ISSN 2616-5767 Vol.7, Issue 3: 166-172, DOI: 10.25236/AJETS.2024.070323

Research on Fault Diagnosis of Transmission Lines Based on Machine Learning

Wang Wei, Jiang Yonglin, Yin Jinyang, Li Xiaoyang, Huang Yingjie, Qiao Huilong

Shandong University of Technology, Zibo, 255000, China

Abstract: *Transmission line maintenance is crucial for the stable operation of power systems, as any faults can lead to severe consequences. With the increasing complexity of power systems, traditional fault diagnosis methods are no longer applicable. Based on this, the article explores the possibility of using machine learning techniques for transmission line fault diagnosis. By analyzing historical fault data, machine learning models can identify complex patterns and correlations to achieve fault prediction and diagnosis. Firstly, different types of short-circuit faults are analyzed, including three-phase short circuits, two-phase short circuits, two-phase ground short circuits, and single-phase ground short circuits, and a classification algorithm based on Linear Discriminant Analysis (LDA) is proposed. Secondly, the Box-Muller transformation is used to generate Gaussian distributed random variables from a uniform distribution to simulate fault data. The algorithm validation results show that the proposed method can effectively diagnose transmission line faults and provide strong technical support for the stable operation of power systems.*

Keywords: *machine learning; fault diagnosis; transmission line fault; LDA algorithm*

1. Introduction

As a critical component of the power system, the stability of transmission lines is directly related to the normal operation of socio-economic activities. Any faults may lead to severe socio-economic consequences, including but not limited to industrial production disruptions, traffic congestion, communication interruptions, and even affecting medical and rescue services. Therefore, ensuring the reliability of transmission lines is the top priority for power engineers and researchers. With technological advancements, traditional fault diagnosis methods are gradually becoming inadequate for modern power systems. These methods often rely on expert systems and rule engines, but their efficiency and accuracy are limited when dealing with large-scale, multivariable, and nonlinear problems. Furthermore, as the complexity of power systems increases, the response speed and accuracy requirements for fault diagnosis systems are also becoming higher. In this context, machine learning techniques exhibit tremendous potential. By leveraging historical data, machine learning models can identify complex patterns and correlations, enabling the prediction and diagnosis of potential faults. This data-driven approach not only improves diagnostic accuracy but also allows for real-time processing of large amounts of sensor data, enabling rapid response[1]. For example, deep learning algorithms can automatically extract fault features by analyzing historical fault data and establish predictive models. These models can provide predictions and warnings before faults occur, enabling operation and maintenance personnel to take prompt measures to avoid or mitigate the impact of faults. Machine learning models can continuously learn and adapt to changes in the power system, constantly improving their diagnostic capabilities. Fault diagnosis of transmission lines based on machine learning is crucial for the stable operation of power systems[2].

2. Classification and Characteristics of Transmission Line Short-Circuit Faults

2.1. Three-Phase Short Circuit

A three-phase short circuit fault is one of the most severe types of faults in transmission lines, involving all three phases simultaneously coming into contact with the ground or each other. This leads to a sudden increase in current and a decrease in voltage, posing a serious threat to the stability and safety of the power system. When a three-phase short circuit fault occurs, there is a sudden increase in

ISSN 2616-5767 Vol.7, Issue 3: 166-172, DOI: 10.25236/AJETS.2024.070323

current in all three phases, with the amplitude of the three-phase currents being close to each other. This is because the resistance at the short-circuit point is very small relative to the power source and can almost be considered zero. Instantaneously after the fault occurs, the three-phase voltage drops sharply to near zero due to the voltage reduction at the short-circuit point and the increased voltage drop along the line caused by the increased current. A three-phase short circuit fault introduces high-frequency components into the power system's frequency spectrum, which can be analyzed using methods such as Fourier transforms.

2.2. Two-Phase Short Circuit

A two-phase short circuit fault is a common type of fault in transmission lines, occurring between two phase lines. It is typically caused by aging of insulation materials, intrusion of external objects, or other mechanical damage. This fault results in unbalanced operation of the power system, affecting grid stability[3]. When a two-phase short circuit fault occurs, the currents in the affected two phases increase significantly, while the current in the unaffected phase may decrease slightly or remain unchanged. The voltages of the affected two phases decrease, while the voltage of the unaffected phase may increase due to the relationship between phase voltage and line voltage in the power system. Additionally, a two-phase short circuit fault causes changes in the energy distribution of the power system, often manifesting as energy concentration near the fault point[4].

2.3. Two-Phase to Ground Short Circuit

A two-phase to ground short circuit fault occurs when two phase lines simultaneously short-circuit to the ground. This fault leads to imbalance in the power system, causing malfunction of protective devices. During a two-phase to ground short circuit, the currents in the two affected phases rise sharply, and the ground current increases as the fault current needs to flow back through the ground. The voltages of the affected two phases drop significantly, while the voltage of the unaffected phase may rise.

2.4. Single-Phase to Ground Short Circuit

A single-phase to ground short circuit is typically caused by insulation damage or external objects, resulting in contact between one phase line and the ground. This fault leads to asymmetrical loading of the power system. During a single-phase to ground short circuit, the current in the faulted phase increases significantly, while the currents in the other two phases remain largely unchanged. Simultaneously, the ground current increases as the fault current needs to flow back through the ground. The voltage of the faulted phase drops to near zero, while the voltages of the other two phases rise.

3. Dataset for Transmission Line Faults

The first step in fault diagnosis lies in constructing a fault feature dataset. Sensor data such as voltage and current from transmission line short-circuit faults need to be aggregated and transmitted to the main station through substation communication interfaces, forming a feature value dataset and an expected result dataset. Through data analysis, key factors that contribute to building dataset features and state identification can be extracted[5]. Constructing an accurate dataset is crucial for transmission line fault diagnosis. The dataset construction process typically involves the following steps:

(1) Data Source Collection: Firstly, raw data is collected from various substations (such as Substation 1 to Substation N) located in different locations. These data sources may include voltage, current, temperature, and other sensor data.

(2) Feature Extraction: After preliminary processing of the raw data, the next step is to extract key feature quantities. These feature quantities serve as inputs to the fault diagnosis model and need to accurately reflect the operating state of the system[6].

(3) State Variable Set: In addition to feature quantities, it is also necessary to determine the state variables of the system, which may include normal operation, various types of faults, and other states.

(4) Secondary Processing: The extracted feature quantities and state variables undergo secondary processing to form a data format suitable for use in machine learning models. This may include data cleaning, normalization, encoding, and other steps[7].

ISSN 2616-5767 Vol.7, Issue 3: 166-172, DOI: 10.25236/AJETS.2024.070323

(5) Dataset Integration: Finally, the feature quantities from multiple sensors are combined with the expected results (i.e., state variables) to form a complete dataset. This dataset will be used to train machine learning models for automatic fault diagnosis and classification. The specific steps are illustrated in Figure 1:

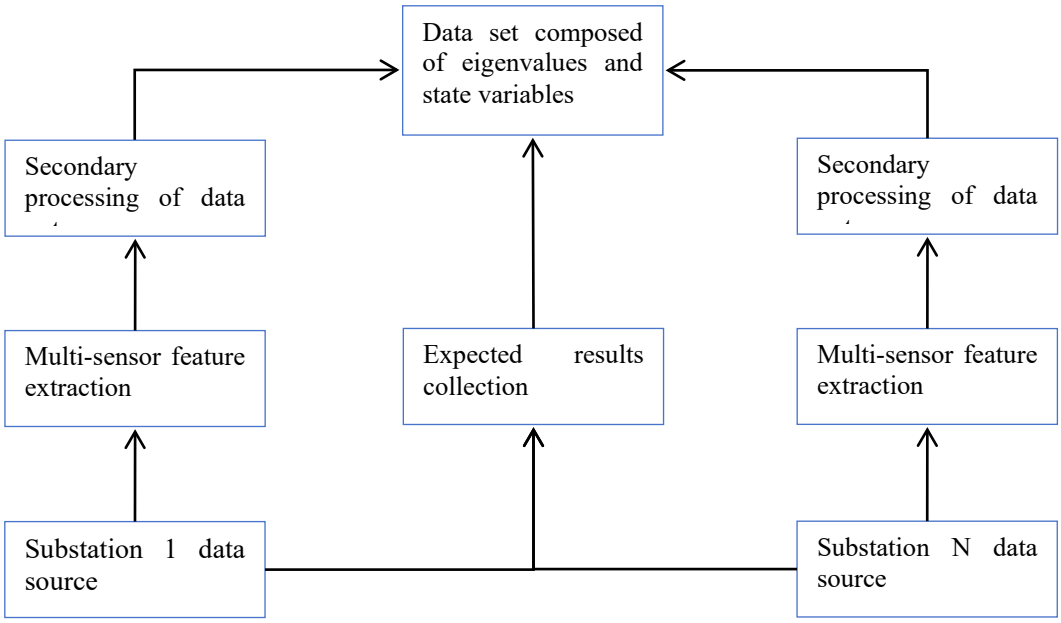

Figure 1: Construction process of fault feature dataset

After the dataset construction steps, a feature matrix containing multi-sensor information and fault states is finally formed. Assuming that the dataset contains N records, M features, and K states, the dataset Y can be represented by the following mathematical expression:

$$Y = \begin{bmatrix} X & S \end{bmatrix}$$

Further expansion yields the feature dataset matrix.

$$Y = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1M} & s_{11} & s_{12} & \cdots & s_{1K} \\ x_{21} & x_{22} & \cdots & x_{2M} & s_{21} & s_{22} & \cdots & s_{2K} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NM} & s_{N1} & s_{N2} & \cdots & s_{NK} \end{bmatrix}$$

Where x_{nm} represents the mth eigenvalue of the nth record, and s_{nk} represents the kth state identifier of the nth record.

4. Transmission Line Fault Feature Classification Algorithm Based on Machine Learning

In the field of machine learning, Linear Discriminant Analysis (LDA) is an important classification algorithm that distinguishes different categories by projecting data from a high-dimensional space to a low-dimensional space. The core idea of LDA is to maximize the between-class scatter matrix and minimize the within-class scatter matrix, so that the projected data has better classification results in the low-dimensional space.The mathematical expression of the LDA algorithm can be simplified as

$$y = W^T X$$

Where W is the dimensionality reduction matrix and X is the feature vector of the training set. In practical fault diagnosis, there are usually multiple feature vectors and multiple categories of fault states, so it is necessary to construct a matrix containing all features and states[8].

The Fisher discriminant is the key to the LDA algorithm, and its goal is to find the optimal

ISSN 2616-5767 Vol.7, Issue 3: 166-172, DOI: 10.25236/AJETS.2024.070323

projection direction so that the distribution of data from different categories has the maximum difference in this direction. This can be achieved by solving the following optimization problem:

$$J(W) = \frac{|W^T S_B W|}{W^T S_W W}$$

Where S_B is the between-class scatter matrix and S_W is the within-class scatter matrix.

The between-class scatter matrix S_B is defined as:

$$S_B = \sum_{i=1}^{C} N_i (\mu_i - \mu)(\mu_i - \mu)^T$$

The within-class scatter matrix SW is defined as:

$$S_W = \sum_{i=1}^{C} \sum_{x \in \omega_i} (x - \mu_i)(x - \mu_i)^T$$

The overall expectation μ of the dataset is calculated as follows:

$$\mu = \frac{1}{N} \sum_{\forall x} x$$

The formula for calculating the mean μ_i of category (i) is:

$$\mu_i = \frac{1}{N_i} \sum_{x \in \omega_i} x$$

where C is the total number of categories, N_i is the number of samples in category i, μ_i is the expectation of category i, μ is the overall mean of all samples, x is the sample point, and ωi is the sample set of category i.

To solve this optimization problem, it is necessary to calculate the overall expectation μ of the fault feature dataset, the expectation μi of each class, and the corresponding scatter matrix. To obtain the maximum value of Fisher's discriminant, let:

$$\frac{d}{d(W)} J(W) = 0$$

Obtain:

$$S_W^{-1} S_B W = \lambda W$$

The final solution W is the optimal dimensionality reduction matrix, where λ is the largest eigenvalue and matrix W^* is used to project high-dimensional data into a low-dimensional space for effective classification.

In power system fault diagnosis, the LDA algorithm can project multi-dimensional fault data into a lower-dimensional space. The reduced data is assumed to follow a Gaussian distribution, allowing the calculation of the probability density function (PDF) of the projected data based on mathematical expectations and variances. From the projected data in the test set, the type of short-circuit fault and its confidence level can be inferred[9]. For example, a dataset consisting of three fault states would be projected into a two-dimensional space by the LDA algorithm. In this two-dimensional space, data points for each fault state would cluster around the center of their respective categories, forming a Gaussian distribution. By leveraging the characteristics of these distributions, the probability of a new test data point belonging to each fault state can be calculated, thereby diagnosing the fault type.

5. Machine learning-based fault feature diagnosis algorithm for transmission lines

Transmission line fault feature data can be modeled using Gaussian-distributed random variables, and fault diagnosis can be performed based on the LDA (Linear Discriminant Analysis) fault

ISSN 2616-5767 Vol.7, Issue 3: 166-172, DOI: 10.25236/AJETS.2024.070323

classification and diagnosis algorithm. Data simulation can be used to generate Gaussian-distributed random variables from a uniform distribution using the Box-Muller transformation. The Box-Muller transformation is a statistical method used to generate random variables that follow a normal distribution from random variables that obey a uniform distribution. This transformation is particularly useful because many statistical algorithms and machine learning models assume that the data follows a normal distribution, whereas in practical applications, we can typically only easily generate uniformly distributed random numbers. Specifically, the Box-Muller transformation requires two independent random variables, U_1 and U_2, that follow a uniform distribution on the interval [0,1]. Then, through the following transformation formulas, two independent random variables, Z_0 and Z_1, that follow a standard normal distribution (with a mean of 0 and a variance of 1) can be generated:

$$Z_0 = \sqrt{-2\ln U_1}\cos(2\pi U_2)$$

$$Z_1 = \sqrt{-2\ln U_1}\sin(2\pi U_2)$$

This process can be visually understood as follows: Firstly, a random point is generated within a unit circle using U_1 and U_2. Then, through polar coordinate transformation, this point is mapped onto a two-dimensional Gaussian distribution. As a result, Z_0 and Z_1 become random variables that follow a two-dimensional standard normal distribution. By employing this method, samples that conform to the distribution of actual fault data can be simulated, enabling the testing and validation of the LDA algorithm. The basic flowchart is shown in Figure 2.

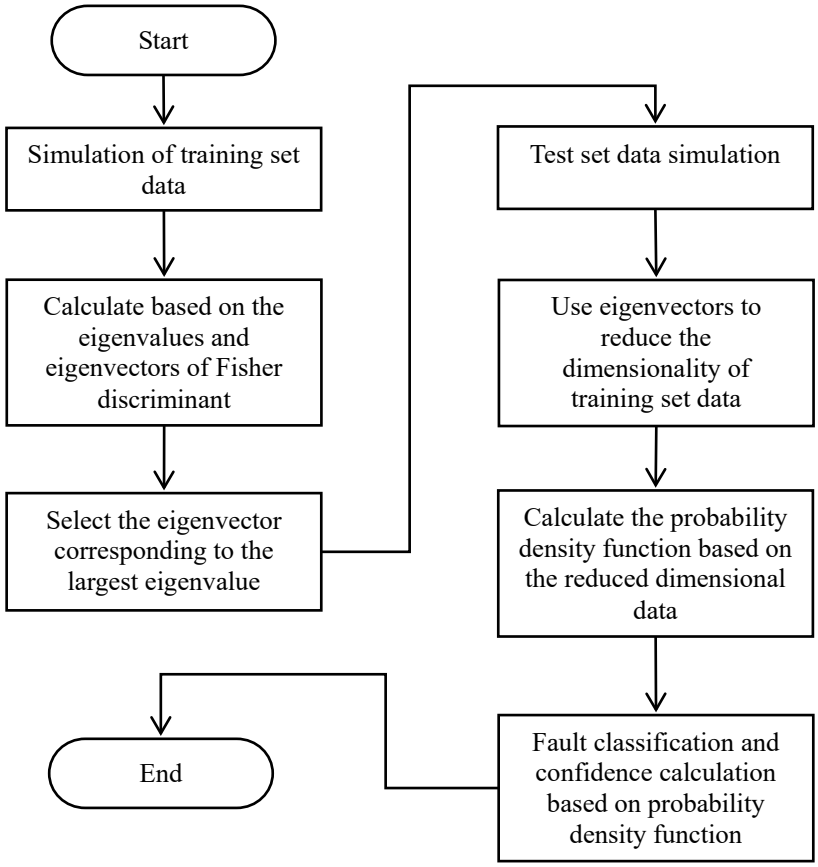

Figure 2:Fault Diagnosis Algorithm Process

6. Algorithm Verification

Firstly, three sets of simulated fault state data are generated by setting specific expectation values and variances to ensure accurate simulation of system behavior under different fault conditions. These three sets of data represent the system's non-fault state, fault type A, and fault type B, respectively. The specific expectation values and variances are set as follows:

The expectation value for the non-fault state is [2.5, 12.5], and the variance is [4.5, -0.5; -2.5, 2.5].

ISSN 2616-5767 Vol.7, Issue 3: 166-172, DOI: 10.25236/AJETS.2024.070323

The expectation value for fault type A is [2, 1], and the variance is [3.5, 0; 0, 3.5].The expectation value for fault type B is [11.5, 9.5], and the variance is [3, 0.5; 2.5, 2].

Secondly, the training dataset is used to calculate the maximum eigenvalue and its corresponding eigenvector. In this example, the maximum eigenvalue is (4945), and the corresponding eigenvector is [0.8932, 0.7343].

Thirdly, the test dataset is used to calculate the Probability Density Function (PDF) to determine the type of relay protection fault and its confidence level in the test data.

Finally, the confidence level of each state is evaluated based on the calculated PDF, as shown in Table 1.

Table 1: Confidence Levels in Three States

PDF / Fault Probability	Normal	Fault State A	Fault State B
Normal (%)	99.78	0.21	0
Fault State A (%)	0.15	97.86	2.02
Fault State B (%)	0	2.84	97.59

As can be seen from Table 1, in the normal state, the system correctly diagnoses the normal state with a high confidence level of 99.78%, while the probability of misdiagnosing the normal state as Fault State A or Fault State B is extremely low (0.21% and 0%, respectively). For Fault State A, the system correctly diagnoses it with a confidence level of 97.86%, with a 0.15% probability of misdiagnosis as normal and a 2.02% probability of misdiagnosis as Fault State B. In the case of Fault State B, the system also demonstrates a high confidence level (97.59%) in correctly diagnosing Fault State B, with only a 2.84% probability of misdiagnosing it as Fault State A and no misdiagnosis as normal. The transmission line fault diagnosis system exhibits high accuracy and reliability in various states. Especially in the normal state, there is almost no misdiagnosis, which is crucial for avoiding unnecessary maintenance and inspections. Meanwhile, even in fault states, the system can accurately diagnose the fault type with high confidence, which helps to quickly and accurately locate the problem for timely repair.

7. Conclusion

In summary, the fault diagnosis model based on machine learning can effectively identify and classify short-circuit faults in transmission lines, providing timely decision support for operation and maintenance personnel. Algorithms such as Linear Discriminant Analysis (LDA) and Box-Muller transformation can efficiently and accurately extract and classify fault features. Therefore, it is necessary to further explore the application of advanced algorithms such as deep learning in fault diagnosis to improve the generalization ability and accuracy of the model. Consider integrating the fault diagnosis system with other intelligent management systems of the power system to form a comprehensive smart grid management platform.

References

[1] Chen Hanxiang. Intelligent Identification of Bird Species Involved in Transmission Line Faults Based on Birdsong Processing and Machine Learning [D]. Jiangxi: Nanchang University, 2021.
[2] Shi Wanyu. Research on Lightning Fault Analysis of Transmission Lines Based on Machine Learning [D]. North China Electric Power University, 2022.
[3] Liu Yingpei, Wang Xiangyu, Wang Xinming, et al. Fault Identification Method for De-energized Transmission Lines Based on Induced Voltage Characteristics [J]. Journal of North China Electric Power University (Natural Science Edition), 2022, 49(4): 14-22, 32.
[4] Chen Huoxing, Zhang Jun, Hailong. Fault Statistics and Analysis of Transmission Lines Based on Machine Learning [J]. Science and Technology Information, 2018, 16(34): 56-57.
[5] Li Ting, Zhang Xiaojun, Pan Hua, et al. Fault Type Diagnosis Method for Transmission Lines Based on Multi-source Information Fusion of Operation and Inspection Control Platform [J]. Southern Power System Technology, 2023, 17(12): 109-118.
[6] Li Shaojun. High-precision Positioning Method for Lightning Faults of Overhead Transmission Lines Based on Wavelet Packet Transform [J]. Telecom Power Technology, 2023, 40(7): 76-79.

ISSN 2616-5767 Vol.7, Issue 3: 166-172, DOI: 10.25236/AJETS.2024.070323

[7] MAHRUKH ANSARI. Fault Prediction of High-voltage Overhead Transmission Lines Based on Clustering Methods [D]. North China Electric Power University, 2021.
[8] Ni Chen. Fast Fault Detection and Identification of Transmission Lines Based on Convolutional Neural Networks [D]. Inner Mongolia: Inner Mongolia University of Technology, 2019.
[9] Li Hang. Optimization of Fault Detection and Maintenance Strategies in Transmission Line Operation and Inspection [J]. Encyclopedia Forum Electronic Magazine, 2023(11): 55-57.