Prediction of the Optimal Umbrella Shape of Insulators Based on Data Mining Technology

Xiangrong Xing

International College, Beijing University of Posts and Telecommunications, Beijing, 102206

*Corresponding author: xing_xiangrong@bupt.edu.cn

Abstract. The insulator flashover process is affected by many factors. In order to better study the factors that affect the insulator flashover voltage, and then accurately predict the insulator flashover voltage. In this paper, 11 different types of composite insulators are studied through artificial pollution tests. Data mining technology has been widely used now. In order to better apply data mining in the prediction of insulation flashover voltage, this paper adopts the BP neural network in data mining technology to predict the flashover voltage of different umbrella-shaped insulators. In addition, in order to further improve the prediction accuracy of the model, this paper uses the global search capability of genetic algorithm to optimize the initial weights and thresholds of the BP neural network, and optimizes the BP neural network prediction model, and established an optimization model based on genetic algorithm to optimize BP neural network. Through this method, the flashover voltage of the insulator can be predicted in time, which is of great significance for maintaining the safe and stable operation of the power system.

Keywords: Data mining, Genetic algorithm, BP neural network, Pollution flash voltage, Predictive model.

1. Introduction

Umbrella skirts provide the main creepage distance for composite insulators, which is one of the important reasons to ensure that composite insulators have excellent pollution flashover resistance. However, there is a problem that cannot be ignored: If the design of composite insulators mainly considers the structural height and creepage distance to meet certain requirements, and does not consider the design of the shed structural parameters of the composite insulator, the supplier can determine the structural height and creepage distance. In the case of, a variety of composite insulators with different umbrella skirt structure parameters are given. If the structure of the umbrella skirt of the composite insulator is unreasonable, the arc bridging between the umbrella skirts is likely to occur, which will lead to pollution flashover accidents and seriously threaten the safe and reliable operation of the power grid. Therefore, it is necessary to study the effect of umbrella skirt structure on composite insulator. The influence of DC pollution flashover characteristics, and get the prediction model of different umbrella insulator flashover voltage.
Kiruthika M [1-3] et al used the solid layer method to conduct artificial pollution tests with four different types of 11kV silicone rubber insulators as materials, and obtained flashover voltages at different equivalent salt densities. Based on the prediction model of artificial neural network and least square support vector machine, the flashover voltage is predicted with input parameters such as height, creepage length, diameter and other input parameters. The final conclusion is LS-SVM has better auxiliary flashover voltage prediction capability than ANN. Chongqing University Shi Yan et al [4] established a radial basis function neural network (RBF) network-based prediction model for flashover voltage of ice-coated insulators based on experimental research on LXZP-160 insulators, which can reflect the ice-coated insulators Non-linear relationship model between flashover voltage and insulator string length, pollution degree, air pressure, conductivity of ice-coated water and other factors. Shu Lichun of Chongqing University et al [5] proposed a support vector machine-based insulator flashover voltage prediction method based on the XZP-160 insulator test data through the artificial climate laboratory, able to meet engineering needs.

Based on this, this article takes the umbrella skirt structure height H, creepage distance L, umbrella spacing S, large umbrella extension $P_1$, small umbrella extension $P_2$ and salt density as parameters, and fully considers the umbrella parameters that affect the flashover voltage of insulators. The Sahanro voltage of the insulator is predicted by data mining technology.

2. BP neural network model optimized by genetic algorithm.

2.1. BP neural network
BP neural network is a type of multi-layer feedforward neural network for learning and training based on error back propagation. According to the "negative gradient descent" theory, the error adjustment direction of the network always follows the direction of the fastest error reduction. Adjust the weights and thresholds of the network until the desired result is entered. The BP neural network mainly includes an input layer, a hidden layer, and an output layer. The basic topology of the three-layer neural network corresponding to its multiple input and single output is shown in Figure 1.

![BP neural network structure](image)

**Figure 1.** BP neural network structure.

2.2. Genetic algorithm optimizes BP neural network

2.2.1. Basic ideas. Genetic algorithm is a search method based on natural selection and genetic inheritance, which can solve many problems that are difficult to solve by traditional optimization methods. Usually based on the huge population size, the algorithm can obtain multiple extreme points
at the same time in the optimization process, so as to ensure that the algorithm is not easy to fall into the local optimal solution. Its main feature is that there is no derivative sum function. The limitation of continuity, parallelism and the ability of global optimization.

The BP neural network is generally initialized to random numbers when assigning weights and thresholds. The initial weights and thresholds obtained in this way have a great impact on the training results of the network. Genetic algorithm has a strong global search ability, which can make up for the limitations of BP neural network local search. Based on this, this paper uses genetic algorithms to optimize the weights and thresholds of the BP neural network, finds the optimal initial weights and thresholds of the network, and then learns and trains the BP neural network model. The specific optimization process is shown in Figure 2.

![Figure 2. Genetic algorithm optimization BP neural network flow chart.](image-url)
2.2.2. Optimization steps.

1) Topological structure of BP neural network. In this paper, the input is salt density and gray density, and the output is pollution flashover voltage, the hidden node of the network is set to 1, and the number of hidden nodes can be determined by the following empirical formula:

\[ l = \sqrt{m + n + a} \]  

(1)

Where: \( m \) and \( n \) are the number of input nodes and output nodes respectively; \( a \) is a constant between 1 and 10.

2) Preprocessing of sample data. The size of data input and output will affect the convergence of the network, so it is necessary to normalize the parameters. Perform normalization processing according to formula (2) to make the change interval in \([-1, 1]\):

\[ x_k = \frac{x_k - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \]

(2)

3) Determine the weights and thresholds of the network and perform population initialization coding.

4) Select operation. This article uses the roulette method to operate, the probability of a single individual \( i \) being selected is

\[ p_i = \frac{f_i}{\sum_{j=1}^{p} f_j} \]

(3)

Where: \( f_i \) is the reciprocal of individual fitness, and \( p \) is the population size.

5) Cross operation. This paper adopts the real number crossover method, the cross operation method of \( k \) chromosome \( a_k \) and \( l \) chromosome \( a_l \) at \( j \) position is as follows

\[
\begin{align*}
    a_{ij} &= a_{ij}(1-b) + a_{lj} \\
    a_{ij} &= a_{ij}(1-b) + a_{lj}b
\end{align*}
\]

(4)

Where: \( b \) is a random number in the interval \([0, 1]\).

6) Mutation operation. Select the \( j \)-th gene \( a_{ij} \) of the \( i \)-th individual for mutation, and the mutation operation method is

\[
\begin{align*}
    a_{ij} &= a_{ij} + (a_{ij} - a_{\text{max}}) \times f(g) \quad r > 0.5 \\
    a_{ij} &= a_{ij} + (a_{\text{min}} - a_{ij}) \times f(g) \quad r \leq 0.5
\end{align*}
\]

(5)

Where: \( a_{\text{max}} \) is the upper bound of gene \( a_{ij} \); \( a_{\text{min}} \) is the lower bound of gene \( a_{\text{max}} \); \( f(g) = r_2(1 - g / G_{\text{max}})^2 \), \( r_2 \) is a random number, \( g \) is the current iteration number, \( G_{\text{max}} \) is the maximum number of evolution; \( r \) is a random number in the interval \([0, 1]\).

7) By solving the optimal chromosome of the genetic algorithm, the optimal weight and threshold of the BP network are obtained, and the optimal value is assigned to the prediction model for training to obtain the optimal prediction value.
3. Test plan

3.1. Samples
In this paper, 11 kinds of composite insulators with different structures are used as the research object, and their numbers and specific umbrella parameters are shown in Table 1.

| Serial number | H/mm | L/mm | S/mm | P1/P2/mm |
|---------------|------|------|------|-----------|
| 1             | 834  | 1400 | 78   | 145/115   |
| 2             | 799  | 1400 | 78   | 145/115   |
| 3             | 815  | 1400 | 78   | 145/115   |
| 4             | 738  | 1600 | 78   | 145/115   |
| 5             | 778  | 1600 | 78   | 145/115   |
| 6             | 760  | 1600 | 50   | 161/116   |
| 7             | 760  | 1600 | 51   | 192/140   |
| 8             | 769  | 1600 | 78   | 145/115   |
| 9             | 864  | 1600 | 31   | 90/72     |
| 10            | 800  | 1600 | 78   | 145/125   |
| 11            | 811  | 1600 | 78   | 145/115   |

3.2. Test device
This article adopts the artificial pollution test, the test is carried out in the artificial fog chamber, the smearing of the sample refers to the quantitative smearing method in the solid coating method recommended by GB/T 4585, and the boosting method adopts the uniform boosting method. The test circuit is shown in Figure 3.

![Test circuit diagram](image)

Figure 3. Test circuit diagram.

3.3. Test plan
This test refers to the recommendation of GB/T 4585 and recommends the quantitative brushing method for contamination, using NaCl to simulate conductive substances and diatomaceous earth to simulate insoluble substances. According to the surface area of the sample and the tested salt density and ash density, calculate and weigh out the required amount of NaCl and diatomaceous earth, and apply the configured dirt evenly on the insulating surface of the sample insulator.

The test uses the uniform boost method. Perform 4~5 flashover tests on each string of insulators; perform pollution 3 times under the same pollution degree, and take the average value of all points with an error of no more than 10% from the average value as the flashover voltage $U_f$ under the pollution degree, which is

$$U_f = \frac{\sum U_i}{N}$$ (6)
\[ \sigma = \sqrt{\frac{\sum_{i=1}^{N} (U_f - U_i)^2}{(N-1)U_f^2}} \times 100\% \]  

(7)

In the formula: \( U_f \) is the average pollution flashover voltage of the insulator, kV; \( U_i \) is the \( i \)-th pollution flashover voltage, kV; \( N \) is the number of tests; \( \sigma \) is the relative standard deviation.

4. Prediction of insulator pollution flashover voltage

4.1. Selection of sample data

When optimizing the design of the insulator, the first consideration is the structural height and the creepage distance. The creepage distance is mainly provided by the umbrella skirt. Under the same structural height, a large number of parameters can be obtained by setting the composite insulator umbrella skirt parameters. Insulators of different umbrella structures. Some research results show that simply increasing the creepage distance by increasing the umbrella spacing cannot increase the pollution flashover voltage; when the umbrella spacing is fixed, increasing the umbrella diameter will increase the creepage distance and the higher the pollution flashover voltage. However, when the umbrella diameter is too large, it is easy to cause bridging between umbrella groups, and the pollution flashover voltage will decrease, so the umbrella diameter should not be too large. Umbrella spacing and umbrella extension have a greater impact on the pollution flashover voltage, and there is a maximum value with the increase of umbrella spacing and umbrella extension. Therefore, when predicting the flashover voltage of an insulator, it is not enough to only consider the height of the structure and the creepage distance.

From the above analysis, we can see that when predicting the flashover voltage of insulators with different umbrella structures, not only the structural height and creepage distance of the insulator, but also the umbrella skirt parameters of the insulator are fully considered. For this reason, the structure height \( H \), the creepage distance \( L \), and the umbrella spacing are selected \( S \), large umbrella extension \( P_1 \), small umbrella extension \( P_2 \) and salt density are used as the input of the prediction model.

In this paper, 11 kinds of composite insulators with different structures are used as the research object. The test salt density is selected as 0.05, 0.1, 0.15, 0.2 mg/cm\(^2\), and 44 sets of pollution flashover voltage data are obtained.

4.2. Establishment of prediction model

The number of BP neural network training is 1000, the learning rate is 0.1, the target error is 0.001, the number of hidden nodes selected is 4~10, and the test one by one shows that when the hidden node is 7, the error is the smallest, so the implicit node is taken as 7. In order to obtain better prediction results, by adjusting the simulation parameters of the genetic network: population number, crossover and mutation probability, it can be concluded that when the population number is 50, the crossover probability is 0.8, and the mutation probability is 0.2, a better prediction result is obtained.

4.3. Simulation results

Through the 44 sets of data obtained from the manual pollution test, 32 sets of data of the first 9 sets of test products are selected as training samples, and 12 sets of data of the last three sets of test products are used as prediction samples. The error training results of the BP neural network and the genetic algorithm optimization BP neural network error training results are shown in Figure 4 and Figure 5, respectively.
It can be seen from Figures 4 and 5 that the number of iterations of the BP neural network is 833, and the number of iterations of the genetic algorithm to optimize the BP neural network is 168. The BP neural network optimized by the genetic algorithm significantly accelerates the convergence speed of the network.

According to the simulation results, it can be concluded that the BP neural network and genetic algorithm optimize the prediction value and relative standard deviation of the BP neural network for the predicted sample pollution flashover voltage as shown in Table 4.2. The optimized training results of the BP neural network and genetic algorithm are shown in Figure 6.

### Table 2. Prediction results of pollution flashover voltage.

| Serial number | Measured value/kV | GA-BP internet | BP internet | GA-BP | BP |
|---------------|-------------------|----------------|-------------|-------|----|
| 1             | 51.9              | 54.1           | 55.4        | 4.23  | 6.63 |
| 2             | 473.              | 45.9           | 45.2        | 2.96  | 4.41 |
| 3             | 42.6              | 41.4           | 40.1        | 2.82  | 5.87 |
| 4             | 36.7              | 37.3           | 35.4        | 1.63  | 3.66 |
| 5             | 62.3              | 64.7           | 65.5        | 3.86  | 5.07 |
| 6             | 58.8              | 59.7           | 56.4        | 1.53  | 4.11 |
| 7             | 54.9              | 51.6           | 58.9        | 6.01  | 7.28 |
| 8             | 47.8              | 45.1           | 44.6        | 5.65  | 6.32 |
| 9             | 66.7              | 60.8           | 60          | 8.85  | 10.04 |
| 10            | 62.1              | 59.5           | 58          | 4.19  | 6.71 |
| 11            | 57.6              | 56.3           | 52.1        | 2.26  | 9.55 |
| 12            | 51.9              | 53.4           | 49.3        | 2.89  | 4.93 |
| average error |                   |                |             | 3.91  | 6.22 |

![Figure 4. BP neural network training results.](image)

![Figure 5. Genetic algorithm optimization of BP neural network training results.](image)
It can be clearly seen from Table 4.2 and Figure 4.5 that the highest error of genetic algorithm optimized BP neural network is 8.85%, the lowest error is 1.53%, and the average error is 3.91%, while the highest error of ordinary BP neural network is 10.04% and the lowest error is 3.66% and the average error is 6.22%. The prediction accuracy of the genetic algorithm optimized BP neural network is generally higher than that of the BP neural network, which meets the actual prediction accuracy requirements.

5. Conclusion

1) In order to predict the flashover voltage of insulators more accurately, this paper proposes to fully consider the parameters that affect the flashover voltage of insulators: umbrella skirt structure height H, creepage distance L, umbrella spacing S, large umbrella extension P1, small umbrella extension P2. It provides a theoretical reference for the prediction of insulation flashover voltage.

2) Based on data mining technology, this paper establishes a prediction model of insulator flashover voltage under different umbrella parameters through genetic algorithm optimization of BP neural network, and verifies the accuracy of the model through artificial pollution tests.

References

[1] Shewehdi M H, Farag A S, Izzularab M A. Estimating the insulation strength of two series non-ceramic dielectricson distribution systems a statistical approach [C]. In: Electrical Insulation Conference and Electrical Manufacturing and Coil Winding Conference Rosemont: EEIC, 1997: 799–804.

[2] Kiruthika M, Sivadasan J. Evaluation and prediction of flashover voltage on contaminated composite insulators [C]. In: 2015 IEEE International Conference of Electrical, Computer and Communication Technologies, Coimbatore: ICECCT, 2015 (2): 978–982.

[3] Kontargyri V T, Gialketsi A A, Tsekouras G J. Design of an artificial neural network for the estimation of the flashover voltage on insulators [J]. Electric Power Systems Research, 2007, 77 (12): 1532-1540.

[4] SHI Yang, JIANG Xingliang. Prediction model of flashover voltage of ice-coated insulator based on RBF network [J]. High Voltage Technology, 2009, 35 (3): 591–596.

[5] Shu Lichun, Bai Qunli Prediction of insulator flashover voltage under complex environmental conditions based on support vector machine [J]. Proceedings of the Chinese Society for Electrical Engineering, 2006, 26 (17): 127-131.