Prediction of Insulator Pollution Flashover Voltage Based on Data Mining Technology

Hongmin Zhou1,*, Yutong Chen2
1School of Automation and Artificial Intelligence, Nanjing University of Posts and telecommunications, Jiangsu, Nanjing, 210023
2School of Computer Science, School of Software, School of Cyber Space Security, Nanjing University of Posts and telecommunications, Jiangsu, Nanjing, 210023
*Corresponding author: hongminzhou@njupt.edu.cn

Abstract. The pollution flashover of insulators is of great significance to the safe and stable operation of power grids. Therefore, studying the pollution flashover voltage of insulators has very important guiding value for the external insulation prevention of power grids. For this reason, this article has obtained the flashover voltage under different salt densities and gray densities through manual pollution tests. Data mining technology is now widely used. In order to better apply data mining to the insulation flashover voltage prediction, this paper adopts the support vector machine regression method in data mining technology to establish a prediction model to predict the insulator pollution flashover voltage. 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, support vector machine regression, flashover voltage, prediction model.

1. Introduction
With the rapid development of our country, there are more and more pollutants in the atmosphere, leading to the gradual accumulation of pollution on the surface of insulators, increasing the number of flashovers of insulators, and characterizing the characteristic parameters of the pollution layer when the salt is dense and the gray is dense [1]. Therefore, studying the flashover voltage under different salt densities and gray densities is of great significance for the safe operation of insulators.

The support vector machine regression method adopts the principle of structural risk minimization, which overcomes the shortcomings of experience and enlightenment in the traditional god network method, improves the confidence level, and solves the unavoidable local extreme value problem in the neural network method. It is considered to be the best theory for classification and regression of small samples. Literature [2] used BP neural network to model the relationship between equivalent salt density on the surface of insulators and meteorological factors. Literature [3] applies BP neural network to external insulation selection. However, the traditional BP neural network has the shortcomings of slow convergence and easy to fall into local minima, and the radial basis function
neural network (RBF) neural network is better than the BP network in terms of approximation ability and learning speed.

2. The basic principles of support vector machine regression

Support Vector Machines (SVM) is proposed by Vapnik and others based on the principle of structural risk minimization in statistical learning theory. Using the SVM algorithm, even if the discriminant function obtained from a limited number of samples is tested on independent samples, good results can still be obtained and the error is minimized. The algorithm first selects a nonlinear transformation, converts the data in the original space into a high-dimensional feature space, and constructs a linear discriminant function in the high-dimensional space to realize the nonlinear discriminant function in the original space. At the same time, it cleverly solves the dimensionality problem, and its algorithm complexity has nothing to do with the sample dimension. In addition, the support vector machine is a convex quadratic optimization problem that can guarantee that the extreme solution found is the global optimal solution. These characteristics make the support vector machine an excellent data-based machine learning algorithm.

The calculation method of the support vector machine algorithm can be described as assuming a training sample set, where \( x \) is the input vector, \( f(x) \) is the corresponding output value, \( n \) is the number of samples, and the function form required to be fitted is

\[
 f(x) = \mathbf{w}^T \phi(x) + b
\]

where \( \phi \) is a nonlinear mapping function, \( \xi \) and \( \xi^* \) are relaxation factors, and \( C>0 \) is a balance factor. The Lagrangian optimization method can be transformed into a dual problem, which is reduced to the condition

\[
 \begin{align*}
 y_i - [\mathbf{w}^T \phi(x_i) + b] &\leq e + \xi^* \\
 [\mathbf{w}^T \phi(x_i) + b] - y_i &\leq e + \xi
 \end{align*}
\]

As shown in formula 2.

\[
 \min \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^{n} (\xi + \xi^*)
\]

Where \( \phi \) is a nonlinear mapping function, \( \xi \) and \( \xi^* \) are relaxation factors, and \( C>0 \) is a balance factor. The Lagrangian optimization method can be transformed into a dual problem, which is reduced to the condition

\[
 \begin{align*}
 \sum_{j=1}^{n} (a_i - a_i^*) = 0, &\quad i = 1, \Lambda, n \\
 0 \leq a_i, a_i^* \leq C
 \end{align*}
\]

Solve under constraints

\[
 \begin{align*}
 \max \left(-\frac{1}{2} \sum_{i,j=1}^{n} (a_i - a_i^*) (a_j - a_j^*) K(x_i, x_j) - \varepsilon (a_i + a_i^*) + \sum_{i=1}^{n} y_i (a_i - a_i^*) \right)
 \end{align*}
\]
Where \( K(x_i, x_j) = \varphi(x_i) \varphi(x_j) \) is the kernel function, \( a_i \) and \( a_i^* \) is the Lagrangian multiplier. The regression function is obtained by solving:

\[
f(x) = \sum_{i=1}^{N} (a_i - a_i^*) K(x_i, x) + b
\]

Among them, \( N \) is the number of support vectors.

3. Test plan

3.1. Test device and sample
In this paper, an artificial pollution test platform for composite insulators is built, and the specific legal person circuit is shown in Figure 1. The main structural parameters are shown in Table 1.

![Test circuit diagram](fig1.png)

**Fig. 1** Test circuit diagram

| Tab. 1 Structural parameters of composite post insulators |
|----------------------------------------------------------|
| parameter                  | Value |
|-----------------------------|-------|
| Rated voltage (kV)          | 100   |
| Rated mechanical load (kN)  | 40    |
| Structure height (mm)       | 2150  |
| Umbrella spacing (mm)       | 75    |
| Big umbrella stretched out (mm) | 380 |
| Small umbrella stretched out (mm) | 280 |
| Electrical clearance (mm)   | 1970  |

3.2. Test principle and method
This test adopts the DL/T859——2004 "Artificial Pollution Test of Composite Insulators for High Voltage AC Systems". This standard can be applied to high-voltage pillar insulators for electrical power stations, cross-arm insulators for high-voltage lines and composite spacer insulators. The test adopts the salt spray flashover method, that is, based on the equivalent salt density method, a certain concentration of sodium chloride solution is used to equate the insulator's natural pollution and moisture situation. The test steps include the washing of the sample, the preparation of the salt solution, the pollution treatment of the sample, and the pressure test of the sample.

According to the recommendations of GB/T 16434-1996, the conductivity of tap water for cleaning the contaminated insulators should be less than or equal to 10\( \mu \)S/cm, and the water consumption depends on the surface area of the insulators, as shown in Table 2. In this paper, from the actual size of the insulator, the surface area of the umbrella skirt is calculated to be 2450cm\(^2\), so the water
consumption in this test is 500mL. According to the national standards (GB/T 5582-1993) and (GB/T 16434-1996), the corresponding salt density values for pollution levels are shown in Table 3. The salt solution concentration can be obtained according to formula (6). The pressure method adopts the uniform boost method to avoid the influence of the boost speed on the development of the local arc and the boost speed is controlled to 3kV/S.

Tab. 2 Relationship between surface area of insulator and water consumption of salt density measurement

| Insulator surface area, A(cm²) | Water consumption(ml) |
|--------------------------------|------------------------|
| ≤500                           | 100                    |
| 500<A≤1000                     | 200                    |
| 1000<A≤1500                    | 300                    |
| 1500<A≤2000                    | 400                    |
| 2000<A≤2500                    | 500                    |

Tab. 3 Salt content and conductivity of equivalent salt density for different pollution grades

| Pollution level | Salt density( mg/cm²) | Salt content( mg/100 mL) | Conductivity( S/m) |
|-----------------|-----------------------|--------------------------|--------------------|
| 0               | <0.03                 | <11.4                    | 0.0188             |
| I               | 0.03~0.06             | 11.4~22.8                | 0.0188~0.0368      |
| II              | 0.06~0.10             | 22.8~38.0                | 0.0368~0.0721      |
| III             | 0.10~0.25             | 38.0~95.0                | 0.0721~0.1754      |
| IV              | 0.25~0.35             | 95.0~133                 | 0.1754~0.2601      |

\[ N_i = \frac{S_A}{Q_W} \]  \hspace{1cm} (6)

In the formula: \(N_i\) is the number of milligrams of salt in a 100 mL solution, mg/100 mL; \(S\) is the amount of salt per unit area on the surface of the insulating part, that is the equivalent salt density, mg/cm²; \(A_S\) is the total area of the insulating parts of the insulator, cm²; \(Q_W\) is the amount of distilled water used to clean dirt, mL.

According to the DL/T859-2004 standard, the salt spray flashover test method should be repeated twice when the flashover process is uniformly boosted. In this paper, the corona initiation voltage and flashover voltage are measured 4 to 5 times, whichever is the error of the average value. The average value of all points not exceeding 10% is regarded as the flashover voltage \(U_f\) under the pollution degree, namely

\[ U_f = \frac{\sum U_i}{N} \]  \hspace{1cm} (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

In this paper, the flashover voltage of the insulators under different salt densities and gray densities is obtained through artificial pollution tests, and the salt gray density and pollution flashover voltage models are established by the support vector machine regression method of data mining technology.

The kernel function used this time is the radial basis (RBF) function $K(x, x_i) = \exp(-\sigma \|x - x_i\|^2)$, where $\sigma > 0$ is the parameter of the kernel function. In this paper, a total of 30 sets of experimental data are obtained, of which 20 sets of sample data are used as training samples, and the remaining 10 sets are prediction samples. The experiment uses svm-train and svm-predict in the software libsvm-3.17 to train and predict the model, and the prediction results are shown in Table 5.

| Tab. 4 Data of the samples |
|---------------------------|
| Salt density (mg/cm²) | Gray dense (mg/cm²) | $U_f$ (kV) |
| 0.05 | 0.7 | 157 |
| 0.15 | 0.71 | 142 |
| 0.18 | 0.7 | 139 |
| 0.06 | 1.01 | 148 |
| 0.19 | 1.03 | 130 |
| 0.289 | 1.01 | 125 |
| 0.22 | 2.9 | 112 |
| 0.152 | 2.2 | 121 |
| 0.06 | 2.11 | 133 |
| 0.113 | 1.62 | 130 |

| Tab. 5 Predictable results of the samples |
|-----------------------------------------|
| Salt density (mg/cm²) | Gray dense (mg/cm²) | Measured value | Predictive value | Error (%) |
| 0.148 | 0.7 | 141 | 135 | 4.3 |
| 0.18 | 2.14 | 119 | 126 | 5.9 |
| 0.41 | 2.13 | 107 | 100 | 6.5 |
| 0.054 | 0.58 | 162 | 160 | 1.2 |
| 0.03 | 1.02 | 158 | 150 | 5.1 |
| 0.053 | 2.5 | 132 | 130 | 1.5 |
| 0.39 | 0.49 | 120 | 124 | 3.3 |
| 0.151 | 1.46 | 127 | 132 | 3.9 |
| 0.22 | 4.1 | 105 | 101 | 3.8 |
| 0.272 | 5.03 | 99 | 106 | 7.1 |

It can be seen from Table 2 that the prediction accuracy of the flashover voltage under different salt densities and gray densities through the support vector machine regression model method is high, which can meet the actual operation requirements, which has very important guiding significance for the safe and stable operation of the power grid.
5. Conclusions
In order to improve the prediction accuracy of insulator pollution flashover voltage, this paper introduces the support vector machine regression method of data mining technology to study the flashover voltage under different salt and gray densities. The prediction result of the support vector machine regression model and the measured value have a maximum error of 7.1%, a minimum of 1.2%, and an average error of 4.3%. The prediction results show that the prediction error is small, which greatly improves the prediction accuracy of insulator pollution flashover voltage, it provides a certain technical reference for the anti-pollution flashover measures of insulators.

References
[1] ZHANG Zhijin, JIANG Xingliang, SUN Caixin. Summary of research on flashover mechanism of polluted insulators [J]. Power System Technology, 2008(16): 37-42.
[2] Farang A S. Estimation of polluted insulators flashover time using artificial neural networks [C] // IEEE Industrial and Commercial Power Systems Technical Conference . Philadelphia, PA, USA :IEEE, 1997:184-192.
[3] YANG Qing, SIMA Wen-xia, JIANG Xing-liang, et al. The building and application of a neural network model for forecasting the flashover voltage of the insulator in complex ambient conditions [J]. Proceedings of the CSEE, 2005, 25(13): 155-159.