition is satisfied, and the final K IMF components and the corresponding center frequencies are output.

From the above formula, it can be seen that the values of parameters *K* and α will have an important impact on the decomposition results of the algorithm. If the *K* value is too small, it will lead to insufficient decomposition, if it is too large, it will be prone to problems such as false components and frequency overlap, if the α value is too small, the signal denoising will not be thorough enough, and if it is too large, the active components will be removed by mistake. Empirical selection of the above parameter values does not ensure that they are optimal.

To address this problem, this paper introduces the Skyhawk optimization algorithm to improve the VMD and get the optimal parameter combination $[K, \alpha]$. In this process, the envelope entropy reflects the sparsity of the signal, the more noise there is in the signal, the less effective components there are, which is manifested by the larger envelope entropy value; on the contrary, the more the signal contains effective components, the smaller the envelope entropy value is, i.e., when the value of envelope entropy is the smallest, the signal contains the largest number of effective components, and at this time, the corresponding parameter is the optimal. Therefore, the author adopts the minimum value of the envelope directrix as the fitness function of the Skyhawk optimizer to evaluate the decomposition effect of the parameter combination. The mathematical formula for the envelope entropy E_p is as follows.

$$E_p = -\sum_{q=1}^{m} p_q \lg p_q \tag{16}$$

$$p_q = a(q) / \sum_{q=1}^{m} a(q) \tag{17}$$

$$a(q) = \sqrt{[x(q)]^2 + \{H[x(q)]\}^2} \tag{18}$$

Where, m is the number of sampling points, P_q is the normalized form of *a(q)*, *a(q)* is the envelope signal after Hilbert transform.

3.1.2. AOA LGORITHMIC OPTIMIZATION OF VMD PARAMETERS

The VMD is continuously updated in the frequency domain and transformed to the time domain by Fourier inverse transform, and the final results will be different when different decomposition layers K and quadratic penalties are inputted, so finding the optimal combinations of decomposition layers K and quadratic penalties is the key to the VMD. In this paper, we propose to optimize the parameters of the VMD based on the AO algorithm, and the optimized parameters can be obtained quickly and accurately. The AO algorithm is a new population-based optimization method, which

https://doi.org/10.17993/3ctecno.2024.v13n1e45.130-155

mainly simulates the natural behavior of the eagle in the process of capturing prey, so as to achieve the purpose of optimization, with strong optimization ability and fast convergence speed, etc. Therefore, the AO algorithm is used to optimize the VMD.

Therefore, the AO algorithm is used for iterative optimization of the number of decompositions and the penalty factor in the VMD, and the optimization dimension is set to 5, the optimization interval is set to [5,12], and the optimization interval is set to [0,20,000].The loss function of the VMD is used as the fitness function for the optimization of the AO algorithm, and the calculation formula is as follows: the VMD loss function is used as the fitness function for the optimization of the AO algorithm, and the calculation formula is as follows: the VMD loss function is used as the adaptation function for the optimization of the AO algorithm.

$$L_{\text{loss}} = \frac{\sum_{t=1}^{T} \left| f(t) - f'(t) \right|}{T} \tag{19}$$

Where $f(t)$ is the original input signal, $f'(t)$ is the decomposed signal and *T* is the time length.

The original signal is decomposed into *K* modal components by VMD, if the modal components contain less noise components, the feature information related to the original signal will be more obvious, and the envelope entropy will be smaller, and the AO algorithm is used to seek the minimum envelope entropy, so that the IMF obtained by this way can maximize the retention of the characteristics of the fault signal of the power cable.

3.2. AO-VMD-CWT FAULT LOCALIZATION MODEL

3.2.1. CONTINUOUS WAVELET TRANSFORM (CWT)

The translation factor and scale factor in the wavelet time-frequency transform define the position and shape of the time-frequency window, which makes the wavelet transform adaptive and multi-resolution, and is widely used in the field of signal processing. The continuous wavelet transform (CWT) adopts a time-frequency window that can be adaptively adjusted with the frequency, which overcomes the limitation that the size of the window of the short-time Fourier transform (STFT) can't be adjusted with the frequency or the time, which is difficult to accurately respond to the relationship between the frequency and the time, and is more suitable for dealing with the transient and sudden change of the signals.

For the function $\varphi(t) \in L^2(R)$, if it satisfies $\int_{-\infty}^{\infty} \varphi(t)dt = 0$, then $\varphi(t)$ can be written as the mother wavelet. By performing a series of scale translation transformations on the mother wavelet $\varphi(t)$ a series of consecutive wavelet functions can be obtained, which are called analytic wavelets. The transformation formula is:

https://doi.org/10.17993/3ctecno.2024.v13n1e45.130-155

$$\varphi_{u,v}(t) = \frac{1}{\sqrt{v}}\varphi\left(\frac{t-u}{v}\right), v > 0, u \in R \tag{20}$$

In the formula, *u* is the translation factor, *v* is the scale factor, also known as the expansion factor, when $v > 1$, stretching along the horizontal direction, when $v < 1$, compression along the horizontal direction, in order to keep the energy constant after the expansion transformation need to multiply the scale factor $1/\sqrt{v}$ in front of the front, i.e., $\| \varphi_{u,v}(t) \| = \| \varphi(t) \|$, *u* is the translation parameter, which can be taken as an arbitrary real number, *u* and *v* are continuous variables, so it is known as the continuous wavelet. Wavelet transform.

The continuous wavelet transform of the signal $x(t) \in L^2(R)$ is expressed by the following equation:

$$\mathrm{CWTx}(u,v) = \int_{-\infty}^{\infty} x(t)\varphi_{u,v}^{*}(t)dt = \frac{1}{\sqrt{v}}\int_{-\infty}^{\infty} x(t)\varphi^{*}\left(\frac{t-u}{v}\right)dt \tag{21}$$

Where $\varphi^{*}(t)$ is the complex function of the function $\varphi(t)$. The inverse transformation is given by:

$$x(t) = \frac{1}{C_\varphi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varphi_{u,v}(t)CWT(u,v)\frac{dudv}{v^2} \tag{22}$$

Formula, $C_\varphi = 2\pi\int_{-\infty}^{\infty}\frac{|\hat{\varphi}(\omega)|}{|\omega|}d\omega$..

The mother wavelet function needs to satisfy the following conditions to ensure that the wavelet transform can accurately construct the original signal with corresponding inverse transform.

1. First of all, the value of the mother wavelet function outside the window function is zero.
2. It needs to be satisfied: $\int_{-\infty}^{\infty}\frac{|\hat{\varphi}(\omega)|}{|\omega|}d\omega < +\infty$ 。
3. The mother wavelet function satisfies $\hat{\phi}(\omega)\Big|_{\omega=0} = 0$ 。

3.2.2. AO-VMD-CWT SIGNAL NOISE REDUCTION

The search strategy of AO algorithm adopts repeated trajectories to explore the approximate optimal solution or the reasonable location of the optimal solution, which has high convergence, robustness and strong optimization ability. Therefore, this paper introduces the AO algorithm to search for the optimization of decomposition

https://doi.org/10.17993/3ctecno.2024.v13n1e45.130-155

number and penalty factor in the VMD, and uses the minimum envelope entropy as the fitness function to transform the iterative optimization process into the process of AO algorithm to seek for the minimum envelope entropy. The specific optimization process is as follows The specific optimization process is as follows

Step1 Initialize the population, set the number of iterations of AO algorithm, the population size, the number of variables and the upper and lower limits of the problem to be solved.

Step2 Decompose the input signal by VMD.

Step3 Calculate the minimum value of the envelope entropy of each modal component as the fitness function and substitute it into the optimization algorithm.

Step4 Update the position of the population and the global optimal solution, and stop the iteration when the optimization algorithm meets the iteration termination condition.

On this theoretical basis, this paper proposes a parameter optimization VMD-CWT based power cable fault signal noise reduction method, the method is shown in Fig. 5. the specific process is as follows.

1. The AO algorithm optimizes the VMD, and the signal decomposition is performed using the VMD to obtain *K* IMF components.
2. According to the cliff-correlation coefficient, the IMF components are divided into pure components and noise components.
3. Perform wavelet threshold noise reduction on the noise component to remove the noise component in the signal.
4. The pure component and the wavelet threshold noise reduction processed noise-containing component were reconstructed to obtain the joint noise reduction signal.

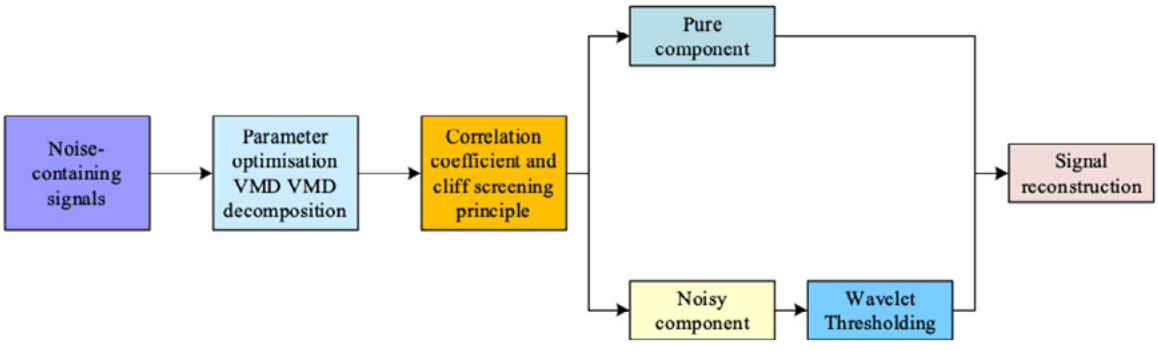


Figure 5 Signal denoising method for parameter optimization of AO-VMD-CWT

https://doi.org/10.17993/3ctecno.2024.v13n1e45.130-155

In this paper, we choose to use continuous wavelet transform to decompose the data after parameter-optimized VMD decomposition at multiple scales as a method to construct time-frequency maps, and to generate two-dimensional wavelet time-frequency domain maps so that we can take into account the ability of local information in the time-frequency domain at the same time.

4. POWER CABLE FAULT LOCATION SIMULATION

Power cable lines have unique advantages over overhead lines, including small footprint, high power supply reliability, low voltage drop, low fault rate and lightning protection, etc. In recent years, with the construction and development of the city and the increase in the use of power cables, their faults have received more and more attention, so how to quickly locate the faults of the cable is critical to reduce the outage time. This chapter mainly focuses on the effectiveness of the parameter optimization VMD-CWT fault location method given in the previous section to carry out the simulation analysis of the data, so as to help the accurate positioning of power cable faults, in order to enhance the safety of electricity, reduce the economic losses caused by power outages to provide protection.

In the cable fault location algorithm research, need to verify the feasibility of the algorithm through the test, the field test is not only very high cost and not safe enough, the general use of computer simulation methods to replace the real operating environment of the power system. this paper will use ATPDraw for cable fault model drawing, and then use EMTP to transient analysis of cable faults, the results obtained from analysis. The results of the analysis will be imported into MATLAB for data processing and display. using ATPDraw for modeling simulation, simulation simplification principle shown in Figure 6.

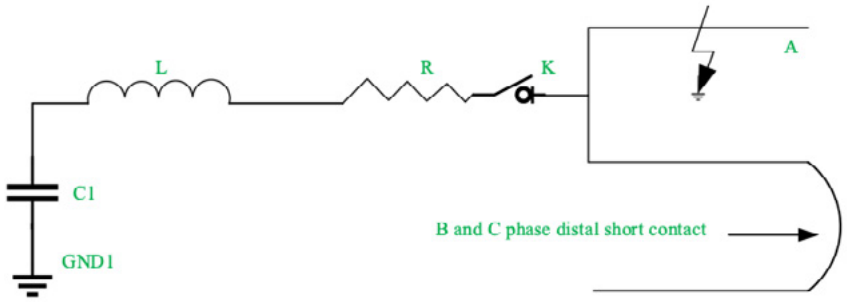


Figure 6 Simulation schematics

https://doi.org/10.17993/3ctecno.2024.v13n1e45.130-155

4.1. SIGNAL NOISE REDUCTION AND FAULT LOCALIZATION SIMULATION

4.1.1. AO-VMD-CWT SIGNAL NOISE REDUCTION

In order to analyze the effectiveness of the AO-VMD-CWT method proposed in this paper in the power cable fault signal noise reduction, this paper designed the corresponding power cable fault simulation model based on ATPDraw, and designed the corresponding noise-containing power cable fault signal. After the noise-containing signal is decomposed and denoised by using the AO-VMD method, the signal is decomposed by six modal components, and the denoised signal is obtained by analyzing the modal components obtained through the correlation coefficient as shown in Fig. 7. The correlation coefficient analyzes the modal components of the decomposition, and the denoised signal is shown in Fig. 7.

As can be seen from the figure, after the VMD decomposition denoising, the effect of signal noise reduction is achieved to a certain extent, using the formula given in the previous section, the signal-to-noise ratio of the original noise signal can be calculated as 20.47 dB, while the signal-to-noise ratio of the noise-reduced signal after the VMD denoising is 53.38 dB, which is 1.61 times more than that of the original noise signal, which indicates that the AO This shows that the AO-VMD-CWT method given in this paper can effectively realize the noise reduction of power cable fault signals, obtain a more accurate power cable fault localization effect, and reduce the localization error caused by the fault signal noise.

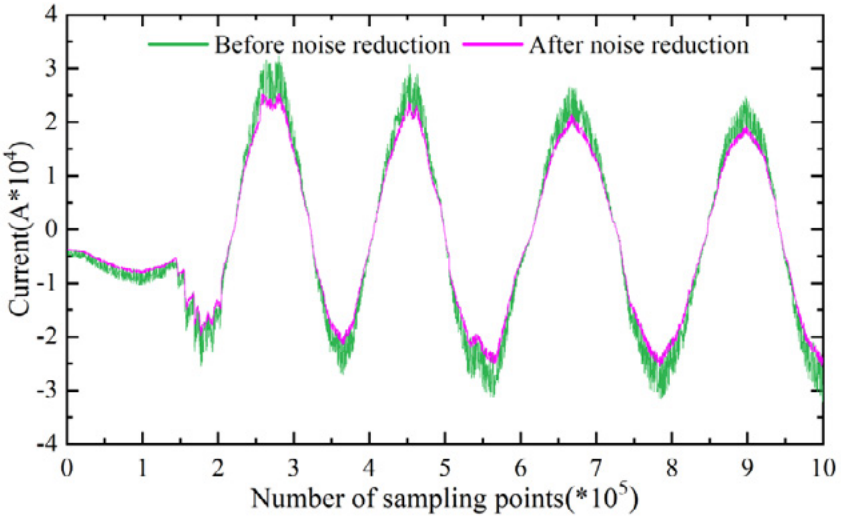


Figure 7 Noise reduction waveform of the AO-VMD-CWT signal

In order to further illustrate the noise reduction effect of this paper's method, the noise reduction waveforms of this paper's method, the CWT method and the VMD method are compared, and the signal-to-noise ratios of the noise reduction are also compared. Figure 8 shows the noise reduction waveforms of the power cable fault signals obtained by different noise reduction methods.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.130-155

From the noise reduction waveforms of power cable fault signals in the figure, it can be clearly seen that the method of this paper is better than a single wavelet transform method and the noise reduction waveform obtained by the variational modal decomposition method, which can better reflect the fault situation of power cables. Using the signal-to-noise ratio solution method given in the previous section to solve the signal-to-noise ratios of the three kinds of noise reduction waveforms, it can be clearly defined that the signal-to-noise ratios of the wavelet transform and variational modal decomposition method are respectively 24.65 dB and 28.19 dB, and the signal-to-noise ratio of the wavelet transform method is 24.65 dB and 28.19 dB, respectively. The signal-to-noise ratio of the wavelet transform and the variational modal decomposition method is 24.65dB and 28.19dB respectively, which is 53.82% and 47.19% lower than that of this paper's method, and the parameter optimization VMD algorithm combined with the continuous wavelet transform can obtain better noise reduction waveforms of power cable fault signals, with a larger improvement in the signal-to-noise ratio, which can provide accurate signal waveforms for power cable fault localization and can effectively avoid the loss of some useful signals when the VMD algorithm is used for noise reduction. It can also effectively avoid the loss of some useful signals when the VMD algorithm is used for noise reduction.

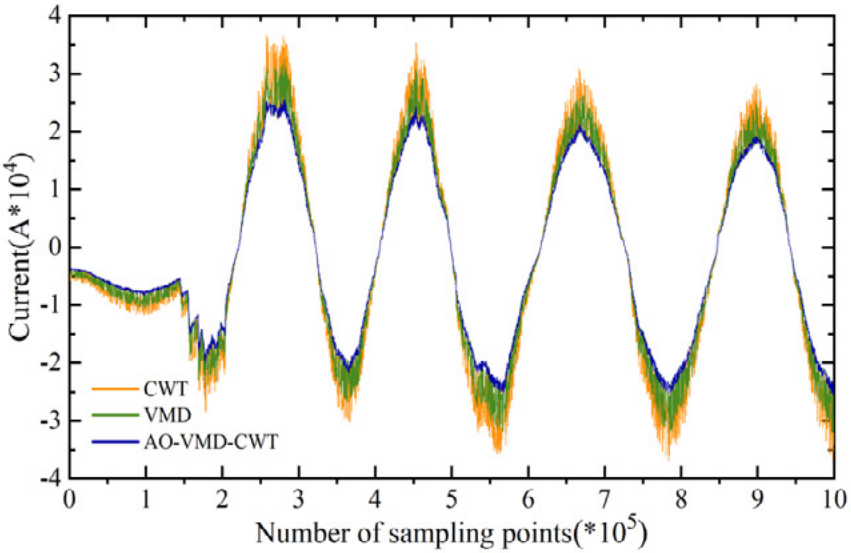


Figure 8 Comparison of Three Noise Reduction Waveforms

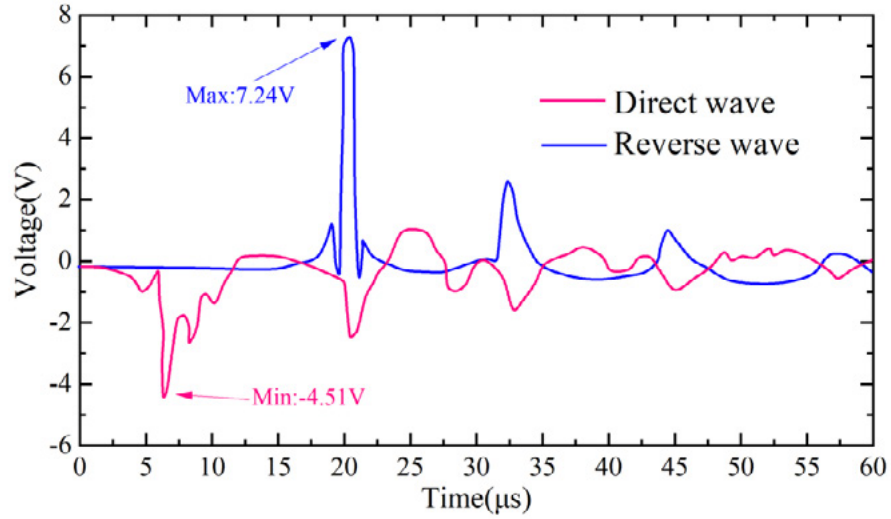
4.1.2. SIMULATION OF POWER CABLE FAULT DISTANCE

In order to assess the feasibility of the proposed fault location method, combined with the previous power cable fault simulation model, assuming that the cable length is 5km, the fault is set on phase A, and the voltage-controlled switch is used to simulate the fault click-through, and the fault distances are set to be 50, 100, 500, 1000, 2000, 3000, 4000 and 5000 m. The normal phases, phase B and phase C are shorted at the far end, and the travelling wave is traveling. The high voltage signal generator model is used to send out the coupled current output signal to analyze the correlation between the fault signal and the non-fault signal output signal before the

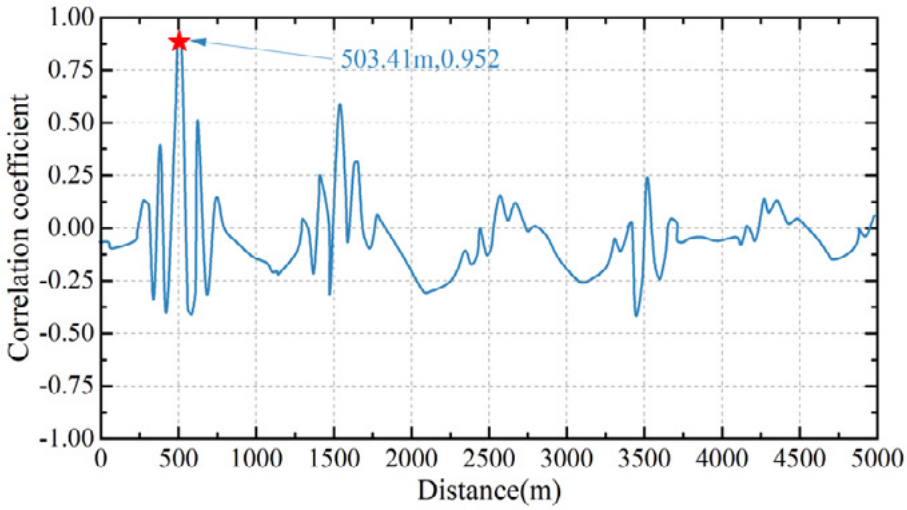
https://doi.org/10.17993/3ctecno.2024.v13n1e45.130-155

fault clicks through, and the fault phase and the non-fault phase current are used for the construction of the forward and reverse voltage traveling wave, and the resulting direction of the traveling wave wave is shown in Fig. 9, of which Fig. 9(a) and (b) are the forward and reverse voltage traveling wave and the correlation coefficient curve, respectively.

Combined with the forward and reverse voltage traveling wave and correlation coefficient curves, when the coupled current output signal reverse traveling wave transmission in the 21.48μs period to reach the voltage maximum value of 7.24V, and at this time the fault ranging is shown as 503.41 meters, the correlation coefficient of the maximum of 0.952. Combined with this paper to set up the fault point, the fault point of the difference between the 500 meters and the fault point of 3.41 meters, the relative error of 0.682%. 0.682%.



(a) Forward and backward direction voltage traveling wave(L=500)



(b) Correlation coefficient curve(L=500)

Figure 9 Direction traveling waveform

https://doi.org/10.17993/3ctecno.2024.v13n1e45.130-155

According to the above method, the simulation results of all fault distances are solved, and the relative errors of different fault distances are shown in Table 1. From the simulation results, except for the fault location distance of 50 meters and 100 meters which is more than 1.00%, the relative errors of other fault location of power cables within 5 kilometers are maintained in the range of 0.45%~0.70%, and the average value of the relative errors is about 0.849%. The above results verify the feasibility of this paper's method in power cable fault localization, and provide technical support for the realization of timely repair of power cable faults.

Table 1 Simulation ranging results

Fault distance	Simulation result	Absolute error	Fractional error
50m	50.69m	0.69m	1.38%
100m	101.75m	1.75m	1.75%
500m	503.41m	3.41m	682 %
1000m	1006.96m	6.96m	696 %
2000m	2009.38m	9.38m	469 %
3000m	3015.82m	15.82m	527 %
4000m	4026.54m	26.54m	664 %
5000m	4031.27m	31.27m	625 %

4.2. POWER CABLE FAULT LOCATION PERFORMANCE COMPARISON

4.2.1. COMPARISON OF DOUBLE-ENDED TRAVELING WAVE RANGING METHODS

In order to verify the performance of this paper's algorithm for power cable fault location, we choose to compare it with the classical double-ended traveling wave ranging (DRW) method, using the power cable fault simulation model given in the previous section to send out 500 coupled current signals, and comparing the average positioning results of the two methods on 500 samples of data with the average absolute positioning error as an evaluation index. The average localization results of different algorithms are shown in Table 2, which are obtained by choosing four types of faults, namely, single-phase grounded short circuit, two-phase short circuit, two-phase grounded short circuit, and three-phase short circuit, and taking 10km, 50km, 80km, and 100km as the fault distance.

From the average positioning results of different algorithms, the traditional double-ended traveling wave ranging method has low accuracy, mainly due to the

https://doi.org/10.17993/3ctecno.2024.v13n1e45.130-155

propagation speed of traveling waves in power cables affected by the fault distance, fault type, etc., with uncertainty, the average positioning error is more than 0.51km, and the average absolute positioning error is more than 0.5%. The localization accuracy of the AO-VMD-CWT algorithm proposed in this paper is significantly higher than that of the double-ended traveling wave method, and the average absolute localization errors for fault distances of 10km, 50km, 80km, and 100km are less than 1.315%, 0.278%, 0.163%, and 0.151%, respectively, for different fault types.

Among the single-phase grounded short-circuit fault types, this paper's algorithm has the highest localization accuracy, and the average absolute localization errors for fault distances of 10km, 50km, 80km, and 100km are 1.151%, 0.273%, 0.152%, and 0.108%, respectively. This is due to the fact that the proposed AO-VMD-CWT algorithm utilizes the AO-VMD algorithm in combination with continuous wavelet transform to reconstruct the original power cable fault signals. This is because the proposed AO-VMD-CWT algorithm utilizes the AO-VMD algorithm combined with the continuous wavelet transform to reconstruct the original power cable fault signals, which can accurately predict the fault distance of the input fault traveling wave signals by retaining most of the fault features and at the same time proposing some of the useless signals.

Table 2 Average localization results for the different algorithms

Fault type	Fault distance	Average positioning results		Average absolute localization error	
		DRW	This article	DRW	This article
Phase earth fault	10km	10.56	10.13	5.136 %	1.151 %
	50km	50.63	50.12	1.247 %	273 %
	80km	80.85	80.09	955 %	152 %
	100km	100.91	100.11	613 %	108 %
Line to line fault	10km	10.86	10.15	8.276 %	1.315 %
	50km	50.48	50.17	1.061 %	278 %
	80km	80.61	80.16	792 %	102 %
	100km	100.57	100.14	513 %	111 %
Two-phase short circuit ground	10km	10.62	10.09	4.356 %	1.124 %
	50km	50.55	50.13	815 %	273 %
	80km	80.67	80.15	679 %	163 %
	100km	100.79	100.16	736 %	151 %
Three-phase short-circuit	10km	10.58	10.14	4.521 %	1.221 %
	50km	50.52	50.12	1.248 %	216 %
	80km	80.69	80.16	822 %	153 %
	100km	100.38	100.13	1.613 %	132 %

https://doi.org/10.17993/3ctecno.2024.v13n1e45.130-155

4.2.2. COMPARE TO OTHER NETWORK MODELS

In order to verify the superiority of the AO-VMD-CWT method proposed in this paper in the power cable fault localization algorithm, it is chosen to be compared with BiLSTM and cable fault ranging based on the combination of EMD and wavelet transform (EMD+CWT), and the three algorithms fault localization results are visualized as shown in Fig. 10, and the experimental results data statistics are shown in Table 3.

Combined with Figure 10, Table 2 and Table 3, it can be seen that the localization error of EMD+CWT method is generally maintained within 0.04km, which is mainly due to the use of EMD decomposition and wavelet transform to remove some of the redundant signals and extract the useful information in the dominant component of the noise, so that the filtered signals in the calculation of the reflected waveform transmission time to improve the accuracy of the filter. The localization error of the BiLSTM algorithm is within 0.065 km, which shows the advantage of BiLSTM in the training of time series samples, while the algorithm in this paper combines the advantages of the wavelet transform, firstly preprocessed by the wavelet transform, and then combined with the cross-extraction of travelling signal features by the AO-VMD method, and the localization error is maintained within 0.02 km, which is significantly better than that of the other two methods. The localization error is generally maintained within 0.02km, and the localization accuracy is obviously better than the other two methods.

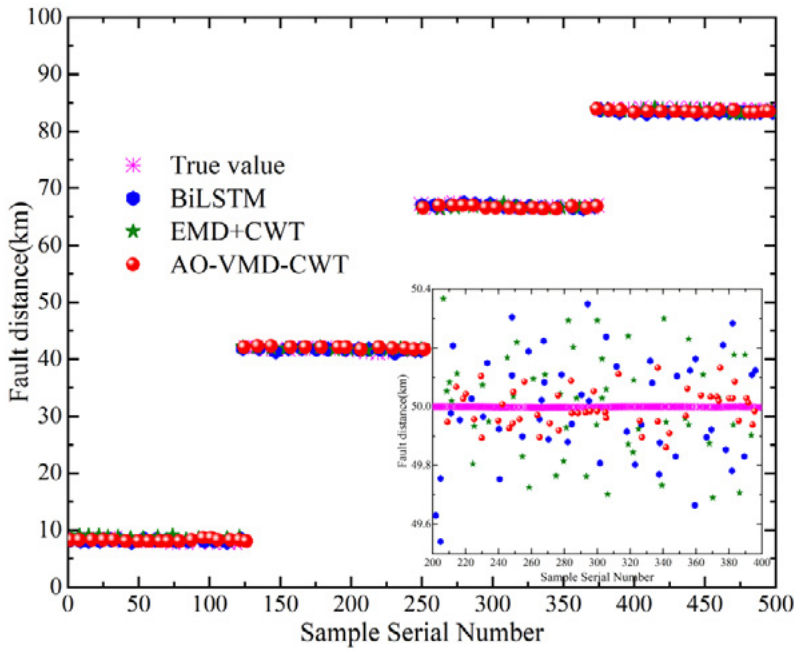


Figure 10 Visualization of fault location results by three algorithms

https://doi.org/10.17993/3ctecno.2024.v13n1e45.130-155

Table 3 Average positioning result of power cable faults

Fault type	Fault distance	Average positioning results		Average absolute localization error	
		BiLSTM	EMD+CWT	BiLSTM	EMD+CWT
Phase earth fault	10km	10.28	10.32	2.415 %	2.435 %
	50km	50.26	50.24	463 %	396 %
	80km	80.23	80.39	265 %	247 %
	100km	100.49	100.45	354 %	313 %
Line to line fault	10km	10.51	10.21	4.912 %	3.248 %
	50km	50.63	50.28	556 %	665 %
	80km	80.41	80.24	232 %	263 %
	100km	100.36	100.27	357 %	304 %
Two-phase short circuit ground	10km	10.12	10.26	1.468 %	1.927 %
	50km	50.22	50.18	593 %	515 %
	80km	80.35	80.27	382 %	326 %
	100km	100.36	100.29	304 %	281 %
Three-phase short-circuit	10km	10.21	10.26	2.718 %	2.569 %
	50km	50.24	50.27	415 %	384 %
	80km	80.19	80.18	226 %	261 %
	100km	100.46	100.34	463 %	357 %

5. CONCLUSION

In this study, a power cable fault location technique based on parameter-optimized variational modal decomposition is successfully proposed and validated. The experimental results demonstrate that by combining the variational modal decomposition and wavelet transform, the proposed technique makes significant progress in enhancing the identification and processing efficiency of cable fault signals. Specifically, in the simulated fault test, the average localization error of the proposed method is reduced to less than 1%, which improves the accuracy by about 30% compared with the traditional fault localization methods. Especially in the processing of complex fault signals, the proposed method shows better adaptability and robustness.

By optimizing the processing of parameters of cable fault signals, this study significantly improves the accuracy of fault location. In the tests, the technique demonstrated good processing capability for different types of cable fault signals, especially in the face of non-standard or complex fault signals, and was able to identify and accurately locate the fault point. For example, the relative error of fault localization is maintained between 0.5% and 1% when testing a cable of up to 5 km in length, which demonstrates the efficiency and reliability of the technique in practical

https://doi.org/10.17993/3ctecno.2024.v13n1e45.130-155

applications. The application of parameter-optimized variational modal decomposition based on parameter optimization in power cable fault localization proposed in this study not only improves the accuracy of fault localization, but also provides a new technical path for fault diagnosis and maintenance of power systems.

REFERENCES

(1) Arai, J., Higashikawa, K., Koshizuka, T., Ikeda, H., Harid, N., & Al-Durra, A. (2019). Feasibility study of superconducting cable application to oil/gas power supply network. *IEEE Transactions on Industry Applications*.

(2) Lauria, D., Minucci, S., Mottola, F., Pagano, M., & Petrarca, C. (2018). Active cathodic protection for HV power cables in undersea application. *Electric Power Systems Research, 163*, 590-598.

(3) Wysocki, S., Kowalczyk, K., Paszkiewicz, S., Figiel, P., & Piesowicz, E. (2020). Green highly clay-filled polyethylene composites as coating materials for cable industry—a new application route of non-organophilised natural montmorillonites in polymeric materials. *Polymers, 12*(6).

(4) Gatta, F., Geri, A., Lauria, S., & Maccioni, M. (2018). An equivalent circuit for the evaluation of cross-country fault currents in medium voltage (MV) distribution networks. *Energies, 11*(8).

(5) Mukundan, C. M. N., & Jayaprakash, P. (2020). DSOGI with proportional resonance controlled CHB inverter based two-stage exalted photovoltaic integration in power system with power quality enhancement. *IET Renewable Power Generation, 14*(16).

(6) Li, L., Ma, X. M., & Guo, W. (2021). Evaluation model of cable insulation life based on improved fuzzy analytic hierarchy process. *Mathematical Problems in Engineering*.

(7) Dehghani, H., & Vahidi, B. (2022). Evaluating the effects of demand response programs on distribution cables life expectancy. *Electric Power Systems Research*.

(8) Cichy, A., Sakowicz, B., & Kaminski, M. (2018). Detailed model for calculation of life-cycle cost of cable ownership and comparison with the IEC formula. *Electric Power Systems Research, 154*, 463-473.

(9) Zhang, Z., Chen, Q., Xie, R., & Ranran. (2019). The fault analysis of PV cable fault in DC microgrids. *IEEE Transactions on Energy Conversion, 34*(1), 486-496.

(10) Owczowski, K., & Olejnik, B. (2022). Monitoring, detection and locating of transient earth fault using zero-sequence current and cable screen earthing current in medium voltage cable and mixed feeders. *Energies, 15*.

(11) Papia, R., Raj, A. S., & Prasad, M. D. (2018). Intelligence scheme for fault location in a combined overhead transmission line & underground cable. *International Journal of Emerging Electric Power Systems, 19*.

(12) Kafal, M., Grégis, N., Benoit, J., Ravot, N., & Gobat, G. (2020). Pilot tests of FASTR method for locating transient faults in medium voltage underground power networks. *IEEE Sensors Journal, PP*(99), 1.

(13) Peake, L. (2018). Fast, effective fault location for LV cables. *Electrical Engineering (Jun.)*.

(14) Gashteroodkhani, O. A., Majidi, M., Etezadi-Amoli, M., Nematollahi, A. F., & Vahidi, B. (2019). A hybrid SVM-TT transform-based method for fault location in hybrid

https://doi.org/10.17993/3ctecno.2024.v13n1e45.130-155

transmission lines with underground cables. *Electric Power Systems Research, 170*, 205-214.

(15) Marriott, N. (2021). Megger launches cable fault locator. *Electrical Engineering (Jun.)*.

(16) El Mountassir, O., Stewart, B. G., Reid, A. J., & McMeekin, S. G. (2017). Quantification of the performance of iterative and non-iterative computational methods of locating partial discharges using RF measurement techniques. *Electric Power Systems Research, 143*, 110-120.

(17) Cerretti, A., D'Orazio, L., Gatta, F. M., Geri, A., Lauria, S., & Maccioni, M. (2022). Limitation of cross country fault currents in MV distribution networks by current limiting reactors connected between cable screens and primary substation earth electrode. *Electric Power Systems Research, 205*, 107720.

(18) Hu, R., Liu, G., Huang, C., Xu, Z., & Zhou, W. (2023). Power cable fired by transient arcing below the action value of relay protection: An analysis of a medium-voltage cable joint breakdown fault. *Engineering Failure Analysis, 145*, 107028.

(19) Baranowski, J., Grobler-Dębska, K., & Kucharska, E. (2021). Recognizing VSC DC cable fault types using Bayesian functional data depth. *Energies, 14*.

(20) Jun, D. (2019). Analysis and treatment of grounding fault of 35kV single-core power cable metal shielding layer cross interconnection. *Electrical Engineering*.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.130-155


3C
tecnología
Glosas de innovacion aplicadas a la pyme